University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Liu, C;Bousman, CA;Pantelis, C;Skafidas, E;Zhang, D;Yue, W;Everall, IP

Title:
Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations

Date:
2017-01-01

Citation:
Liu, C., Bousman, C. A., Pantelis, C., Skafidas, E., Zhang, D., Yue, W. & Everall, I. P. (2017). Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations. Translational Psychiatry, 7 (2), https://doi.org/10.1038/tp.2017.8.

Persistent Link:
https://hdl.handle.net/11343/258294

License:
CC BY

OPEN

Citation: Transl Psychiatry (2017) 7, e1037; doi:10.1038/tp.2017.8

www.nature.com/tp

ORIGINAL ARTICLE

Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations

C Liu[1,10], CA Bousman[1,2,3,4,10], C Pantelis[1,2,5,6], E Skafidas[1,2,5], D Zhang[7,8,9], W Yue[7,8,11] and IP Everall[1,2,5,6,11]

Genome-wide association studies have confirmed the polygenic nature of schizophrenia and suggest that there are hundreds or thousands of alleles associated with increased liability for the disorder. However, the generalizability of any one allelic marker of liability is remarkably low and has bred the notion that schizophrenia may be better conceptualized as a pathway(s) disorder. Here, we empirically tested this notion by conducting a pathway-wide association study (PWAS) encompassing 255 experimentally validated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways among 5033 individuals diagnosed with schizophrenia and 5332 unrelated healthy controls across three distinct ethnic populations; European-American (EA), African-American (AA) and Han Chinese (CH). We identified 103, 74 and 87 pathways associated with schizophrenia liability in the EA, CH and AA populations, respectively. About half of these pathways were uniquely associated with schizophrenia liability in each of the three populations. Five pathways (serotonergic synapse, ubiquitin mediated proteolysis, hedgehog signaling, adipocytokine signaling and renin secretion) were shared across all three populations and the single-nucleotide polymorphism sets representing these five pathways were enriched for single-nucleotide polymorphisms with regulatory function. Our findings provide empirical support for schizophrenia as a pathway disorder and suggest schizophrenia is not only a polygenic but likely also a poly-pathway disorder characterized by both genetic and pathway heterogeneity.

Translational Psychiatry (2017) **7,** e1037; doi:10.1038/tp.2017.8; published online 21 February 2017

INTRODUCTION

Schizophrenia is a severe psychiatric disorder characterized by significant genetic heterogeneity and is commonly referred to as a polygenic disorder.[1] This polygenicity was most recently highlighted in the largest genome-wide association study (GWAS) of schizophrenia that identified 128 single-nucleotide polymorphisms (SNPs) at 108 loci associated with the disorder,[2] although there are likely thousands of SNPs that contribute to its liability, many of which are population-specific. Thus, identifying schizophrenia-associated SNPs that are generalizable to diverse populations of people with the disorder is unlikely and suggests alternative approaches at identifying genetic markers of schizophrenia liability are needed.

One such approach is to examine SNPs within the biological pathway(s) in which they reside. Underlying this approach is the notion that schizophrenia is a pathway(s) disorder,[3] whereby one or a number of SNPs within a pathway could result in an increased liability to schizophrenia by altering sensitivity to environmental insults and/or disruption of brain development. In the context of schizophrenia, this pathway approach has been applied in a variety of forms ranging from pathway clustering analysis,[4] where SNPs in key genes within a single pathway are examined, to *post-hoc* pathway enrichment analyses of candidate SNP-sets using the SNP ratio test[5] or bioinformatics resources (for example, Ingenuity Pathway Analysis).[6–8] These approaches undoubtedly have a pathway focus but provide an incomplete examination of the compendium of known human biological pathways. Our primary aim was to conduct a comprehensive pathway-wide association study (PWAS) of schizophrenia. Here, we report results of that analysis in which we tested 255 biological pathway-based SNP-sets for their association and potential function in schizophrenia in three ancestral distinct populations.

MATERIALS AND METHODS

Data sources

GWAS data from individuals with schizophrenia ($n = 5033$) and healthy controls ($n = 5332$) across three distinct ethnic populations; European-American (EA) (2455 schizophrenia, 2826 controls), Han Chinese (CH) (1625 schizophrenia, 1527 controls) and African-American (AA) (953 schizophrenia, 979 controls) were obtained (Table 1; Supplementary Table S1). EA and AA GWAS data were collected by the Genetic Association Information Network (GAIN) and nonGAIN projects and were obtained through the

[1]Department of Psychiatry, Melbourne Neuropsychiatry Centre, The University of Melbourne and Melbourne Health, Carlton South, VIC, Australia; [2]Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, VIC, Australia; [3]Department of General Practice, The University of Melbourne, Parkville, VIC, Australia; [4]Centre for Human Psychopharmacology, Swinburne University of Technology, Hawthorn, VIC, Australia; [5]Department of Electrical and Electronic Engineering, Centre for Neural Engineering (CfNE), University of Melbourne, Carlton South, VIC, Australia; [6]NorthWestern Mental Health, Melbourne, VIC, Australia; [7]Institute of Mental Health, The Sixth Hospital, Peking University, Beijing, China; [8]Key Laboratory of Mental Health, Ministry of Health & National Clinical Research Center for Mental Disorders (Peking University), Beijing, China and [9]Peking-Tsinghua Joint Center for Life Sciences/PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. Correspondence: Dr CA Bousman, Department of Psychiatry, Melbourne Neuropsychiatry Centre, University of Melbourne, 161 Barry Street, Level 3, Carlton South, VIC 3053, Australia.
E-mail: cbousman@unimelb.edu.au
[10]Joint first authorship.
[11]Joint last authorship.
Received 19 December 2016; accepted 30 December 2016

Table 1. Sample size and SNPs available for analysis following quality control procedures

Ancestry	*Cases*			*Controls*			*Platform*	*SNPs*	*Source*
	N	*Age (s.d.)*	*Male/female*	N	*Age (s.d.)*	*Male/female*			
European-American									
Discovery	972	43.71 (11.29)	681/291	1248	50.61 (17.05)	570/678	Affymetrix 6.0	691 822	GAIN
Replication	879	42.26 (11.93)	609/270	1132	49.91 (15.77)	569/563	Affymetrix 6.0	691 822	Non-GAIN
Han Chinese									
Discovery	1125	35.97 (7.82)	555/570	1034	36.60 (10.35)	476/558	Illumina Zhonghua 8	800 509	CSCG
Replication	454	36.48 (7.98)	262/192	411	36.40 (8.16)	187/224	Illumina Zhonghua 8	800 509	CSCG
African-American	896	43.30 (10.12)	558/338	906	45.16 (13.03)	344/562	Affymetrix 6.0	818 941	GAIN

Abbreviations: CSCG, Chinese Schizophrenia Collaboration Group; GAIN, Genetic Association Information Network; SNPs, single-nucleotide polymorphisms.

database of Genotypes and Phenotypes (dbGaP, phs000021.v1.p1 and phs000167.v1.p1)[9] with ethics approval by The University of Melbourne Human Ethics Committee (#1340723). CH GWAS data were collected from multiple collaborating hospitals included in the Chinese Schizophrenia Collaboration Group (see Supplementary Materials for details).[10] For the EA and CH cohorts two independent datasets were available. One was used as a discovery dataset (EA: 1215 schizophrenia cases and 1442 healthy control subjects; CH: 1159 schizophrenia cases and 1089 healthy control subjects) and the other a validation dataset (EA: 1240 schizophrenia and 1384 controls; CH: 466 schizophrenia and 438 controls).

Quality control and population stratification

We adhered to a previously published quality control protocol[11] with the exception of procedures related to identity by descent and population stratification (see Supplementary Materials for details). Population stratification was mitigated using spatial ancestry analysis (SPA).[12] The SPA European model was used for the analysis of the EA data, whereas the SPA worldwide model was used for the AA and CH datasets. We used all available genotypes to calculate the geographic coordinates of latitude and longitude and set inclusion boundaries (Supplementary Figure S1). The final sample sizes and SNPs available for analysis following quality control are presented in Table 1 (see Supplementary Table S1 for details on the number of individuals and SNPs removed at each quality control step).

Mapping SNPs to genes and pathways

SNPs surviving quality control were mapped to gene loci using the annotation provided by the National Center for Biotechnology Information (see Supplementary Methods for details). Genes were then mapped to pathways curated by the Kyoto Encyclopedia of Genes and Genomes (KEGG, Release 76.0, 1 October, 2015),[13] which includes 301 human pathways from six main categories (metabolism, genetic Information processing, environmental information processing, cellular processes, organismal systems and human diseases). A mega KEGG pathway (metabolic pathways, hsa01100) that encompasses several other pathways was excluded, leaving 300 pathways available for further analysis.

Pathway-wide association analysis

The analysis pipeline used to assess each of the 300 KEGG pathways for their association with schizophrenia is depicted in Figure 1, evolving from our previously published pathway analysis pipeline.[14] For each pathway the discovery dataset for the EA and CH cohorts as well as the single dataset available for the AA cohort were randomly split (maintaining the case:control ratio of the full dataset) 100 times into two subsets, a SNP (that is, feature) selection set (80% of the participants) and a test set (20% of participants). Within each SNP selection set, 80% of participants were randomly selected 10 times (maintaining the case:control ratio of the full dataset) and the resulting subsets were subjected to the maximum relevance minimum redundancy (mRmR) feature selection procedure (blue box, Figure 1).[15] The mRmR procedure was chosen as an alternative to *P*-value-based feature selection procedures that are dependent on sample size and do not necessarily result in feature sets that maximize relevance and minimize redundancy (that is, increase mutual information; see Supplementary Materials and Supplementary Figure S2 for details and a comparison of the two feature selection methods in our datasets). This procedure resulted in 300 SNP sets, one for each of the KEGG pathways (Supplementary Table S2). Among these 300 SNP sets, 45 sets containing less than two features (SNPs) at one or more of the 100 iterations were excluded from further analysis, as our algorithm requires two or more features to fit a model.

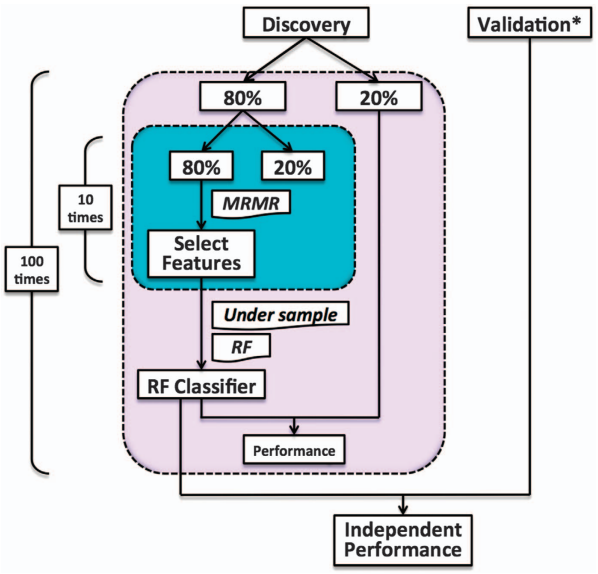


Figure 1. Pathway-wide association analysis pipeline. *Only the European American population and the Han Chinese population have independent validation dataset. MRMR, maximum relevance minimum redundancy; RF, random forest.

The 255 SNP sets were then used to build 255 classifiers, one for each KEGG pathway, via a random forest algorithm with default parameters (R package: 'randomForest') using 80% of the discovery dataset followed by testing of the classifiers in the remaining 20% of the discovery dataset. To address inherent imbalances in the case:control ratios of our datasets, under-sampling of the majority class (cases or controls) of each dataset was employed before running the random forest algorithm, as this strategy has previously been shown to be useful for classification in the presence of imbalanced classes.[16,17]

To assess the overall performance of each pathway classifier, the random forest model derived from the selected SNPs for each pathway within the 80% discovery set was applied to the 20% test set as well as the independent validation dataset, with the exception of the AA cohort for

which an independent validation dataset was not available. In addition, within the 20% test set and independent validation dataset, case–control labels of all individuals were randomly permuted and the random forest model for each pathway derived from the 80% discovery set was applied, with the exception of the AA cohort. Point estimates and 95% confidence intervals for five performance metrics (accuracy, sensitivity, specificity, area under the receiver operating characteristic curve (AUC) and odds ratio (OR)) were calculated for each of the pathways using the independent validation dataset for EA and CH cohorts, and the 20% test set for the AA cohort. *P*-values for each of the 255 pathways were generated by comparing the mean OR (based on 100 iterations) from the independent (validation for AA) dataset with the corresponding mean OR from the permutation dataset using a *t*-test. The Benjamini–Hochberg (BH) procedure was used to adjust for multiple comparisons.[18] Furthermore, using the independent validation dataset for the EA and CH cohorts, we calculated the Nagelkerke R^2 (see Supplementary Materials for details) to estimate the variance in schizophrenia liability explained by each of the 255 SNP sets used to construct the pathway classifiers. The complete annotated computer script used to conduct the pathway-wide association analysis is available upon request.

To further evaluate our pathway analysis pipeline, we selected 129 previously identified gene ontology (GO) pathways associated with schizophrenia and applied our pipeline to each of the 129 pathways in all three populations.

Functional analysis of SNPs in candidate pathways

To assess whether the mRmR feature selection approach was capable of selecting informative features and to evaluate the potential functional relevance of selected features, we utilized the brain expression quantitative trait loci (eQTLs) dataset obtained from the genotype-tissue expression (GTEx) portal v6.0,[19] as well as the functional annotation information obtained from the RegulomeDB, a database that annotates SNPs with known and predicted regulatory elements.[20] We hypothesized that the selected features with greater appearance rates within significant schizophrenia liability pathways would be enriched for functional SNPs compared with SNP sets derived from non-significant pathways.

eQTL analysis. SNP sets representing pathways associated with schizophrenia in all three cohorts were further assessed as potential *cis*-eQTLs using genotype and gene expression data derived from human postmortem frontal cortex (Brodmann area 9) of 92 donors included in the GTEx portal.[19] For each of the three cohorts, SNPs within each of our candidate pathways was assigned an appearance rate based on the number of times (out of 100 iterations) the SNP represented the candidate pathway during our feature selection procedure (that is, mRmR) described above. SNPs were then grouped into quartiles based on their appearance rate (that is, 0–25%; 26–50%; 51–75% and 76–100%) and the proportion of SNPs within each quartile associated (alpha threshold = 0.05) with expression of its corresponding gene was calculated. For comparison, the same analysis was conducted on 152, 181 and 168 non-significant pathways in the EA, CH and AA population, respectively. A one-sample *t*-test was used to determine if the proportion of eQTLs observed in our overlapping pathway SNP sets differed from the SNP sets derived from non-significant pathways.

Regulome analysis. To investigate the potential functional significance of selected features beyond eQTLs, including DNA–protein interaction (TF-binding motif, DNase footprint) and DNA–RNA interaction (microRNA-binding motif, long non-coding RNA), we utilized the RegulomeDB (http://www.regulomedb.org).[20] Similar to the eQTL analysis, SNPs were grouped into quartiles based on their appearance rate and a weighted Regulome score was computed for each quartile group as well as SNP sets derived

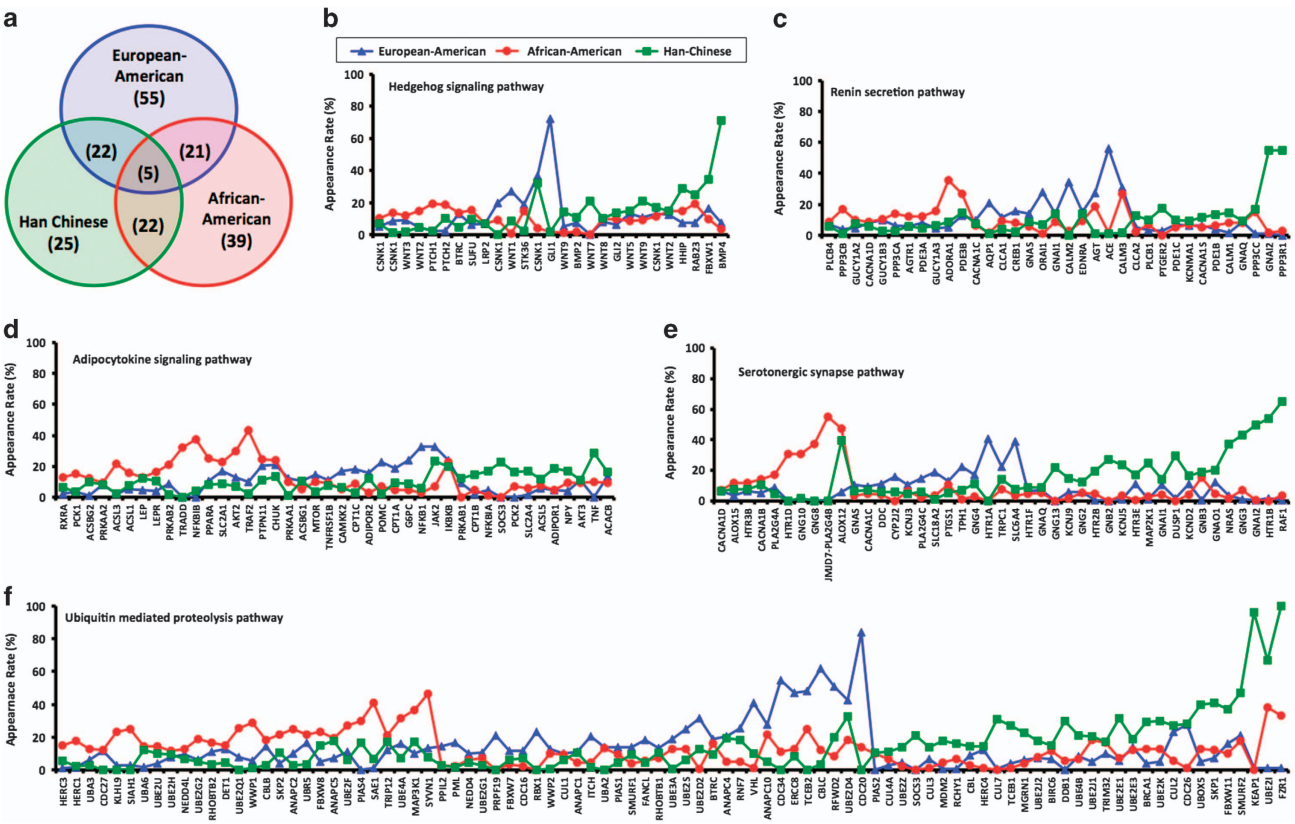


Figure 2. (**a**) Venn diagram showing the degree of overlap in pathways identified as a liability for schizophrenia across the three ancestrally distinct populations. (**b–f**) Relative contribution (measured by appearance rate in our feature selection procedure) of genes within the five common schizophrenia liability pathways by ancestry. Only genes with appearance rates summing to 20% or greater across the three populations are shown for each pathway (see Supplementary Table S6 for appearance rates of all single-nucleotide polymorphisms (SNPs) in each of the five pathways). Each gene's appearance rate is normalized to the number of haplotype-tagging SNPs (threshold $R^2 = 0.50$) within the gene.

Table 2. Performance metrics for the four common pathways across the three ethnically distinct cohorts

Point estimate (95% CI)	*Ubiquitin-mediated proteolysis*			*Serotonergic synapse*			*Hedgehog signaling*			*Adipocytokine signaling pathway*			*Renin secretion*		
	EA	*AA*	*CH*	*EA*	*AA*	*CH*	*EA*	*AA*	*CH*	*EA*	*AA*	*CH*	*EA*	*AA*	*CH*
Sensitivity	0.530 (0.526–0.534)	0.515 (0.508–0.522)	0.519 (0.514–0.523)	0.522 (0.519–0.525)	0.506 (0.499–0.513)	0.491 (0.486–0.496)	0.513 (0.509–0.516)	0.510 (0.503–0.517)	0.495 (0.491–0.500)	0.520 (0.517–0.523)	0.515 (0.508–0.522)	0.491 (0.486–0.496)	0.524 (0.521–0.527)	0.508 (0.501–0.516)	0.490 (0.485–0.494)
Specificity	0.490 (0.487–0.493)	0.505 (0.498–0.512)	0.506 (0.502–0.511)	0.491 (0.487–0.494)	0.508 (0.501–0.515)	0.524 (0.519–0.529)	0.504 (0.500–0.508)	0.504 (0.496–0.512)	0.522 (0.518–0.527)	0.505 (0.501–0.508)	0.507 (0.499–0.514)	0.525 (0.520–0.529)	0.492 (0.488–0.495)	0.505 (0.497–0.513)	0.524 (0.520–0.529)
AUC	0.515 (0.513–0.517)	0.518 (0.514–0.523)	0.520 (0.517–0.523)	0.508 (0.506–0.510)	0.509 (0.505–0.514)	0.512 (0.509–0.515)	0.511 (0.509–0.513)	0.511 (0.505–0.516)	0.511 (0.508–0.515)	0.517 (0.514–0.519)	0.514 (0.509–0.518)	0.509 (0.506–0.512)	0.511 (0.509–0.514)	0.508 (0.503–0.514)	0.511 (0.508–0.514)
Accuracy	0.512 (0.510–0.515)	0.510 (0.506–0.515)	0.512 (0.510–0.515)	0.509 (0.507–0.510)	0.507 (0.503–0.511)	0.508 (0.505–0.511)	0.509 (0.507–0.511)	0.507 (0.502–0.512)	0.509 (0.506–0.512)	0.513 (0.511–0.515)	0.511 (0.506–0.515)	0.509 (0.506–0.511)	0.510 (0.508–0.512)	0.509 (0.502–0.512)	0.508 (0.505–0.511)
Odds ratio	1.087 (1.069–1.104)	1.104 (1.065–1.144)	1.113 (1.088–1.138)	1.057 (1.041–1.073)	1.072 (1.038–1.107)	1.070 (1.043–1.098)	1.073 (1.056–1.091)	1.077 (1.036–1.117)	1.080 (1.054–1.105)	1.108 (1.09–1.127)	1.109 (1.069–1.150)	1.072 (1.046–1.098)	1.069 (1.053–1.086)	1.080 (1.036–1.124)	1.065 (1.041–1.090)
Adjusted-*P*	1.07×10^{-7}	0.0139	0.00242	0.000285	0.0492	0.00420	1.55×10^{-7}	0.00615	2.41×10^{-5}	2.89×10^{-14}	0.0103	0.0121	3.03×10^{-8}	0.0139	0.00597
Nagelkerke R^{2a}	0.0005	—	0.0057	0.0017	—	0.0015	0.0003	—	0.0033	0.0011	—	0.0005	0.0010	—	0.0012

Abbreviations: AA, African-American cohort; AUC, area under the curve; CH, Han Chinese cohort; EA, European-American cohort. [a]Refer to the Supplementary Materials for details.

from non-significant pathways. A one-sample *t*-test was used to determine if the weighted Regulome score in our candidate SNP sets differed from SNP sets derived from non-significant pathways (see Supplementary Materials for details).

RESULTS

Pathway-wide association analysis

Of the 255 pathways examined, 103, 74 and 87 pathways were significantly associated with schizophrenia liability in the EA, CH and AA cohorts, respectively (Supplementary Figure S3; Supplementary Table S3–5). Examination of the overlap between the cohorts showed 55, 25 and 39 pathways were uniquely associated with schizophrenia liability in the EA, CH and AA cohorts, respectively (Figure 2a). Five pathways (serotonergic synapse, ubiquitin mediated proteolysis, hedgehog signaling, adipocytokine signaling and renin secretion) were shared across all three cohorts (Figure 2a). However, the relative contribution of the SNPs and genes within these five pathways differed considerably by ancestry (Figures 2b–e; Supplementary Table S6) and a small subset of genes were shared across two or three common pathways (Supplementary Figure S4). Sensitivity, specificity, AUC, accuracy and ORs were modest for each of the five pathways and no pathway explained more than one percent ($R^2 = 0.03–0.57\%$) of the variance in the liability to schizophrenia (Table 2). Combining the selected features from the five shared pathways had minimal impact on the variance explained ($R^2 = 0.31–0.57\%$), although when features from all significant pathways were assessed the variance explained ranged from 0.66% in the EA cohort to 2.46% in the CH cohort. Furthermore, among the 129 previously identified schizophrenia-associated GO pathways our analysis pipeline replicated 45, 20 and 56 of these pathways in the EA, CH and AA populations (Supplementary Table S7).

Functional analysis of SNPs in candidate pathways

Analysis of SNPs selected to represent the five pathways that overlapped in the three populations showed SNPs with greater appearance rates had a greater probability of being an eQTL or having some other regulatory function (Figure 3; Supplementary Figures S5 and S6). SNPs that appeared >50% of the time during our feature selection procedure were more likely to be functional compared with 100 random SNP sets of equal size, with the exception of the serotonergic synapse SNPs in EAs. Likewise, SNPs with appearance rates >75% also had a higher probability to be functional, although for six of the pathway-population pairs (Figure 3) our feature selection procedure did not identify SNP sets enriched for functional SNPs.

DISCUSSION

The notion that schizophrenia is a pathway disease has only recently been proposed[3] and as such empirical testing of this notion is limited. We conducted a PWAS of schizophrenia in three ancestrally distinct cohorts. We found evidence of pathway heterogeneity in schizophrenia liability, identified five pathways conferring liability across populations and showed that the SNP sets representing these five pathways were enriched for SNPs with regulatory functions.

Pathway heterogeneity has only recently been discussed in the context of schizophrenia[4] but has been well characterized in other diseases such as cancer.[21,22] Pathway heterogeneity builds on and encompasses the concept of genetic heterogeneity in that it postulates a disorder is a result of one or more perturbations in one or more of a multiple number of pathways. Supporting this notion, we found that nearly half (47%) of the pathways we tested were uniquely associated with schizophrenia liability in only one of the three populations we examined—raising the possibility that

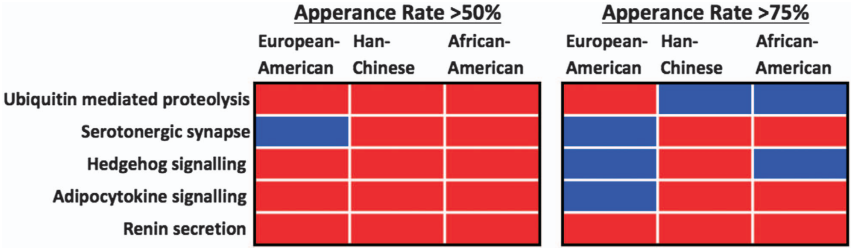


Figure 3. Summary of the functional analysis performed on single-nucleotide polymorphisms (SNPs) with an appearance rate of >50% and >75% during feature selection in the five shared pathways across the three populations. Red boxes indicate the proportion of functional SNPs (based on GTEx or RegulomeDB data) was significantly greater compared with SNP sets derived from non-significant pathways within that population. Blue boxes indicate the proportion of functional SNPs was significantly lower compared with SNP sets derived from non-significant pathways within that population.

schizophrenia is not only a polygenic but also a poly-pathway disorder. In fact, all pathways had an OR < 1.35, suggesting multiple pathways of small effect collectively contribute to schizophrenia liability.

Furthermore, our results suggest disruption of certain pathways may be necessary (but perhaps not sufficient) for the development of schizophrenia across populations. About one-fourth (27%) of the pathways we tested were associated with liability to schizophrenia in two or more of the populations, among which five pathways were associated with schizophrenia liability in all three cohorts. These pathways included the serotonergic synapse, ubiquitin mediated proteolysis, hedgehog signaling, renin secretion and adipocytokine signaling, all of which have been implicated in schizophrenia and/or related phenotypes.

A number of post-mortem, functional neuroimaging and peripheral biomarker studies have implicated the serotonergic system in the pathophysiology of schizophrenia (for review see: ref. 23) and many atypical antipsychotic agents (for example, clozapine, olanzapine) are potent serotonin receptor 2A antagonists.[24] Thus, identification of the serotonergic synapse pathway in the current study is perhaps not surprising. In fact, the largest schizophrenia GWAS to date found SNPs in three genes (*CACNA1C*, *ITPR3* and *CYP2D6*) within the serotonergic synapse pathway reached GWAS significance ($P < 5 \times 10^{-8}$) and SNPs in another 17 genes within this pathway were nominally significant ($P < 1 \times 10^{-5}$).[2] Furthermore, a recent gene-set enrichment analysis of the SZGene database[25] identified 24 pathways significantly enriched for schizophrenia candidate genes among which the serotonin receptor signaling pathway was ranked second.[7]

The ubiquitin mediated proteolysis pathway (UPP), a critical system for the removal of damaged/toxic proteins in the cell, has been shown to be dysregulated at the transcript[26–29] and protein levels[30] in both peripheral and central tissue among individuals with schizophrenia. In addition, peripheral transcript levels within the UPP have been associated with positive symptom severity[31] and a recent copy number variant meta-analysis in schizophrenia, autism and intellectual disability revealed that two ubiquitin-related gene-ontologies were highly enriched with schizophrenia-associated copy number variants.[32] Furthermore, animal studies have suggested that UPP has an important role in regulating synaptic growth and neural circuits,[33,34] and demand on the UPP at pre-synaptic and post-synaptic terminals may in part link dysfunction of the UPP to increased schizophrenia liability.[35]

The hedgehog signaling pathway is a key regulator of oligodendrocyte production,[36,37] dopaminergic neuron development,[38] and promotes brain expression of *disc1*, a candidate gene in schizophrenia.[39] The pathway's most well characterized ligand, sonic hedgehog, regulates the generation of functional synaptic contacts,[40,41] and is abundant in the adult human central nervous system.[42,43] Furthermore, hedgehog signaling interacts with the UPP[44] and has been implicated in the 'two-hit' hypothesis of schizophrenia by which disruption of the pathway during brain development primes the central nervous system for a pathologic response to a second hit in later life.[45]

The renin secretion pathway is typically associated with regulation of arterial blood pressure, thirst and thermoregulation via the kidney secreted enzyme renin and its interaction with the renin–angiotensin–aldosterone system. Epidemiological studies have reported up to 25% of schizophrenia have polydipsia (excessive thirst)[46] and in general exhibit dysregulation of body temperature.[47] Rodent studies have demonstrated renin is also synthesized in the brain[48] and has considerable effects on anxiety-related behaviors and cognition (for example, memory).[49] In the brain, renin is proposed to enzymatically process angiotensinogen to angiotensin, which is then further processed by angiotensin-converting enzyme (ACE).[50] ACE activity has been shown to modulate dopamine turnover[51] and abnormal levels of ACE in cerebrospinal fluid have been reported in individuals with schizophrenia,[52,53] albeit potential neurotropic and length of illness effects have been noted.[54,55] The interaction between angiotensin II (AT II), a neuropeptide substrate for ACE, and central dopamine has also been associated with schizophrenia.[56,57] Moreover, numerous genetic studies suggested polymorphisms in *ACE* are associated with susceptibility to schizophrenia and major depression.[58–62]

The adipocytokine signaling pathway is a collective destination of cytokines secreted by the adipose tissue. Since the first adipocytokine leptin was discovered in 1994,[63] hundreds of adipocytokines have been found, such as adiponentin, tumor necrosis factor-alpha and members of the interleukin family. Increased expression of tumor necrosis factor-alpha and a number of interleukins have recently been proposed as markers of schizophrenia in brain[64] and blood.[65] Furthermore, adipocytokines are recognized not only as regulators of energy metabolism, but also as factors that may be associated with mental disorders. Decreased serum levels of adiponectin have been identified in major depressive disorder and schizophrenia,[66–68] and serum levels of leptin correlate with less severe positive symptoms in schizophrenia patients[69] and may regulate the mesolimbic dopamine system.[70]

Despite the novelty and many strengths of our study, our findings should be interpreted in the context of several limitations. First, the detection of population differences in schizophrenia liability at the pathway level may, in part, be a result of sampling, allelic frequency and/or linkage disequilibrium differences across the populations studied. These potential confounding factors may also explain why we only identified five overlapping pathways rather than the expected 10 overlaps given the number of significant pathways identified in each population. Although we

attempted to reduce these confounding influences by selecting features independently in each population using an mRmR approach, complete restraint of these confounds is not possible and as such our results should be interpreted with caution. Second, an independent validation dataset was not available for the AA population and as such all estimates were based on the 20% holdout dataset derived from the discovery dataset. This may have resulted in over-estimation of associations within this population, although the random forest algorithm we employed internally mitigates this potential bias via the out-of-bag error estimate mechanism.[71] Third, the dopamine hypothesis of schizophrenia is an enduring, widely accepted, idea but among the three populations we studied the dopaminergic synapse pathway was only a significant liability pathway in the EA (OR = 1.108, 95% CI = 1.088–1.127; B-H $P = 8.18E-13$) and CH (OR = 1.054, 95% CI 1.027-1.08; B-H $P = 0.050$) populations. Although, we failed to detect this pathway in the AA population, the expected trend was evident (OR = 1.056, 95% CI = 1.011–1.102; B-H $P = 0.429$) and is likely a result of the smaller sample size available for this population. Fourth, our analysis did not look at potential clinical subtypes of schizophrenia based on symptom profiles, despite recognition that schizophrenia is a broader spectrum disorder including a range of symptoms. A PWAS of clinical subtypes may lead to stronger associations by reducing the noise associated with the broad schizophrenia phenotype. However, the clinical symptom data available for this study was inconsistent or minimal across the three cohorts inhibiting such an analysis. Fifth, SNPs eligible for inclusion in our analysis were limited to those that were within a gene using a narrow '5 and 3' intergenic window (2 and 0.5 kbp, respectively). Intergenic SNPs are known to play functional regulatory roles on genes[72] and as such exclusion of more distal intergenic SNPs may have biased our results. In the most recent schizophrenia GWAS,[2] 45% (57) of the 128 SNPs identified were intergenic but 52% (30) of these intergenic SNPs are in linkage disequilibrium ($R^2 \geqslant 0.50$) with one or more SNPs within a gene according to our intergenic window. Thus, our intergenic window was capable of capturing a majority (79%, $n = 101$) of the 128 SNPs, suggesting the bias conferred by our SNP inclusion criteria are likely modest. Finally, our feature selection procedure (mRmR) resulted in the loss of many pathways containing smaller SNP pools. In total, 45 pathways cataloged within KEGG were not included in our pathway association study as the number of SNPs selected to represent these pathways was fewer than the required number of SNPs (that is, 2) to run our random forest algorithm.

In conclusion, our results empirically support the notion that schizophrenia is a pathway disorder and further suggest that there is a considerable amount of pathway heterogeneity within and across different ethnic populations. We also identified five pathways that may serve as harbors of genotypic markers for schizophrenia across populations. However, future application of our pathway-wide association approach in larger cohorts as well as among ethnic groups not examined here are required before firm conclusions can be drawn.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We acknowledge the financial support of the Brain and Behavior Research Foundation (NARSAD) Young Investigator Award (CAB, Grant# 20526), National Key Research and Development Program of China (2016YFC1307001) and National Key Technology R&D Program of China (2015BAI13B01). CAB was supported by a University of Melbourne Ronald Phillip Griffith Fellowship. CP was supported by a NHMRC Senior Principal Research Fellowship (628386 and 1105825).

Funding support for the Genome-Wide Association of Schizophrenia Study was provided by the National Institute of Mental Health (R01 MH67257, R01 MH59588, R01 MH59571, R01 MH59565, R01 MH59587, R01 MH60870, R01 MH59566, R01 MH59586, R01 MH61675, R01 MH60879, R01 MH81800, U01 MH46276, U01 MH46289 U01 MH46318, U01 MH79469 and U01 MH79470) and the genotyping of samples was provided through the Genetic Association Information Network (GAIN). The datasets used for the analyses described in this manuscript were obtained from the database of Genotypes and Phenotypes (dbGaP) found at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000021.v1.p1 and phs000167.v1.p1. Samples and associated phenotype data for the Genome-Wide Association of Schizophrenia Study were provided by the Molecular Genetics of Schizophrenia Collaboration (PI: Pablo V. Gejman, Evanston Northwestern Healthcare (ENH) and Northwestern University, Evanston, IL, USA).

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171) and Science Care (X10S172). The Laboratory, Data Analysis and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 and DA033684 and to contract N01MH000028. Statistical methods development grants were made to the University of Geneva (MH090941 and MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina—Chapel Hill (MH090936 and MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St. Louis (MH101810) and the University of Pennsylvania (MH101822). The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 5/12/2016.

REFERENCES

1 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.
2 Ripke S, Neale BM, Corvin A, Walters JT, Farh K-H, Holmans PA *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
3 Sullivan PF. Puzzling over schizophrenia: schizophrenia as a pathway disease. *Nat Med* 2012; **18**: 210–211.
4 Hatzimanolis A, McGrath JA, Wang R, Li T, Wong PC, Nestadt G *et al.* Multiple variants aggregate in the neuregulin signaling pathway in a subset of schizophrenia patients. *Transl Psychiatry* 2013; **3**: e264.
5 O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D *et al.* Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* 2011; **16**: 286–292.
6 Inada T, Koga M, Ishiguro H, Horiuchi Y, Syu A, Yoshio T *et al.* Pathway-based association analysis of genome-wide screening data suggest that genes associated with the gamma-aminobutyric acid receptor signaling pathway are involved in neuroleptic-induced, treatment-resistant tardive dyskinesia. *Pharmacogenet Genomics* 2008; **18**: 317–323.
7 Sun J, Jia P, Fanous AH, van den Oord E, Chen X, Riley BP *et al.* Schizophrenia gene networks and pathways and their applications for novel candidate gene selection. *PLoS One* 2010; **5**: e11351.
8 Jia P, Wang L, Meltzer HY, Zhao Z. Common variants conferring risk of schizophrenia: a pathway analysis of GWAS data. *Schizophr Res* 2010; **122**: 38–42.
9 Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M *et al.* New models of collaboration in genome-wide association studies: the genetic association information network. *Nat Genet* 2007; **39**: 1045–1051.
10 Yu H, Yan H, Li J, Li Z, Zhang X, Ma Y *et al.* Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. *Mol Psychiatry* 2016; advance online publication, 6 December 2016; doi: 10.1038/mp.2016.212 (e-pub ahead of print).
11 Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564–1573.
12 Yang WY, Novembre J, Eskin E, Halperin E. A model-based approach for analysis of spatial structure in genetic data. *Nat Genet* 2012; **44**: 725–731.
13 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**: 27–30.

14 Skafidas E, Testa R, Zantomio D, Chana G, Everall I, Pantelis C. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Mol Psychiatry* 2012; **19**: 504–510.
15 Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell* 2005; **27**: 1226–1238.
16 Garcia S, Herrera F. Evolutionary undersampling for classification with imbalanced datasets: proposals and taxonomy. *Evol Comput* 2009; **17**: 275–306.
17 Drummond C, Holte R. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling. Workshop on Learning from Imbalanced Datasets II. ICML: Washington DC, 2003.
18 Benjamini Y, Hochberg Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc B Met* 1995; **57**: 289–300.
19 GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013; **45**: 580–585.
20 Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–1797.
21 Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 2008; **321**: 1801–1806.
22 Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012; **490**: 61–70.
23 Chana G, Bousman CA, Money TT, Gibbons A, Gillett P, Dean B *et al.* Biomarker investigations related to pathophysiological pathways in schizophrenia and psychosis. *Front Cell Neurosci* 2013; **7**: 95.
24 Meltzer H. Mechanism of action of atypical antipsychotic drugs. In: Davis KL, Charney D, *et al.* (eds). *Neuropsychopharmacology—5th Generation of Progress*. Lippincott, Williams, & Wilkins: Philadelphia, PA, USA, 2002.
25 Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008; **40**: 827–834.
26 Middleton FA, Mirnics K, Pierri JN, Lewis DA, Levitt P. Gene expression profiling reveals alterations of specific metabolic pathways in schizophrenia. *J Neurosci* 2002; **22**: 2718–2729.
27 Altar CA, Jurata LW, Charles V, Lemire A, Liu P, Bukhman Y *et al.* Deficient hippocampal neuron expression of proteasome, ubiquitin, and mitochondrial genes in multiple schizophrenia cohorts. *Biol Psychiatry* 2005; **58**: 85–96.
28 Bousman CA, Chana G, Glatt SJ, Chandler SD, Lucero GR, Tatro E *et al.* Preliminary evidence of ubiquitin proteasome system dysregulation in schizophrenia and bipolar disorder: convergent pathway analysis findings from two independent samples. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 494–502.
29 Arion D, Corradi JP, Tang S, Datta D, Boothe F, He A *et al.* Distinctive transcriptome alterations of prefrontal pyramidal neurons in schizophrenia and schizoaffective disorder. *Mol Psychiatry* 2015; **20**: 1397–1405.
30 Rubio MD, Wood K, Haroutunian V, Meador-Woodruff JH. Dysfunction of the ubiquitin proteasome and ubiquitin-like systems in schizophrenia. *Neuropsychopharmacology* 2013; **38**: 1910–1920.
31 Bousman CA, Chana G, Glatt SJ, Chandler SD, May T, Lohr J *et al.* Positive symptoms of psychosis correlate with expression of ubiquitin proteasome genes in peripheral blood. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 1336–1341.
32 Pescosolido MF, Gamsiz ED, Nagpal S, Morrow EM. Distribution of disease-associated copy number variants across distinct disorders of cognitive development. *J Am Acad Child Adolesc Psychiatry* 2013; **52**: 414–430 e414.
33 DiAntonio A, Haghighi AP, Portman SL, Lee JD, Amaranto AM, Goodman CS. Ubiquitination-dependent mechanisms regulate synaptic growth and function. *Nature* 2001; **412**: 449–452.
34 Hamilton AM, Zito K. Breaking it down: the ubiquitin proteasome system in neuronal morphogenesis. *Neural Plast* 2013; **2013**: 196848.
35 Sheng ZH, Cai Q. Mitochondrial transport in neurons: impact on synaptic homeostasis and neurodegeneration. *Nat Rev Neurosci* 2012; **13**: 77–93.
36 Orentas DM, Hayes JE, Dyer KL, Miller RH. Sonic hedgehog signaling is required during the appearance of spinal cord oligodendrocyte precursors. *Development* 1999; **126**: 2419–2429.
37 Cunliffe VT, Casaccia-Bonnefil P. Histone deacetylase 1 is essential for oligodendrocyte specification in the zebrafish CNS. *Mech Dev* 2006; **123**: 24–30.
38 Luo SX, Huang EJ. Dopaminergic neurons and brain reward pathways: from neurogenesis to circuit assembly. *Am J Pathol* 2016; **186**: 478–488.
39 Boyd PJ, Cunliffe VT, Roy S, Wood JD. Sonic hedgehog functions upstream of disrupted-in-schizophrenia 1 (disc1): implications for mental illness. *Biol Open* 2015; **4**: 1336–1343.
40 Angot E, Loulier K, Nguyen-Ba-Charvet KT, Gadeau AP, Ruat M, Traiffort E. Chemoattractive activity of sonic hedgehog in the adult subventricular zone modulates the number of neural precursors reaching the olfactory bulb. *Stem Cells* 2008; **26**: 2311–2320.
41 Hor CH, Tang BL. Sonic hedgehog as a chemoattractant for adult NPCs. *Cell Adh Migr* 2010; **4**: 1–3.
42 Dellovade T, Romer JT, Curran T, Rubin LL. The hedgehog pathway and neurological disorders. *Annu Rev Neurosci* 2006; **29**: 539–563.
43 Ahn S, Joyner AL. *In vivo* analysis of quiescent adult neural stem cells responding to Sonic hedgehog. *Nature* 2005; **437**: 894–897.
44 Maniatis T. A ubiquitin ligase complex essential for the NF-kappaB, Wnt/Wingless, and Hedgehog signaling pathways. *Genes Dev* 1999; **13**: 505–510.
45 Maynard TM, Sikich L, Lieberman JA, LaMantia AS. Neural development, cell-cell signaling, and the "two-hit" hypothesis of schizophrenia. *Schizophr Bull* 2001; **27**: 457–476.
46 Verghese C, de Leon J, Josiassen RC. Problems and progress in the diagnosis and treatment of polydipsia and hyponatremia. *Schizophrenia Bull* 1996; **22**: 455–464.
47 Chong TW, Castle DJ. Layer upon layer: thermoregulation in schizophrenia. *Schizophr Res* 2004; **69**: 149–157.
48 Dzau VJ, Ingelfinger J, Pratt RE, Ellison KE. Identification of renin and angiotensinogen messenger RNA sequences in mouse and rat brains. *Hypertension* 1986; **8**: 544–548.
49 Wright JW, Harding JW. Brain angiotensin receptor subtypes in the control of physiological and behavioral responses. *Neurosci Biobehav Rev* 1994; **18**: 21–53.
50 Morimoto S, Cassell MD, Sigmund CD. The brain renin–angiotensin system in transgenic mice carrying a highly regulated human renin transgene. *Circ Res* 2002; **90**: 80–86.
51 Jenkins TA, Mendelsohn FA, Chai SY. Angiotensin-converting enzyme modulates dopamine turnover in the striatum. *J Neurochem* 1997; **68**: 1304–1311.
52 Beckmann H, Saavedra JM, Gattaz WF. Low angiotensin-converting enzyme activity (kininase II) in cerebrospinal fluid of schizophrenics. *Biol Psychiatry* 1984; **19**: 679–684.
53 Owen F, Lofthouse R, Crow TJ. Angiotensin-converting enzyme in substantia nigra of schizophrenics. *N Engl J Med* 1980; **303**: 528–529.
54 Wahlbeck K, Rimon R, Fyhrquist F. Elevated angiotensin-converting enzyme (kininase II) in the cerebrospinal fluid of neuroleptic-treated schizophrenic patients. *Schizophr Res* 1993; **9**: 77–82.
55 Wahlbeck K, Ahokas A, Nikkila H, Miettinen K, Rimon R. Cerebrospinal fluid angiotensin-converting enzyme (ACE) correlates with length of illness in schizophrenia. *Schizophr Res* 2000; **41**: 335–340.
56 Davis KL, Kahn RS, Ko G, Davidson M. Dopamine in schizophrenia: a review and reconceptualization. *Am J Psychiatry* 1991; **148**: 1474–1486.
57 Johnston CI. Biochemistry and pharmacology of the renin–angiotensin system. *Drugs* 1990; **39**: 21–31.
58 Baghai TC, Binder EB, Schule C, Salyakina D, Eser D, Lucae S *et al.* Polymorphisms in the angiotensin-converting enzyme gene are associated with unipolar depression, ACE activity and hypercortisolism. *Mol Psychiatry* 2006; **11**: 1003–1015.
59 Crescenti A, Gasso P, Mas S, Abellana R, Deulofeu R, Parellada E *et al.* Insertion/deletion polymorphism of the angiotensin-converting enzyme gene is associated with schizophrenia in a Spanish population. *Psychiatry Res* 2009; **165**: 175–180.
60 Kucukali CI, Aydin M, Ozkok E, Bilge E, Zengin A, Cakir U *et al.* Angiotensin-converting enzyme polymorphism in schizophrenia, bipolar disorders, and their first-degree relatives. *Psychiatr Genet* 2010; **20**: 14–19.
61 Hui L, Wu JQ, Ye MJ, Zheng K, He JC, Zhang X *et al.* Association of angiotensin-converting enzyme gene polymorphism with schizophrenia and depressive symptom severity in a Chinese population. *Hum Psychopharmacol* 2015; **30**: 100–107.
62 Wu Y, Wang X, Shen X, Tan Z, Yuan Y. The I/D polymorphism of angiotensin-converting enzyme gene in major depressive disorder and therapeutic outcome: a case-control study and meta-analysis. *J Affect Disord* 2012; **136**: 971–978.
63 Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. Positional cloning of the mouse obese gene and its human homologue. *Nature* 1994; **372**: 425–432.
64 Fillman SG, Cloonan N, Catts VS, Miller LC, Wong J, McCrossin T *et al.* Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol Psychiatry* 2013; **18**: 206–214.
65 Fillman SG, Weickert TW, Lenroot RK, Catts SV, Bruggemann JM, Catts VS *et al.* Elevated peripheral cytokines characterize a subgroup of people with schizophrenia displaying poor verbal fluency and reduced Broca's area volume. *Mol Psychiatry* 2016; **21**: 1090–1098.
66 Cohn TA, Remington G, Zipursky RB, Azad A, Connolly P, Wolever TM. Insulin resistance and adiponectin levels in drug-free patients with schizophrenia: A preliminary report. *Can J Psychiatry* 2006; **51**: 382–386.
67 Diniz BS, Teixeira AL, Campos AC, Miranda AS, Rocha NP, Talib LL *et al.* Reduced serum levels of adiponectin in elderly patients with major depression. *J Psychiatr Res* 2012; **46**: 1081–1085.
68 Lehto SM, Huotari A, Niskanen L, Tolmunen T, Koivumaa-Honkanen H, Honkalampi K *et al.* Serum adiponectin and resistin levels in major depressive disorder. *Acta Psychiatr Scand* 2010; **121**: 209–215.

69 Takayanagi Y, Cascella NG, Santora D, Gregory PE, Sawa A, Eaton WW. Relationships between serum leptin level and severity of positive symptoms in schizophrenia. *Neurosci Res* 2013; **77**: 97–101.
70 Opland DM, Leinninger GM, Myers MG Jr. Modulation of the mesolimbic dopamine system by leptin. *Brain Res* 2010; **1350**: 65–70.
71 James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*, Vol. 6. Springer: New York, NY, USA, 2013.
72 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012; **489**: 57–74.


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)