

ARTICLE

Open Access

Male increase in brain gene expression variability is linked to genetic risk for schizophrenia

Junfang Chen¹, Han Cao¹, Andreas Meyer-Lindenberg¹ and Emanuel Schwarz¹

Abstract

Schizophrenia shows substantial sex differences in age of onset, course, and treatment response, but the biological basis of these effects is incompletely understood. Here we show that during human development, males show a regionally specific decrease in brain expression similarity compared to females. The genes modulating this effect were significantly co-expressed with schizophrenia risk genes during prefrontal cortex brain development in the fetal period as well as during early adolescence. This suggests a genetic contribution to a mechanism through which developmental abnormalities manifest with psychosis during adolescence. It further supports sex differences in brain expression variability as a factor underlying the well-established sex differences in schizophrenia.

Introduction

Schizophrenia is a severe developmental mental illness with an incidence approximately 1.4 times higher in men compared to women¹. The disorder is substantially heritable and a large number of common and rare variants have been associated with illness risk^{2–5}. A widely accepted neurodevelopmental hypothesis posits that genetically determined alterations in early brain development interact with developmental changes during adolescence in the prefrontal cortex to lead to the manifestation of psychosis^{6,7}. Consistent with this, developmentally changing prefrontal cortex expression has been found to be linked to neuronal differentiation and maturation, as well as genetic schizophrenia risk⁸.

In men, the illness has a more severe course characterized by more pronounced negative symptoms as well as cognitive impairment^{9,10}, although evidence has been reported that substance abuse in men may confound such clinical differences¹¹. Males with schizophrenia have also, albeit inconsistently, been reported to have a lower age of onset, show more pronounced alterations of brain

morphology, and poorer response to antipsychotic medication^{9,11–13}. Genetic risk associations, as well as molecular profiles, contain sex-dependent factors^{14,15} and sex hormones are thought to play an important role for illness course^{9,16}, but again little is known about the underlying neurobiological mechanisms.

We pursued a novel strategy to explore how biological sex differences may impact on the manifestation of genetic risk and the clinical sex differences of schizophrenia. Inspired by a recent study on the human brain connectome¹⁷, we tested whether during development human brain gene expression is more variable in males than females. We hypothesized that such increased expression variability might contribute to a predisposition of males for heritable neurodevelopmental disorders. A similar hypothesis has previously been explored for HIV, where gene expression variability has been suggested as a modulator for susceptibility to infection¹⁸. Our study is further motivated by previous identification of sexual dimorphisms of brain expression^{19–21}, protein abundance²², as well as genetic and epigenetic factors modulating gene expression noise^{23,24}, supporting the possibility of links between polygenic risk and expression variance. The longitudinal exploration of variability

Correspondence: Emanuel Schwarz (emanuel.schwarz@zi-mannheim.de)

¹Medical Faculty Mannheim, Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Heidelberg University, Mannheim, Germany

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

differences is further motivated by previous identification of differential variance of transcriptional regulators during human embryonic development²⁵. Analysis of gene expression variability has also been successfully applied to identify genes and pathways implicated in several illnesses and highlighted such variability as an informative biological signal^{26,27}.

Expression variability as genetic risk mediator can capture polygenic effects beyond sex differences of expression. To investigate this, we identified genes driving brain region and age-specific variability differences between sexes and tested whether these were associated with expression of schizophrenia risk genes.

Materials and methods

Data preprocessing

To characterize brain expression throughout the human lifespan, we used data from the BrainSpan: Atlas of the Developing Human Brain (funded by ARRA Awards 1RC2MH089921-01, 1RC2MH090047-01, and 1RC2MH089929-01 and available from: <http://developinghumanbrain.org>), as well as Braincloud microarray data (GSE30272²⁸, available from the GEO database²⁹).

The primary analysis was performed on BrainSpan exon microarray data (GSE25219, preprocessed as described in ref. ³⁰) due to availability of a larger sample number. BrainSpan RNA sequencing (RNAseq) data was used for replication and Braincloud data for validation of findings. BrainSpan data comprised transcriptome-wide expression information on subjects between the 6th post-conceptual week (PCW) and 40 years of age (Table 1, Supplementary Tables 2, 8, and 9). We did not consider older subjects, as sex effects on risk are not likely to manifest beyond the typical age of onset that ranges between late adolescence and early adulthood. As performed by Willsey et al.³¹, subjects were grouped in age-bins by a windowing approach that joins three consecutive age periods into a single group.

Preprocessing of all datasets followed a similar sequence of steps (Supplementary Fig. 1). Procedures performed on all datasets comprised: RNA Integrity Number (RIN) filtering (for BrainSpan exon microarray data, all donors were removed that had more than 25% of microarray samples with $RIN < 7.5$, as in ref. ³⁰; for BrainSpan RNAseq data and Braincloud data, a more stringent filtering was performed by removing all samples with $RIN < 7.5$); removal of subjects >40 years; log₂ transformation of data; extraction of autosomal genes (without minimum expression filter); quantile normalization; surrogate variable determination; covariate adjustment; and outlier detection. This data contained the respective median values if multiple replicates per subject were present. Following a previously described pipeline²⁰, processing of RNAseq data included two additional steps: gene-level

reads per kilobase million mapped reads (RPKM) were normalized for GC content using conditional quantile normalization based on the R library *cqn*³² and all genes with less than 1 RPKM in more than 50% of male or female samples were removed. Surrogate variable analysis was performed to account for the potential effects of unobserved confounders³³. The number of surrogate variables were automatically determined using the *num.sv* function of the R package *sva*³³, using the approximation method by Leek^{33,34}. The underlying full model matrix contained gender, whose effects on expression variability should be preserved, as well as age, PMI, RIN, and brain pH (as well as an array indicator for Braincloud data). The null-model matrix contained all covariates but gender. Age was used as a covariate, to prevent artifactual correlations between genes due to their joint association with age. This is particularly important for age-bins covering a broader range of ages, where significant correlations between age and expression can be expected. The number of surrogate variables determined for BrainSpan exon microarray was 0, 2 for BrainSpan RNAseq and 0 for the Braincloud data. Covariate adjustment was performed via residualization against all covariates described above (except for gender) using linear models. Missing brain pH values were replaced by the mean of non-missing values.

Outlier detection

After preprocessing, principal component analysis was used to exclude outliers (Supplementary Fig. 2). For this, we identified separately for males and females observations that deviated more than 3 standard deviations from the mean of the respective first two principal components. This removed 7 samples in the BrainSpan exon microarray data (6 from male donors), 11 observations in the BrainSpan RNAseq data (6 from male donors), and 1 outlier (from a female donor) in the Braincloud data.

Schizophrenia risk genes

Schizophrenia risk variants, loci, and associated genes were taken from ref. ⁵ (Supplementary Table 3). Previous analyses have pursued different approaches to identify genes linked to genetic schizophrenia risk. Among these approaches is the selection of all genes or those within a certain distance from a given locus⁵, or genes affected by index variant eQTLs³⁵. For the present study, we aimed to identify a single gene per locus. This was due to the risk of introducing statistical bias from including multiple genes per locus, caused by (1) the undue influence of loci harboring a larger number of genes and (2) the gene–gene correlation of genes in close chromosomal proximity. Therefore, for loci harboring multiple genes, we here used the gene in closest chromosomal proximity to the genome-wide significant index variant. If a locus contained more than one index variant, we selected the gene

Table 1 BrainSpan exon microarray sample numbers for males and females across 11 age-bins and 4 brain regional clusters after data preprocessing

Age-bin	Regional cluster	Males				Females			
		1	2	3	4	1	2	3	4
1		12 (3)	21 (4)	11 (4)	1 (1)	20 (4)	24 (4)	12 (4)	7 (4)
2		23 (5)	31 (6)	16 (6)	5 (4)	20 (4)	24 (4)	12 (4)	7 (4)
3		23 (5)	27 (5)	14 (5)	5 (4)	33 (7)	39 (7)	20 (7)	13 (7)
4		23 (5)	27 (5)	14 (5)	8 (5)	27 (6)	31 (6)	16 (6)	12 (6)
5		15 (3)	18 (3)	9 (3)	6 (3)	27 (6)	31 (6)	16 (6)	12 (6)
6		30 (6)	36 (6)	18 (6)	12 (6)	14 (3)	16 (3)	8 (3)	6 (3)
7		24 (5)	29 (5)	12 (4)	10 (5)	5 (1)	6 (1)	3 (1)	2 (1)
8		24 (5)	28 (5)	12 (4)	10 (5)	10 (2)	9 (2)	6 (2)	4 (2)
9		19 (4)	20 (4)	8 (3)	7 (4)	15 (3)	15 (3)	8 (3)	5 (3)
10		20 (4)	21 (4)	10 (4)	7 (4)	20 (4)	21 (4)	11 (4)	7 (4)
11		36 (8)	42 (8)	19 (7)	14 (8)	33 (7)	39 (7)	19 (7)	12 (7)

1: VIC-STC, 2: PFC-MS, 3: STR-HIP-AMY, 4: MD-CBC (see Supplementary Table 1 for details). Subject numbers are shown in brackets

in closest chromosomal proximity to the most significant index variant. Chromosomal locations were determined from the R library org.Hs.eg.db., vs. 3.1.2 (genome build hg19, assembly GRCh37). Genes within the MHC region were not considered due to their significant linkage disequilibrium pattern. Two loci mapped to the genes *IMMP2L* and *TCF4*, and these were considered only once for subsequent analyses. *C10orf32*, *C12orf79*, and *VPS14C* were not annotated by the library org.Hs.eg.db. and not considered for further analysis. The final set of schizophrenia risk genes contained 100 genes, of which 97 were autosomal. Of these, 87 were part of the BrainSpan dataset (see Supplementary Table 3).

Analysis of expression similarity

First, all samples were identified for a given brain regional cluster and age-bin. Based on such data subset, we performed a three stage resampling approach separately for males and females. The objective of this resampling was to quantify the expression similarity (and its confidence interval) between subjects while accounting for the non-independence of multiple samples taken from the same donor:

1. First, we randomly selected a single sample per subject to prevent an impact of sample non-independence on results.
2. Second, we took a bootstrap sample of subjects by sampling with replacement and chose the unique set of subjects. This was performed to prevent the perfect correlation between multiply selected samples.

3. Finally, we subsampled the selected subjects, such that the same number of subjects was chosen for males and females. This was aimed at preventing an influence of unequal sample numbers on results.

Then separately for males and females, we determined the pairwise Pearson correlation coefficients between all subject pairs using expression values from all genes. The mean of these estimates was used as an estimate of expression similarity between subjects for a given regional cluster age-bin combination. Only the upper triangular matrix of a given correlation matrix was used for estimation. This entire resampling was repeated 100 times and the mean value (for confidence intervals the upper and lower 2.5% percentile) of obtained estimates used to quantify expression similarity.

The difference between males and females was then quantified as the mean difference between the point estimates of each regional cluster age-bin combination. To assess significance, the resampling procedure was repeated 1000 times. During each repetition, gender information was permuted for a given regional cluster age-bin combination, such that different samples of the same subject were always assigned the same gender. The frequency of bootstrapping point estimates at least as high as the one obtained from non-permuted data was used as empirical *P*-value and corrected for multiple comparisons according to the method of Bonferroni. To perform two-sided tests, absolute values were used for this calculation.

Identification of genes driving expression similarity differences

We anticipated that genes driving the difference of expression similarity between males and females would likely show strong differences in expression variance between sexes. For each regional cluster age-bin combination, we therefore performed the same resampling strategy as described above. For a given set of subjects (males and females separately), we then determined the standard deviation of expression for a given gene. These estimates were averaged over 100 resampling repetitions. We then determined the ratio of these averages between males and females and used the 100 genes (arbitrary cut-off) with the highest ratio as “variability genes”. To test whether these gene sets were also “variability genes” in replication (BrainSpan RNAseq data) and validation (Braincloud) data, we determined the difference of expression similarity estimates (using the resampling strategy described above) between males and females. An empirical *P*-value was then determined by comparing this estimate against those derived from random “variability genes” identified as described below (1000-fold resampling, one-sided test).

Testing associations with schizophrenia risk genes

To explore associations between variability genes and schizophrenia susceptibility genes, the co-expression between the two gene sets was determined for a given regional cluster age-bin combination, by calculating a matrix of all pairwise Pearson correlation coefficients using expression values from both gene sets. The median value of this correlation matrix was then used as a measure of co-expression. Again, these calculations were determined as part of the resampling procedure described above, with the exception of the third step (undersampling to obtain equal numbers of male and female subjects), since calculations were performed using males only.

Significance was determined using 1000 fold resampling. During each repetition and for each regional cluster age-bin combination, the low number of donors prevented meaningful permutation of gender information. Therefore, random “variability genes” were selected such that for each real variability gene, one gene with a standard deviation of expression within 5% of the original gene was randomly chosen. The resulting co-expression values were then used to form null-distributions. Empirical *P*-values were determined as the frequency of co-expression values at least as high as that observed from real data (one-sided test). Since a total of 22 sets of variability genes were tested, *P*-values were corrected for the Family Wise Error Rate according to the method of Bonferroni.

Analysis of schizophrenia specificity

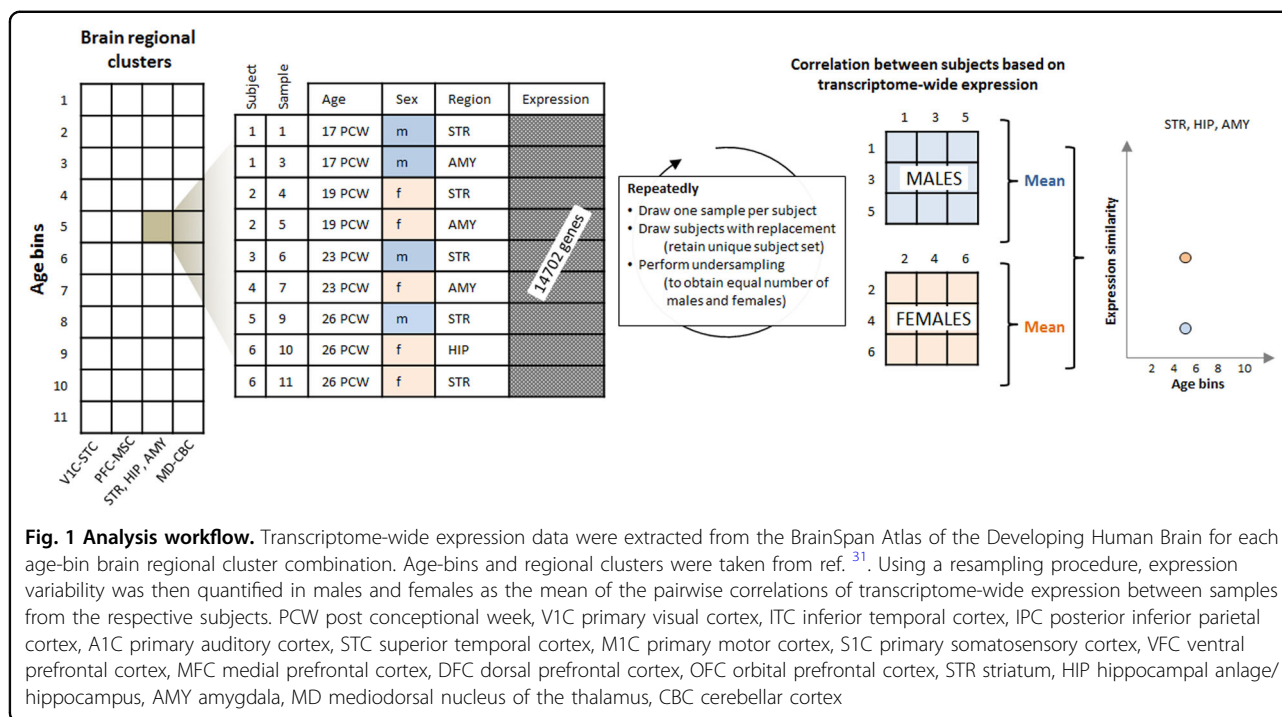
To test the specificity of the co-expression between variability genes and schizophrenia susceptibility genes, five additional analyses were performed, using different selections of “susceptibility genes”: (I) Random selection of schizophrenia susceptibility genes for a given locus (instead of based on physical proximity to the index SNP). (II) Random selection of genes from loci with comparable DNA sequence variability compared to the schizophrenia loci. For this analysis, the number of common (MAF > = 1%) variants recorded in dbSNP (GRCh37, available from <https://genome.ucsc.edu/>) was used as a proxy for DNA sequence variability. For each schizophrenia locus, a locus of the same size was selected from the same chromosome and retained if the DNA sequence variability was within 10% of the original locus. A random gene was then selected from the locus, extended by 20 kbp, using the R library *biomaRt*³⁶. (III) Random selection of genes from the same chromosome as a given schizophrenia gene, irrespective of DNA sequence variability. (IV) Selection of genes in proximity to SNPs associated with major depressive disorder (35 genes; closest gene selected to a given index SNP, as described in ref. ³⁷). (V) Selection of genes in proximity to SNPs associated with a non-psychiatric phenotype (coronary artery disease; 35 genes; random gene selected from a given susceptibility locus, as described in ref. ³⁸).

Exploratory age-windowing

To perform a “fine-mapping” of effects within a set of age-bins, we performed separate analyses for subjects within a given age-window (Supplementary Table 7). The width of the window was determined as four consecutive age entries among the recorded ages in weeks. Differences of expression similarity and co-expression with schizophrenia susceptibility genes were determined separately for each age-window as described above. Genes identified as “variability genes” of the investigated age-bins were combined and used for this analysis.

Functional analysis

To explore biological functions of genes contributing to differences of expression similarity between sexes, we used the DAVID functional annotation tool using default settings (<https://david.ncifcrf.gov/home.jsp>)³⁹. In this tool, enrichment is quantified based on a modified Fisher’s exact test. The 14,702 autosomal genes part of the BrainSpan exon microarray data were used as background for functional analysis. We retained all functional annotation clusters with at least one annotation term passing the False Discovery Rate corrected *P*-value threshold of 0.05.



Code availability

Code is available from the corresponding author upon request.

Results

Expression similarity differences in BrainSpan exon microarray data

The filtered dataset contained autosomal, transcriptome-wide expression data on healthy subjects between the 6th PCW and 40 years of age¹⁹ (42 donors, 23 males, 14,702 autosomal genes; Fig. 1). We tested whether gender was confounded by ethnicity, but found no association ($P = 0.77$, Chi-squared test). Subjects were binned into 11 age groups and the 16 brain areas were aggregated into 4 regional clusters with similar expression values (Supplementary Tables 1 and 2, regional clustering was taken from ref. ³¹ and based on hierarchical clustering of fetal transcriptome profiles; for abbreviations, see Fig. 1): (1) the V1C-STC cluster; (2) the prefrontal and primary motor-somatosensory cortex or PFC-MSC cluster; (3) the STR-HIP-AMY cluster; and (4) the MD-CBC cluster.

Figure 2a shows that despite substantial variability, males had significantly lower expression similarity compared to females in three of the four brain regional clusters ($P_{V1C-STC} < 0.004$, $P_{PFC-MSC} < 0.004$, $P_{STR-HIP-AMY} = 0.003$, $P_{MD-CBC} = 0.080$; FWER corrected). Due to the more pronounced differences in the regional clusters V1C-STC and PFC-MSC, subsequent analyses focused on these areas. Figure 2a further shows that in females, expression similarity tended to decrease across

developmental time points, suggesting that inter-subject similarity was lower in adulthood compared to younger age. We aimed to explore whether sex differences in expression similarity were associated with genetic schizophrenia risk, to pinpoint a potential biological mechanism for the well-known sex differences of the disorder.

Identification of genes driving sex differences in expression similarity

For each age-bin regional cluster combination we identified the 100 “variability genes” with the greatest ratio (male divided by female) of standard deviations of expression (see Supplementary Fig. 3 and Supplementary Dataset for a list of all “variability genes”). Figure 2b shows that expression similarity determined from these genes differed strongly between sexes.

Co-expression between variability and schizophrenia susceptibility genes

Next, we investigated potential relationships between these variability genes and genes harbored by the 108 well-established schizophrenia susceptibility loci⁵. This analysis was performed in males, since the lack of variance in female expression levels would prevent meaningful association analyses. Across 22 sets of variability genes (11 age-bins in the 2 regional clusters V1C-STC and PFC-MSC), we found that variability genes derived from both clusters were significantly co-expressed with schizophrenia susceptibility genes in age-bins 8 (4 months–4 years, $\rho_{V1C-STC} = 0.05$,

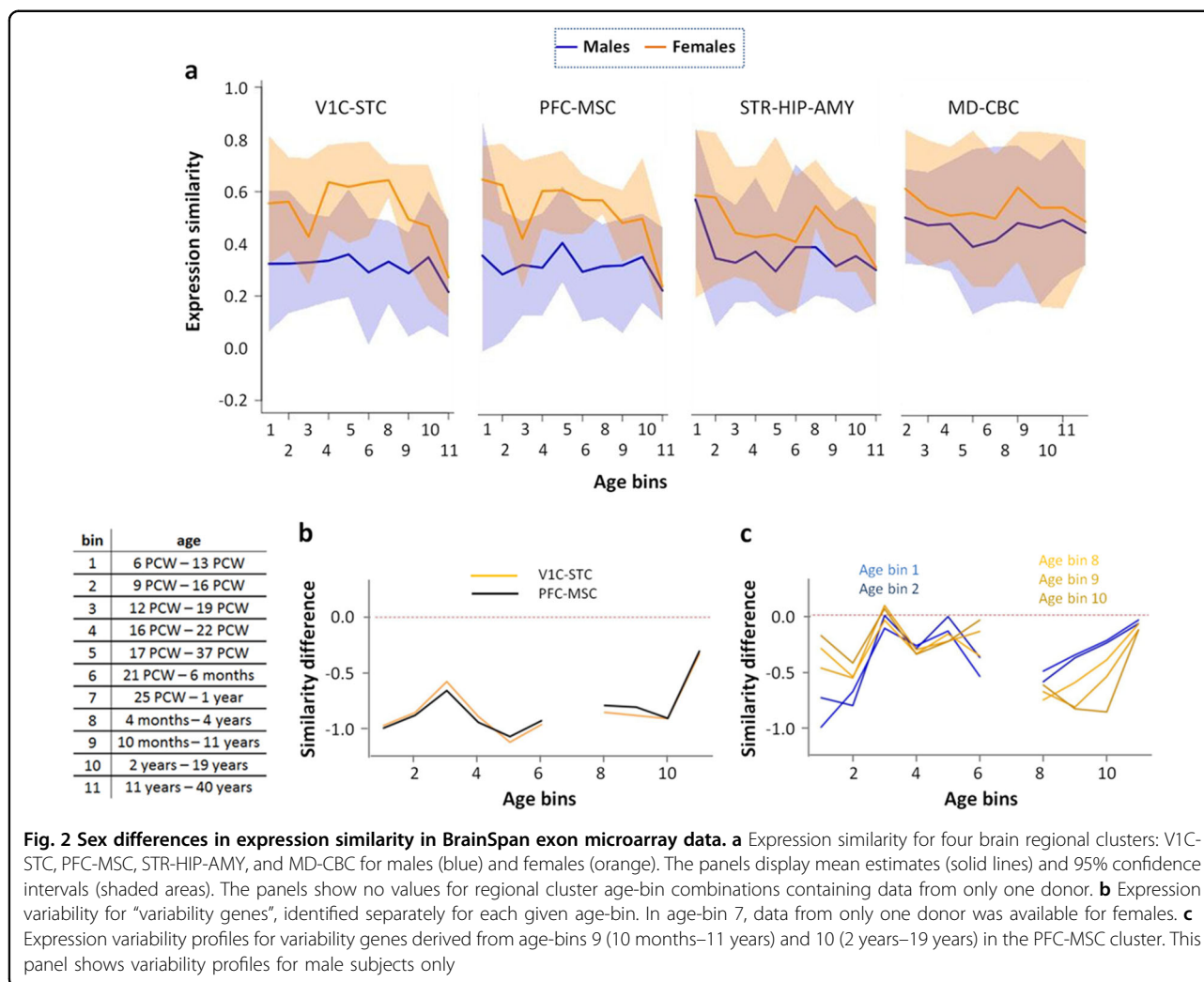


Fig. 2 Sex differences in expression similarity in BrainSpan exon microarray data. **a** Expression similarity for four brain regional clusters: VIC-STC, PFC-MSC, STR-HIP-AMY, and MD-CBC for males (blue) and females (orange). The panels display mean estimates (solid lines) and 95% confidence intervals (shaded areas). The panels show no values for regional cluster age-bin combinations containing data from only one donor. **b** Expression variability for “variability genes”, identified separately for each given age-bin. In age-bin 7, data from only one donor was available for females. **c** Expression variability profiles for variability genes derived from age-bins 9 (10 months–11 years) and 10 (2 years–19 years) in the PFC-MSC cluster. This panel shows variability profiles for male subjects only

$\rho_{\text{PFC-MSC}} = 0.05$), 9 (10 months–11 years, $\rho_{\text{VIC-STC}} = 0.10$, $\rho_{\text{PFC-MSC}} = 0.12$), and 10 (2 years–19 years, $\rho_{\text{VIC-STC}} = 0.07$, $\rho_{\text{PFC-MSC}} = 0.13$; all $P_{\text{FWER}} < 0.022$, Fig. 3a, b). Significant co-expression was additionally observed for the PFC-MSC in age-bin 1 (6 PCW–13 PCW, $\rho = 0.11$) and 2 (9 PCW–16 PCW, $\rho = 0.08$; all $P_{\text{FWER}} < 0.022$).

Age-bin specificity and pathway analysis

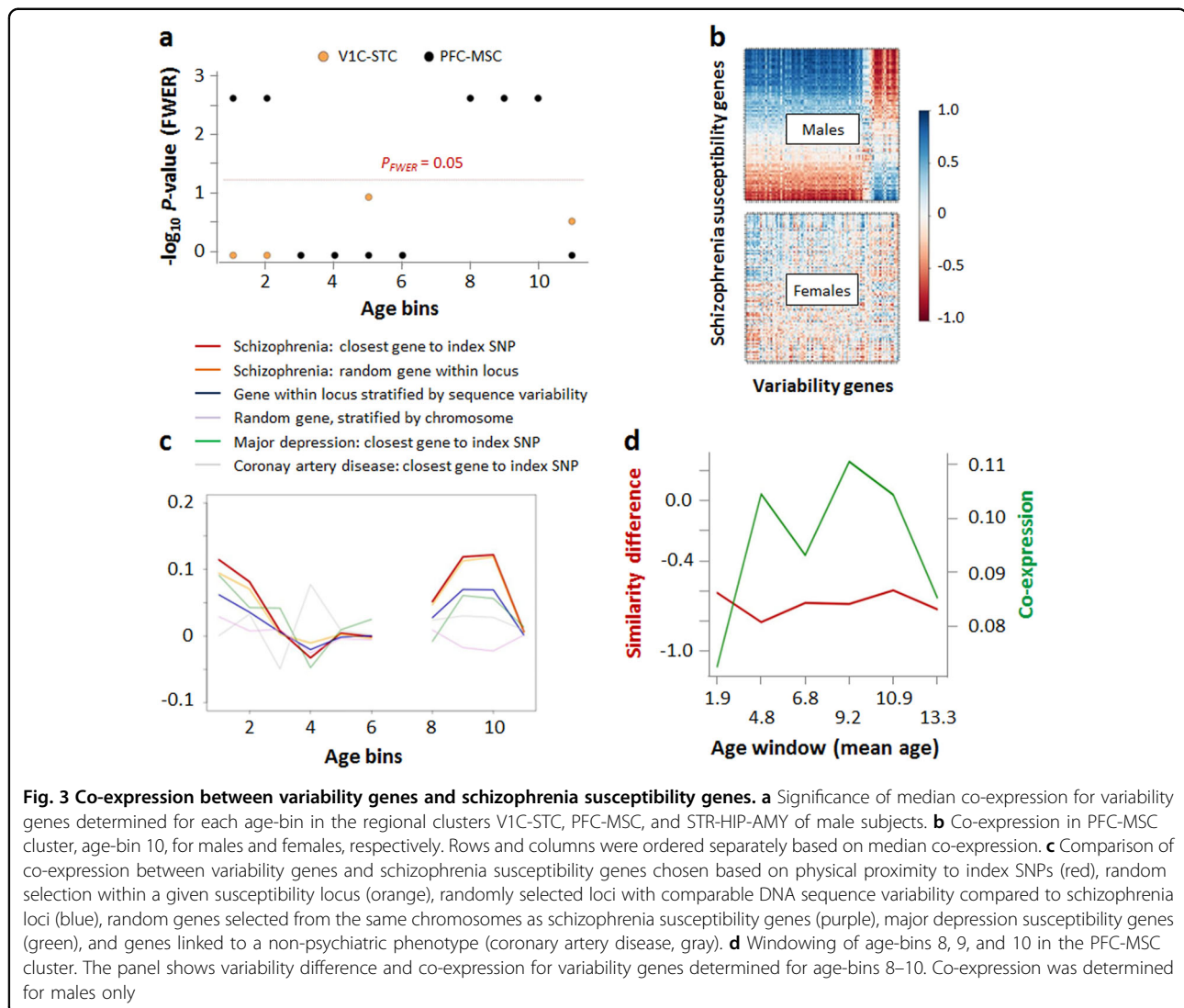
Next, we explored whether differences in expression similarity were age-bin specific. Figure 2c shows that PFC-MSC variability genes of age-bin blocks 1–2 and 8–9–10 were also associated, albeit to a lesser extent, with decreased male expression similarity in the respectively other age-bin blocks.

In this brain regional cluster, the 257 genes of age-bins 8–10 were significantly linked to synaptic processes and (calcium-)ion signaling (Supplementary Table 4). Notably, the 138 variability genes from age-bins 1 and 2 in the PFC-MSC cluster were associated with similar ontological categories, including “post-synaptic

membrane” and “synapse” (Supplementary Table 5). Interestingly, the genes from age-bins 1–2 and age-bins 8–10 showed only a minimal overlap (8 genes shared). These ontological associations showed regional specificity for the PFC-MSC cluster, as the VIC-STC variability genes (age-bins 8–10) that also showed significant co-expression with susceptibility genes were not associated with similar ontological categories (Supplementary Table 6). Furthermore, the ontological overlap between age-bins 1–2 and age-bins 8–10 in the PFC-MSC cluster is consistent with the correlation of the male expression similarity profiles (Fig. 2c).

Schizophrenia specificity

To explore the specificity of co-expression results for schizophrenia, analysis was repeated using (I) schizophrenia susceptibility genes randomly selected for a given locus (instead of based on physical proximity to the index SNP), (II) genes randomly selected from loci with comparable DNA sequence variability compared to the



schizophrenia loci, (III) genes randomly selected from the same chromosome as a given schizophrenia gene, irrespective of DNA sequence variability, (IV) genes in proximity to SNPs associated with major depressive disorder, (V) genes in proximity to SNPs associated with a non-psychiatric phenotype (coronary artery disease). Figure 3c shows that random and proximity-based selection of genes from schizophrenia loci yielded similar results. Despite a similar co-expression profile across age-bins, schizophrenia gene co-expression showed specificity against DNA sequence variability-stratified gene selection in age-bins 8 and 9 ($P=0.05$) and a trend toward specificity in age-bins 2 and 10 ($P=0.06$). Randomly selected genes (procedure III) showed substantially lower mean co-expression, leading to specificity of schizophrenia results (age-bins 8–10, $P \leq 0.05$; age-bin 2, $P=0.06$). For both random selection procedures,

schizophrenia specificity could not be observed in age-bin 1 ($P=0.12$ and $P=0.11$, for procedures II and III, respectively). Genes in the proximity of SNPs linked to major depression or grip strength led to lower co-expression values in age-bins 2 and 8–10; in age-bin 1, major depression genes showed higher co-expression than the schizophrenia genes.

Age-windowing

Finally, since age-bins 8–10 covered a broad age range (4 months–19 years), we performed an exploratory “fine-mapping” of PFC-MS effects using an age-windowing approach. While based on small sample numbers, this analysis suggested that co-expression had a broad plateau from a mean age of 4.8–10.9 years (Fig. 3d). Differences in expression similarity between sexes were consistent across all age windows (Fig. 3d).

Replication in BrainSpan RNAseq data

Preprocessed BrainSpan RNAseq data comprised expression information on 11,514 autosomal genes in 400 samples (37 subjects, 20 males). The transcriptome-wide expression similarity showed similar profiles as observed for exon microarray data (Supplementary Fig. 4a). Similarly, the variability genes identified from exon microarray data were also variability genes in RNAseq data (Supplementary Fig. 4b, $P < 0.001$). These genes were significantly correlated with schizophrenia susceptibility genes in the PFC-MSC regional cluster for age-bins 2 ($\rho = 0.01$, $P < 0.001$), 9 ($\rho = 0.05$, $P < 0.001$), and 10 ($\rho = 0.03$, $P < 0.001$), validating exon microarray observations. For the V1C-STC cluster, we found significant associations for age-bins 3 ($\rho = 0.03$, $P < 0.001$), and a trend toward nominal significance in age-bins 8 ($P = 0.07$) and 10 ($P = 0.05$).

Validation in Braincloud data

Filtered Braincloud data contained dorsolateral prefrontal cortex expression information on 14,773 autosomal genes from 112 subjects (75 males). In covariate-corrected data, expression similarity is dependent on expression variance. Therefore, we compared the standard deviation of expression across all genes overlapping with BrainSpan exon microarray data. We found these estimates to be strongly correlated across datasets ($\rho = 0.40$, $P < 2.2 \times 10^{-16}$, Spearman correlation), suggesting that preprocessing resulted in high cross-dataset comparability. Since the Braincloud data contained no subjects in age groups 1 and 2 (i.e., age-bin 1 only consisted of subjects in age group 3), age-bin 1 was not used for further analysis. Assessment of expression similarity differences using BrainSpan exon microarray PFC-MSC variability genes validated the decreased similarity in males ($P < 0.001$, Supplementary Fig. 5), which was less pronounced in Braincloud data and driven by genes from age-bins 8 and 9. Consistent with BrainSpan results, co-expression with schizophrenia susceptibility genes was significant in age-bin 2 ($\rho = 0.03$, $P < 0.001$), age-bin 9 ($\rho = 0.07$, $P < 0.001$), and age-bin 10 ($\rho = 0.03$, $P < 0.001$) and showed a trend toward significance in age-bin 8 ($\rho = 0.01$, $P = 0.08$).

Discussion

The present results demonstrate that the similarity of gene expression profiles in males shows a brain region-specific decrease compared to females. Some of the genes driving this effect were co-expressed with schizophrenia susceptibility genes, in a regionally specific and age-dependent manner. Importantly, co-expression was found in the brain regional cluster encompassing the prefrontal cortex during fetal brain development, confirming a core prediction of the neurodevelopmental hypothesis of

schizophrenia⁶. Additionally, and again as predicted by this hypothesis, significant co-expression was further found during adolescence. Similar differences of expression similarity were found in RNAseq data acquired on a subset of the same samples. In this dataset, we further replicated associations between variability and susceptibility genes in data from adolescent donors, but found no associations during the fetal period. Expression similarity differences were further validated in the independent Braincloud data and significant co-expression was found in samples from fetal, as well as adolescent donors.

Co-expression did not depend on how genes were selected from a given susceptibility locus and exceeded that observed for major depression (in age-bin 1 by a small margin) and coronary artery disease in the early fetal phase, as well as during adolescence. We observed that genes selected from randomly chosen loci stratified for DNA sequence variability showed a broadly similar, although less pronounced, co-expression trend compared to schizophrenia genes. In contrast, genes selected randomly without consideration of DNA sequence variability were not co-expressed with variability genes, on average. This may suggest that sequence variability associated with schizophrenia susceptibility loci impacted on diversification of gene expression and the sex differences observed in the present study.

Genes from the fetal and adolescent periods were involved in synaptic processes, which have been implicated in schizophrenia by a range of genetic, histopathological, neuroimaging, pharmacological, and neurotransmitter studies^{40–44}. They are affected by genetic and environmental risk in particular during early life, leading to subsequent impairments in synaptic plasticity and connectivity⁴⁵. The lack of overlap between variability genes from the fetal period and adolescence may hint at biologically divergent risk processes that converge on the same synaptic pathways.

The main limitation of the present study is sample size. The primary analysis of the BrainSpan data reported that brain region and age are stronger modulators of gene expression compared to sex or inter-individual variation¹⁹. Therefore, the present study focused on analyses that are stratified by regional clusters and age-bins, with significant impact on sample numbers available for a given analysis. In the BrainSpan dataset, data from multiple brain regions was available for most donors. We performed a donor-wise bootstrapping procedure during all resampling analyses, to account for the non-independence of the samples. This procedure further accounted for potential effects arising from differences in donor numbers between sexes, further reducing the effective sample size. The low donor number per regional cluster age-bin combination prevented meaningful permutation of gender. Therefore, random “variability genes”

were created by randomly sampling genes, stratified by expression variance. This may have led to bias, due to the potential correlation among the actual variability genes that is not captured by the procedure employed here. The low sample number in all three investigated datasets also limits the power to identify and validate significant associations, including expression similarity differences and co-expression between variability and susceptibility genes. This may have contributed to the partial non-replication of findings across datasets.

Another limitation is that we selected a single susceptibility gene per locus to prevent statistical bias, but this selection may not accurately reflect genetic schizophrenia risk. By comparison, other studies have previously selected susceptibility genes by extracting all genes within a given locus⁵ or by focusing on effectors of index variant eQTLs³⁵. Another interesting aspect is that the present findings may relate to underlying, variable phenotypes, such as personality traits and comorbid psychiatric conditions. Furthermore, we aimed to account for the effects of known and unknown confounders during all analyses, but this may not have comprehensively captured experimental artefacts that may have influenced between-subject or gene–gene correlations. Finally, we did not use genetic association data to correct for potential subject relatedness or population structure, due to data availability and sample size limitations.

In conclusion, this study indicates sex-specific genetic mechanisms operating during fetal brain development linked to the variability of prefrontal brain gene expression during adolescence, as predicted by the neurodevelopmental hypothesis of schizophrenia. These effects may contribute to the well-established clinical sex differences of schizophrenia and underlying gene sets may be valuable for biologically stratified exploration of the illness's etiology.

Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (DFG), SCHW 1768/1-1. A.M.-L. was supported by the Deutsche Forschungsgemeinschaft (DFG) (Collaborative Research Center SFB 636, subproject B7); the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement no. 602450 (IMAGEMEND, IMAGING GENetics for MENTal Disorders); the German Federal Ministry of Education and Research (BMBF) through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders) under the auspices of the e:Med Programme (BMBF Grant 01ZX1314G); and the Innovative Medicines Initiative Joint Undertaking (IMI) under Grant Agreements 115300 (European Autism Interventions—A Multicentre Study for Developing New Medications) and 602805 (European Union-Aggressotype). We would like to thank Dr. Dietmar Lippold for his suggestions regarding data analysis and preparation of this manuscript.

Conflict of interest

A.M.-L. has received consultant fees from Blueprint Partnership, Boehringer Ingelheim, Daimler and Benz Stiftung, Elsevier, F. Hoffmann-La Roche, ICARE Schizophrenia, K. G. Jebsen Foundation, L.E.K Consulting, Lundbeck International Foundation (LINF), R. Adamczak, Roche Pharma, Science Foundation, Synapsis Foundation—Alzheimer Research Switzerland, System Analytics, and has received lectures including travel fees from Boehringer Ingelheim, Fama Public Relations, Institut d'investigacions Biomèdiques August

Pi i Sunyer (IDIBAPS), Janssen-Cilag, Klinikum Christophsbad, Göppingen, Lilly Deutschland, Luzerner Psychiatrie, LVR Klinikum Düsseldorf, LWL PsychiatrieVerbund Westfalen-Lippe, Otsuka Pharmaceuticals, Reunions i Ciencia S. L., Spanish Society of Psychiatry, Südwestrundfunk Fernsehen, Stern TV, and Vitos Klinikum Kurhessen. The remaining authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41398-018-0200-0>).

Received: 28 May 2018 Accepted: 8 June 2018

Published online: 01 August 2018

References

- Aleman, A., Kahn, R. S. & Selten, J. P. Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Arch. Gen. Psychiatry* **60**, 565–571 (2003).
- Sullivan, P. F., Daly, M. J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat. Rev. Genet.* **13**, 537–551 (2012).
- International Schizophrenia, C. et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
- Ripke, S. et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).
- Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
- Weinberger, D. R. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44**, 660–669 (1987).
- Birbaumer, R. & Weinberger, D. R. Genetic insights into the neurodevelopmental origins of schizophrenia. *Nat. Rev. Neurosci.* **18**, 727–740 (2017).
- Jaffe, A. E. et al. Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat. Neurosci.* **18**, 154–161 (2015).
- Leung, A. & Chue, P. Sex differences in schizophrenia, a review of the literature. *Acta Psychiatr. Scand. Suppl.* **401**, 3–38 (2000).
- Maric, N., Krabbendam, L., Vollebbergh, W., de Graaf, R. & van Os, J. Sex differences in symptoms of psychosis in a non-selected, general population sample. *Schizophr. Res.* **63**, 89–95 (2003).
- Abel, K. M., Drake, R. & Goldstein, J. M. Sex differences in schizophrenia. *Int. Rev. Psychiatry* **22**, 417–428 (2010).
- Morgan, V. A., Castle, D. J. & Jablensky, A. V. Do women express and experience psychosis differently from men? Epidemiological evidence from the Australian National Study of Low Prevalence (Psychotic) disorders. *Aust. N. Z. J. Psychiatry* **42**, 74–82 (2008).
- Pinals, D. A., Malhotra, A. K., Missar, C. D., Pickar, D. & Breier, A. Lack of gender differences in neuroleptic response in patients with schizophrenia. *Schizophr. Res.* **22**, 215–222 (1996).
- Goldstein, J. M., Cherkerzian, S., Tsuang, M. T. & Petryshen, T. L. Sex differences in the genetic risk for schizophrenia: history of the evidence for sex-specific and sex-dependent effects. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 698–710 (2013).
- Ramsey, J. M. et al. Distinct molecular phenotypes in male and female schizophrenia patients. *PLoS ONE* **8**, e78729 (2013).
- Markham, J. A. Sex steroids and schizophrenia. *Rev. Endocr. Metab. Disord.* **13**, 187–207 (2012).
- Kaufmann, T. et al. Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nat. Neurosci.* **20**, 513–515 (2017).
- Li, J., Liu, Y., Kim, T., Min, R. & Zhang, Z. Gene expression variability within and between human populations and implications toward disease susceptibility. *PLoS Comput. Biol.* **6**, e1000910 (2010).
- Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).

20. Werling, D. M., Parikshak, N. N. & Geschwind, D. H. Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat. Commun.* **7**, 10717 (2016).
21. Trabzuni, D. et al. Widespread sex differences in gene expression and splicing in the adult human brain. *Nat. Commun.* **4**, 2771 (2013).
22. Raser, J. M. & O'Shea, E. K. Noise in gene expression: origins, consequences, and control. *Science* **309**, 2010–2013 (2005).
23. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811–1814 (2004).
24. Alemu, E. Y., Carl, J. W. Jr, Corrada Bravo, H. & Hannenhalli, S. Determinants of expression variability. *Nucleic Acids Res.* **42**, 3503–3514 (2014).
25. Hasegawa, Y. et al. Variability of gene expression identifies transcriptional regulators of early human embryonic development. *PLoS Genet.* **11**, e1005428 (2015).
26. Ran, D. & Daye, Z. J. Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq. *Nucleic Acids Res.* **45**, e127 (2017).
27. Ho, J. W., Stefani, M., dos Remedios, C. G. & Charleston, M. A. Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* **24**, i390–i398 (2008).
28. Colantuoni, C. et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011).
29. Edgar, R., Domrachev, M. & Lash, A. E. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
30. Goyal, M. S., Hawrylycz, M., Miller, J. A., Snyder, A. Z. & Raichle, M. E. Aerobic glycolysis in the human brain is associated with development and neonatal gene expression. *Cell Metab.* **19**, 49–57 (2014).
31. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
32. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
33. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
34. Leek, J. T. Asymptotic conditional singular value decomposition for high-dimensional genomic data. *Biometrics* **67**, 344–352 (2011).
35. Gamazon, E. R. et al. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
36. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
37. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
38. Schunkert, H. et al. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338 (2011).
39. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
40. Network, Pathway Analysis Subgroup of Psychiatric Genomics C. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).
41. Schwarz, E., Izmailov, R., Lio, P. & Meyer-Lindenberg, A. Protein interaction networks link schizophrenia risk loci to synaptic function. *Schizophr. Bull.* **42**, 1334–1342 (2016).
42. Tsai, G. & Coyle, J. T. Glutamatergic mechanisms in schizophrenia. *Annu. Rev. Pharmacol. Toxicol.* **42**, 165–179 (2002).
43. Glausier, J. R. & Lewis, D. A. Dendritic spine pathology in schizophrenia. *Neuroscience* **251**, 90–107 (2013).
44. McGlashan, T. H. & Hoffman, R. E. Schizophrenia as a disorder of developmentally reduced synaptic connectivity. *Arch. Gen. Psychiatry* **57**, 637–648 (2000).
45. Lewis, D. A. & Levitt, P. Schizophrenia as a disorder of neurodevelopment. *Annu. Rev. Neurosci.* **25**, 409–432 (2002).