Gene expression changes associated with trajectories of psychopathology in a longitudinal cohort of children and adolescents

Vanessa Kiyomi Ota1,2,3,4, Marcos Leite Santoro1,2,3,4, Leticia Maria Spindola1,2,3, Pedro Mario Pan1,2,3, Andressa Simabucuro1,3,4, Gabriela Xavier1,3,4, Tamiris Vieira-Fonseca1,2,4, Evelin Aline Zanardo5, Felipe Rodolfo Camargo dos Santos6,7, Julia Luiza Schäfer8, Leslie Domenici Kulikowski5, Pedro A. F. Galante6, Paula Fontes Asprino6, Elisa Brietzke2,9,10, Rodrigo Grassi-Oliveira11, Luis Augusto Rohde3,12, Euripides Constantino Miguel3,13, Ary Gadelha2,13, Jair Jesus Mari2,3, Rodrigo Affonseca Bressan1,2,3, Giovanni Abrahao Salum3,8 and Sintia Iole Belangero1,2,3,4

Abstract
We aimed to identify blood gene expression patterns associated to psychopathological trajectories retrieved from a large community, focusing on the emergence and remission of general psychiatric symptoms. Hundred and three individuals from the Brazilian High-Risk Cohort Study (BHRCS) for mental disorders were classified in four groups according to Child Behavior Checklist (CBCL) total score at the baseline (w0) and after 3 years (w1): low–high (L–H) (N = 27), high–low (H–L) (N = 12), high–high (H–H) (N = 34) and low–low (L–L) groups (N = 30). Blood gene expression profile was measured using Illumina HT-12 Beadchips, and paired analyses comparing w0 and w1 were performed for each group. Results: 98 transcripts were differentially expressed comparing w0 and w1 in the L–H, 33 in the H–L, 177 in the H–H and 273 in the L–L. Of these, 66 transcripts were differentially expressed exclusively in the L–H; and 6 only in the H–L. Cross-Lagged Panel Models analyses revealed that RPRD2 gene expression at w1 might be influenced by the CBCL score at w0. Moreover, COX5B, SEC62, and NDUFA2 were validated with another technique and were also differently regulated in postmortem brain of subjects with mental disorders, indicating that they might be important not only to specific disorders, but also to general psychopathology and symptoms trajectories. Whereas genes related to metabolic pathways seem to be associated with the emergence of psychiatric symptoms, mitochondrial inner membrane genes might be important over the course of normal development. These results suggest that changes in gene expression can be detected in blood in different psychopathological trajectories.

Introduction
Mental disorders lack of validity and poor diagnosis stability is a major concern, especially in samples from young patients. Children and adolescents with mental disorders have a higher risk to present psychiatric problems during adulthood. Moreover, there are several transitions of symptomatic presentations over development1, and investigating these trajectories would help to develop prevention and early intervention policies for mental disorders.
Comorbidity levels are high among mental disorders2 and this may reflect the shared underlying etiologic factors, such as genetic factors. The heritability estimates of mental disorders vary from 40 to 80%3,4, and family
studies suggest strong genetic correlations between several groups of these illness\textsuperscript{2,6}. Recent studies have also shown that mental disorders share molecular mechanisms\textsuperscript{7}, leading to a high genetic correlation among them. The most recent study from the Cross-Disorder Group of the Psychiatric Genomics Consortium identified 109 pleiotropic loci affecting multiple of them and these genes play major roles in neurodevelopmental processes, among others\textsuperscript{8}. Another study showed that polygenic risk scores calculated for major depressive disorder (MDD), schizophrenia and attention deficit/hyperactivity disorder (ADHD) influence developmental trajectories of depression in youth, indicating that genetic liabilities to different mental disorders may affect developmental trajectories\textsuperscript{9}.

Notably, overlapping patterns of gene expression among mental disorders were also observed. Investigating whole blood gene expression across adult and childhood ADHD, autism spectrum disorders (ASD), MDD, and healthy controls, De Jong et al.\textsuperscript{10} identified two immune-related gene co-expression modules inversely correlated with MDD and adult ADHD disease status and one module positively correlated with both MDD and childhood ADHD status\textsuperscript{10}. In another study, Gandal et al.\textsuperscript{11} analyzed published gene-expression microarray studies of the cerebral cortex across five major neuropsychiatric disorders: ASD, schizophrenia, bipolar disorder (BD), MDD and alcoholism. Modules enriched for glial cell differentiation, fatty-acid metabolism, neuronal or mitochondrial pathways were shared across ASD, schizophrenia, and bipolar disorder\textsuperscript{11}. These overlapping patterns of gene expression indicate the importance of investigating general psychiatric symptoms instead of specific mental disorders. We hypothesize that some genes have their expression affected by the course of general psychiatric symptoms. However, a comprehensive effort to investigate the association of gene expression and general psychiatric symptoms over the course of development has not been performed. To contribute to this issue, here we identify blood gene expression patterns associated to trajectories of psychopathology retrieved from a large community cohort, focusing on the emergence and remission of general psychiatric symptoms. We leverage our study by carefully analyzing selected subjects from the Brazilian High-Risk Cohort Study (BHRCS), a large community study with rich genotypic and phenotypic data. Considering that genetic factors and symptoms on mental disorders transcend diagnostic boundaries, investigating general psychiatric symptoms may help to obtain a more comprehensive picture. Moreover, analyzing these symptoms and gene expression patterns in a longitudinal design instead of a cross-sectional study might reveal clues about why overall symptoms emerge and disappear over the course of development.

Materials and methods

Study procedures and participant selection

We selected a subsample from a large prospective community school-based study in Brazil (\(n = 2512\)), the Brazilian High-Risk Cohort Study (BHRCS) for mental disorders. Briefly, the cohort included families recruited in two large Brazilian cities, Sao Paulo and Porto Alegre, and was assessed in two time points: wave 0—w0 and after three years of follow-up—w1. We collected biological data for a subsample of 621 individuals. Of these, 319 blood samples were available for both w0 and w1. After classifying the subjects into groups according to psychopathology, as described below, and selecting only samples from Sao Paulo, our final study sample was composed of 103 subjects. The Research Ethics Committee approved the research protocol. Parents provided written informed consent before the inclusion and children also provided written informed consent. The cohort characteristics and study design are detailed elsewhere\textsuperscript{12}.

Dimensional psychopathology was assessed using the Child Behavior Checklist (CBCL)\textsuperscript{13}, which was administered to the children’s biological parents. The CBCL is a broadly used inventory that provides parent-report information on a wide array of behavioral problems in youth, composed by 120 items rated as not true (0), somewhat or sometimes true (1), or very true or often true (2). CBCL total score was used to measure general psychiatric symptoms in children, considering that it comprises Internalizing behavior problems, which include withdrawn, somatic complains and anxious/depressed, and externalizing behavior problems, that correspond to delinquent behavior and aggressive behavior. Blood samples were also collected in the same day.

For the present study, we classified the participants into four groups according to total CBCL scores: low–high (L–H): those that had an increase in psychopathology at the end of the three years (CBCL at \(w0 \leq 30\) and CBCL at \(w1 > 30\)); high–low (H–L): those with a decrease in psychopathology at \(w1\) moment (CBCL at \(w0 > 30\) and CBCL at \(w1 \leq 30\)); low–low (L–L) and high–high (H–H): individuals that maintained low (CBCL at \(w0\) and \(w1 \leq 30\)) or high (CBCL at \(w0\) and \(w1 > 30\)) general psychopathology, respectively, in both waves. The CBCL threshold (30) was chosen based on a receiver operating characteristic (ROC) curve analysis (\(n = 2512\)) using CBCL score as a predictor of any mental disorder using the Development and Well Being Behavior Assessment (DAWBA), rated by trained psychiatrists. A CBCL threshold of 30 was able to predict any mental disorder based on DAWBA with a sensitivity of 75.6%, a specificity of 73.7% (Youden’s \(J = 0.019\)).

In addition, to fulfill the criteria, the L–H group should present an increase of at least 15 points (correspondent to 0.5 SD) in total CBCL score (\(\Delta\text{CBCL} = \text{CBCL}_{w1} - \text{CBCL}_{w0}\); \(\Delta\text{CBCL} \geq 15\)). Similarly, the H–L group should
include those a decrease of at least 15 points in CBCL score (ΔCBCL ≤ −15). On the other hand, L–L group and H–H group, should have a variation of −15 < ΔCBCL < 15. These criteria were also used in a previous study14.

Using these criteria, we selected those with RNA samples available at both waves (w0 and w1) and from Sao Paulo, to avoid site effects: 27 L–H, 12 H–L, 34 H–H subjects and 30 L–L, totaling 103 subjects.

Genetic analyses

RNA preparation

Blood was collected into PAXgene RNA tubes (BD) and RNA was isolated using PAXgene blood RNA kit (Qiagen). Quality and quantity were assessed using NanoDrop™ 1000 Spectrophotometer (ThermoFisher Scientific) and Qubit 2.0 fluorometer with Qubit RNA BR Assay Kit (ThermoFisher Scientific), respectively. Samples presented 260/280 > 2.0 and mean concentration of 161.67 ng/µL (min = 21, max = 422, SD = 70.58). DNA contamination was verified in 1.5% agarose gel and DNA integrity was assessed using Agilent 2100 Bioanalyzer and RNA 6000 Nano Kit (Agilent). Samples presented mean RNA Integrity Number (RIN) of 8.65 (min = 6.7, max = 9.8, SD = 0.62).

Gene expression analyses

A total of 200 ng of RNA was used with Illumina® Total Prep™ RNA Amplification Kit (Life Technologies) to synthesize cRNA, which was hybrized to Human HT-12 v4 Expression BeadChips. Paired samples (w0 and w1) were included in the same chip, but groups were randomly allocated into each chip. The investigator was blinded to the group allocation during the experiment. BeadChips were scanned using the Illumina iScan System (San Diego, California, USA), and all the experiments were performed according to the manufacturer’s protocol.

Initial quality control and processing probes and samples

The results were downloaded from the iScan and preanalyzed using the GenomeStudio software. We checked the average signal and the number of detected genes for each sample.

Data was imported to R (https://www.r-project.org/) and quality control was performed using lumi package15, ending up with 6,322 probes with high-quality data available for further analyses. Then, we performed a background correction using the Maximum Likelihood Estimation (MLE) of the Model-Based Background Correction R package16. To ensure that the different BeadChips are comparable among each other, we used a robust spline normalization (RSN), which combines the features of quantile and loess normalization and is designed to normalize the variance-stabilized data. Finally, we identified the potential batch effects and corrected for the RIN, the input cRNA concentration and the barcode of each chip. All microarray data will be available once the article has been published at https://www.ebi.ac.uk/arrayexpress/.

Differentially Expressed Genes (DEGs) To identify differentially expressed genes (DEGs) we used the Linear Models for Microarray Data (Limma) R Package17. It estimates the fold changes and standard errors by fitting a linear model for each gene and then, applies empirical Bayes smoothing to the standard errors. We have performed both within-subject and between-subject comparisons. To perform the within-subject comparison, it was estimated for each gene the spatial correlation between the samples from W0 and W1 using the residual maximum likelihood (REML). Then, the between-subject comparisons were performed among the four groups using generalized least squares. We considered as significant those genes with a False Discovery Rate (FDR) lower than 0.05.

DEGs were validated by next generation sequencing using TruSeq® Targeted RNA Expression protocol. A total of 100 ng of RNA were converted to cDNA using ProtoScript II Reverse Transcriptase (New England Biolabs) and libraries were prepared using TruSeq Targeted RNA Expression kit and TruSeq® Targeted RNA custom oligonucleotide pool (Illumina) (Supplementary Table S1). For 21 of the 25 L-H subjects that were included in the microarrays analyses after QC, and 7 of the 11 H-L subjects, libraries were pooled and sequenced on the NextSeq 500 instrument (Illumina) using MidOutput v2 kit (150 cycles). The sequencing runs consisted of 100 single-end sequencing cycles. The average % Q30 scores for individual sequencing runs ranged from 89.5–92.0%, with % PF (passing filter) ranging from 82.9–88.92 %. We used the BaseSpace® TruSeq® Targeted RNA v1.0.1 app (Illumina) to analyze the raw data, including alignment to the provided manifest file using a banded Smith Waterman algorithm. This Targeted RNA app includes the following module versions: Isis (Analysis Software): 2.5.57.4.TREx, SAMtools: 0.1.19-isis-1.0.3, Scipy: 0.14.0, Pandas: 0.14.1. The Targeted RNA app generated a target hits file, which displays total reads per amplicon per sample, and this file was imported to R and analyzed using edgeR package18.

Enrichment analysis

For each probe, we selected only those with FDR < 0.05 from the microarray DEG analyses (comparing the w0 and w1) and calculated a metric based on the -log (raw p-value) and logFC (fold change) in each group (L–H, H–L, L–L and H–H). We uploaded the Entrez gene ID and metric table in WebGestalt (WEB-based Gene SeT AnaLysis Toolket) 201719, selecting as enrich method the Gene Set Enrichment Analysis (GSEA), 1000
permutations, setting the minimum and maximum number of genes in the category as 5 and 2000, respectively, and as collapse method, the mean between duplicate genes. We performed gene ontology (GO) analyses, KEGG (Kyoto Encyclopedia of Genes and Genomes), Panther and Reactome pathway analyses, and FDR < 0.05 was considered significant.

Co-expression network analysis Data-driven clustering was performed using weighted gene co-expression network analysis (WGCNA) in each group results comparing w0 and w1. For this analysis, we used the WGCNA package in R.

Cross-lagged panel model analysis Significant DEGs that were exclusively found in the L–H and H–L groups were selected to test bidirectional effects of gene expression and CBCL scores through Cross-Lagged Panel Models (CLPM) with baseline and 3-year follow-up assessments. We used the lavaan package in R and FDR < 0.05 was considered significant.

Results From 206 initial samples (103 paired samples), 10 samples from different individuals were excluded during quality control, remaining 196 biological samples. In addition, 10 samples (of these 196) were removed because they had only microarray data from one of the waves (w0 or w1), which would not allow paired analyses. Thus, 186 samples from 93 participants were included in the final analyses.

There were no statistically significant differences in sex or age (w0 or w1) among the groups. As expected, the CBCL score at w0, w1, and ΔCBCL score were significantly different among the groups (Table 1). No statistically significant correlation between CBCL score and age was observed for both waves (w0: \( r = -0.013, p = 0.890 \); w1: \( r = 0.154, p = 0.113 \)) or between ΔCBCL (CBCL w1—CBCL w0) and Δ age (age w1—age w0) (\( r = -0.154, p = 0.113 \)). Moreover, CBCL scores did not increase overtime in the whole BHRCS (\( rp = 0.071; p\)-value = 0.074; Supplementary Fig. S1).

Differentially expressed genes (DEGs) Considering FDR < 0.05, 98 transcripts were differentially expressed in the L–H group (Supplementary Table S2), 33 in the H–L group (Supplementary Table S3), 177 in the H–H group (Supplementary Table S4) and 273 in the L–L group (Supplementary Table S5). Genes that were differentially expressed comparing w0 and w1 in the L–L group are probably related to normal development; and those DEGs in the H–H group might be related to the maintenance of high CBCL scores. Thus, concentrating on genes potentially associated with emergence or remission of psychiatric symptoms, 66 transcripts (and 65 genes, considering that PEX16 gene had two differentially expressed transcripts) were exclusive of L–H group and six of H–H group (Fig. 1). Of these 71 DEGs, 12 (all from L–H comparison) were correlated with age (Supplementary Table S6), remaining 59 genes that seem to be related to the emergence and remission of general psychiatric symptoms.

We also performed cross-sectional analyses comparing individuals with high psychiatric symptoms (CBCL > 30) with those with low psychiatric symptoms (CBCL ≤ 30) at each wave. However, no significant association was found after multiple comparisons correction (Supplementary Tables S7, S8).

For 51 of the 65 DEGs exclusive of L–H and 4 of the 6 DEGs exclusive of H–L, we validated our results using targeted RNA sequencing data in 21 L–H (out of 25) and 7 H–L (out of 11) subjects. A total of 19.66% (10/51) and 25% (1/4) were significant in both techniques, all of them with similar logFC. However, three genes (TOP2B, BCOR, and GCHFR) have their expression levels correlated with

| Variable         | L–H (n = 25) | H–L (n = 11) | H–H (n = 30) | L–L (n = 27) | Statistics |
|------------------|-------------|-------------|-------------|-------------|------------|
| Sex (%)          | M: 15       | M: 5        | M: 19       | M: 15       | \( \chi^2 = 1.169; p = 0.761 \) |
|                  | F: 10       | F: 6        | F: 11       | F: 12       |            |
| Age in years at w0, average (SD) | 10.30 (1.67) | 10.05 (1.22) | 10.02 (1.78) | 9.73 (1.54) | \( F = 0.521; p = 0.669 \) |
| Age in years at w1, average (SD) | 14.12 (1.57) | 14.11 (1.21) | 13.83 (1.82) | 13.46 (1.51) | \( F = 0.856; p = 0.467 \) |
| Total CBCL score at w0, average (SD) | 18.72 (7.69) | 48.18 (15.80) | 55.83 (20.11) | 14.70 (8.10) | \( F = 55.25; p = 2.92 \times 10^{-19} \) |
| Total CBCL score at w1, average (SD) | 47.24 (11.10) | 15.54 (9.23) | 56.93 (21.18) | 15.15 (8.18) | \( F = 51.92; p = 1.71 \times 10^{-19} \) |
| ΔCBCL score(SD)  | 28.52 (9.63) | −32.64 (12.97) | 1.10 (6.90) | 0.44 (7.12) | \( F = 136.39; p = 3.57 \times 10^{-33} \) |

L–H: low Child Behavior Checklist (CBCL) scores at wave 0 (w0) and high at wave 1 (w1); H–L: high CBCL scores at w0 and low at w1; H–H: high CBCL scores at w0 and w1; L–L: low CBCL scores at w0 and w1. 
M: male, F: female, SD: standard deviation.
adjusting for multiple comparisons. The remaining validated genes are presented in Table 2. Of note, 70.6% (36/51) and all (4/4) were in the same direction in both the microarray and RNA sequencing analyses.

**Enrichment and co-expression network analyses**

Genes differentially expressed between w0 and w1 in each group were selected for enrichment analyses. No significant enriched category was found for the genes differentially expressed in the H–L or H–H groups. For the L–H group, we found that the Metabolic pathways (hsa01100, FDR = 0.028, normalized enrichment score—NES = 1.97, Supplementary Table S9) in KEGG and the Metabolism pathway in Reactome (R-HAS-1430728, FDR = 0.016, NES = 2.1, Supplementary Table S9) were significantly positively related categories; and, for the L–L group, only Mitochondrial inner membrane (GO:0005743, FDR = 0.049, NES = −1.97, Supplementary Table S10) from GO cellular component was a significantly negative related category. Considering weighted gene co-expression network analysis (WGCNA), no significant co-expression module was detected at any group comparison, after adjusting for multiple comparisons.

**Cross-lagged panel model analyses**

We performed cross-lagged panel model analyses for all 65 DEGs exclusive of L–H group and 6 of H–L group using the whole sample. CBCL scores were substantially stable over the course of three years in every cross-lagged model tested (β range = 0.606, 0.648; β mean = 0.621). Autoregressive analyses revealed that gene expression was only stable for the TBC1D9, PCNL4, UBAH3B, NCK2, ETS1, ACSL3, and RPS12 genes (Table 3).

Longitudinal analyses showed that SCBL scores at w0 predicted the expression of the RPRD2 gene at w1 (Table 3); however, no gene at w0 was associated with CBCL scores at w1 after adjustment for multiple comparisons. Results from all cross-lagged analyses are described in Supplementary Table S11.

**Discussion**

This study found differentially expressed genes in blood, related to different psychopathological trajectories during childhood and adolescence, focusing on the emergence and remission of psychiatric symptoms in a 3-year follow-up. Comparing w0 and w1, we found 98 transcripts differentially expressed in the L–H group, 33 in the H–L group, 177 in the H–H group and 273 in the L–L group. DEGs found in the last two groups might be related to age, whereas those found in the first two are suggested here as associated to changes in psychopathology. Although others analyzed shared blood and brain transcriptomes in different major mental disorders or the conversion to specific mental disorders, no other study has investigated gene expression and changes in psychopathology during childhood and adolescence.

**DEGs comparing w0 and w1 in the H–H and L–L groups**

A total of 55 DEGs (of 177) in the H–H group and 74 (of 273) in the L–L group (30 were in common for both groups) were associated with age in the discovery stage of a previous study. RSRC2 (Arginine And Serine Rich Coiled-Coil 2) and NLIPI60 (Nucleoporin 160), both differentially expressed in the H–H and L–L groups, were previously correlated with age in peripheral blood leukocytes. Other genes included S100A4 (S100 Calcium Binding Protein A4), which was elevated in w1 in L–L, was previously reported as overexpressed with age. Also, at
Table 3 Significant cross-lagged panel models results.

| Gene                  | CBCL → GENE | Autoregressive | Adj.p value | 95% CI  | p value | Adj.p value | 95% CI  |
|-----------------------|-------------|----------------|-------------|---------|---------|-------------|---------|
| TBC1D9                | -0.061      | 1.44778e-06    | 0.004       | -0.035  | 0.037   | -0.176      | 0.095   |
| PCNXL4                | 0.061       | 0.235          | 0.019       | 0.149   | 0.049   | 0.461       | 0.011   |
| RPRD2                 | -0.174      | 0.271          | 0.035       | 0.174   | 0.035   | 0.166       | 0.011   |
| UBASH3B               | 0.023       | 0.395          | 0.019       | 0.419   | 0.042   | 0.136       | 0.011   |
| NCK2                  | -0.273      | 0.136          | 0.035       | 0.273   | 0.035   | 0.116       | 0.011   |
| ETS1                  | 0.089       | 0.367          | 0.019       | 0.089   | 0.019   | 0.116       | 0.011   |
| ACSL3                 | -0.272      | 0.136          | 0.035       | -0.272  | 0.035   | 0.116       | 0.011   |

DEGs comparing w0 and w1 in the L–H group

A total of 66 transcripts (and 65 genes) seems to be related to the emergence of psychopathology; however, none was among the most significant differentially expressed genes in previous studies that investigated shared gene expression patterns in cross-disorder analyses. Moreover, only the expression of TBC1D9, PCNLX4, UBASH3B, NCK2, ETS1, and ACSL3 seems to be stable over the course of three years. Interestingly, cross-lagged analyses revealed that the expression of RPRD2 (Regulation Of Nuclear Pre-mRNA Domain Containing 2) at w1 can be influenced by CBCL at w0, though very few studies have investigated this gene and none in mental disorders. Among the 65 DEGs, 12 were correlated with age (Supplementary Table S6), and might reflect age-related genes, even though we excluded those DEGs in L–L group. Notably, 10 of 51 DEGs were also validated by another technique, though three were related to age and were not presented in Table 2.

Differentially expressed genes in the L–H group seem to be enriched for metabolic pathways. Of note, COXSB (Cytochrome C Oxidase Subunit 5B, adjusted p = 0.034, logFC = 0.271), which was also validated by RNA sequencing, encodes one of the nuclear-coded polypeptide chains of cytochrome c oxidase, the terminal oxidase in mitochondrial electron transport. Its protein levels seem to be increased in resilience to stress in a rat model of depression. Interestingly, it is differentially expressed in postmortem brain of subjects with ASD or schizophrenia in the study of Gandal et al., which performed a...
meta-analysis of transcriptomic studies. In the same metabolic pathway, SIN3A (SIN3 Transcription Regulator Family Member A, adjusted \( p = 0.048, \log FC = -0.184 \)), although not significantly validated \( ( p = 0.062; \log FC = -0.253) \), seems to have a trend to decrease its expression levels with the increase of psychiatric symptoms. This gene encodes a transcriptional repressor that seems to play a role in cell cycle and proliferation. Witteveen et al. found that its haploinsufficiency is associated with mild syndromic intellectual disability and that SIN3A is essential for cortical brain development. Additionally, SIN3A seems to act as a co-repressor for RE-1 silencing transcription factor (REST), which is also known as neuron-restrictive silencer factor (NRSF). Therefore, SIN3A is a potential gene to play a role in the emergence of psychiatric symptoms, considering that REST/NRSF is known to regulates neurogenesis and neural differentiation.

Although S100A6 (S100 Calcium Binding Protein A6, adjusted \( p = 0.042, \log FC = 0.357 \)) was not analyzed in the RNA sequencing, it seems to be an interesting gene. It encodes a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs and seems to be involved in cellular calcium signaling. Its expression levels were increased in high hallucinations states; however, another study identified a lower expression in peripheral blood lymphocytes from schizophrenia patients compared to controls.

TSC22D4 (TSC22 Domain Family Member 4, adjusted \( p = 0.038, \log FC = 0.276 \)) is a member of the TSC22 domain family of leucine zipper transcriptional regulators. Although few studies investigated this gene, Pfaffenseller et al., investigating differential expression of transcriptional regulatory units in the prefrontal cortex of patients with bipolar disorder (BD), identified that TSC22D4 regulatory unit was increased in BD. Moreover, a SNP near this gene, rs2406253, was associated to self-reported educational attainment in multiple genome-wide association studies.

CCM2 gene (Cerebral Cavernous Malformations 2, adjusted \( p = 0.022, \log FC = 0.331 \)) encodes a scaffold protein that functions in the stress-activated p38 Mitogen-activated protein kinase (MAPK) signaling cascade. Mutations in this gene have been associated to cerebral cavernous malformations, which are brain vascular lesions that may lead to neurologic problems. Although it has been consistently related to this disorder, studies investigating the role of this gene in other disorders are scarce. The peripheral blood expression of CCM2 seems to be related to chronic academic stress in healthy medical students.

SMAD4 gene (SMAD Family Member 4, adjusted \( p = 0.022, \log FC = -0.182 \)) encodes a protein involved in signal transduction of the transforming growth factor-beta superfamily (TGFβ) and bone morphogenetic proteins (BMP). Mutations in this gene are related to Myhre syndrome, a connective tissue disorder with multisystem involvement with or without intellectual disability. Cases with this disorder and SMAD4 mutation and ASD have been described and a SNP within this gene was associated with psychosis in Korean families. This gene seems to play a role in neuronal differentiation and cerebellar development.

CRLF3 (Cytokine Receptor Like Factor 3, adjusted \( p = 0.034, \log FC = -0.187 \)) encodes a largely uncharacterized orphan cytokine receptor that is expressed in various human tissues, including brain. Its function is not well known; however, it has been associated with cell cycle regulation, neuronal morphology, and amyotrophic lateral sclerosis.

SEC62 (SEC62 Homolog, Preprotein Translocation Factor, adjusted \( p = 0.047, \log FC = -0.192 \)) was suggested to play a role in protein translocation, calcium homeostasis and the recovery from endoplasmic reticulum stress. Although no study investigated this gene in mental disorders, SEC62 was among the differentially expressed genes in ASD brain samples, showing decreased transcript levels.

**DEGs comparing w0 and w1 in the H–L group**

Six genes were associated with the remission of psychiatric symptoms, with NDUFA2 being the most significant (adjusted \( p \)-value = 0.014, \( \log FC = -0.466 \)) and the only one validated with another technique (Table 2). NDUFA2 (NADH: Ubiquinone Oxidoreductase Subunit A2) encodes a subunit of the hydrophobic protein fraction of the NADH:ubiquinone oxidoreductase (complex 1), the first enzyme complex in the electron transport chain located in the inner mitochondrial membrane. This gene seems to be upregulated in the prefrontal cortex of an adolescent social isolation, a typical paradigm for schizophrenia, rat model, in line with our results that showed a downregulation in the remission of psychiatric symptoms. Moreover, NDUFA2 was among the differentially expressed genes in postmortem schizophrenia brain samples.

**Strengths and limitations**

Our study presents some strengths and limitations that deserve attention. First, our study design was longitudinal, and it combined the collection of clinical assessment and RNA samples at baseline and at a 3 years follow-up; this can help to elucidate the temporality of the relationship between the development of psychopathology and gene expression changes. Second, psychopathology exists on a continuum in the general population, and population-based studies have demonstrated that symptoms, when considered dimensionally, vary with neurobiological features, providing further support for the examination
of dimensional psychopathology. Third, we studied youths, a much less studied population than adults, even though the psychopathology could be severe, persistent, and with a significant impact over the functional trajectory across the life. Moreover, psychiatric symptomatology during childhood and adolescence predicts persistent mental illness later in life.

The results of this study should be interpreted at light of some limitations. First the relatively small sample size could be an issue, thus, further replication studies are needed. Moreover, even not validated by RNA sequencing, some of the DEGs found in the microarray might be interesting to be replicated in other studies with increased sample sizes, such as PGAM1 and BCL2, which have been associated to mental disorders and were in the same direction of association in the RNA sequencing and microarray. Second, gene expression is tissue-specific, and our findings may not mirror gene expression changes in brain, though some of our results were also found in a meta-analysis of human postmortem brain transcriptomic studies. We would also like to highlight that we were focused in searching potential peripheral markers. Furthermore, it is unknown whether the gene expression differences observed in this study of blood samples will remain stable over time or will change at the onset of a full-blown mental disorder, since RNA levels are dynamic throughout development. However, the low stability and high comorbidity patterns of categorical psychopathology in this age range, as assessed by our current classification (e.g., DSM), support our approach to psychopathology. Finally, it is not possible to know if other factors that may change gene expression influenced our results, such as body mass index, or physical exercise.

Conclusions

In conclusion, we identified 72 transcripts from 71 genes related to the emergence and remission of general psychiatric symptoms during adolescence, though 12 might be age-related. A set of genes seems to be related to metabolic processes, neurodevelopment and mental disorders. One of them (RPRD2) might have its expression at w1 influenced by the CBCL scores at w0. Three (COX5B, SEC62, and NDUFA2) were validated with another technique and were also differentially regulated in postmortem brain of subjects with mental disorders. Our findings support the further exploration of changes in transcription of these set of genes as peripheral markers of emergence or remission of mental disorders and their dimensions.

Acknowledgements

We would like to thank the participants and their families. This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant numbers: 2017/05339-7, 2016/04983-7 and 2014/07280-1 as research projects for, respectively, V.K.O., J.J.M., and S.I.B.), CAPES (Finance Code 001) and FINEP/CT-INFRA-01/2011 160/12 (from LDK). BHRCs was funded by the National Institute of Developmental Psychiatric for Children and Adolescent (INPD), supported by CNPq-FAPESP (573974/2008-0; 465550/2014-2 and 2008/57896-8) and INCT (FAPESP/CNPq, 2014/50917-0).

Author details

1. L.INC - Interdisciplinary Laboratory of Clinical Neurosciences, Universidade Federal de São Paulo (UNIFESP), São Paulo, Brazil. 2. Post-Graduation Program in Psychiatry and Medical Psychology, UNIFESP, São Paulo, Brazil. 3. Genetics Division, Department of Morphology and Genetics, UNIFESP, São Paulo, Brazil. 4. Laboratory of Cytogenomics, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo (FMUSP), Sao Paulo, Brazil. 5. Centro de Oncologia Molecular, Hospital Sírio-Libanès, São Paulo, Brazil. 6. Universidade de São Paulo (USP), Sao Paulo, Brazil. 7. Department of Psychiatry, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil. 8. Kingston General Hospital and Providence Care Hospital, Department of Psychiatry, School of Medicine, Queen’s University, Kingston, ON, Canada. 9. Centre for Neuroscience Studies (CNS), Queen’s University, Kingston, ON, Canada. 10. Medicine and Medical Sciences Graduate Program, School of Medicine, Brain Institute (InsCer), Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil. 11. Laboratórios de TDAH e Psiquiatria do Desenvolvimento, Hospital de Clínicas de Porto Alegre, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Brazil. 12. Department of Psychiatry, FMUSP, São Paulo, Brazil

Conflict of interest

L.A.R. has received grant or research support from, served as a consultant to, and served on the speakers’ bureau of Eli Lilly and Co., Janssen, Medice, Novartis and Shire. The ADHD and Juvenile Bipolar Disorder Outpatient Programs chaired by L.A.R. have received unrestricted educational and research support from the following pharmaceutical companies: Eli Lilly and Co., Janssen, and Novartis. L.A.R. has received authorship royalties from Oxford Press and ArtMed and travel grants from Shire to take part in the 2018 APA annual meeting and from Novartis to take part of the 2016 AACAP annual meeting. E.B. has received honoraria as speaker/consults from Daiichi-Sankyo and from the elaboration of educational material from Daiichi-Sankyo, Ache, Eurofarma, and Sanofi. In the last five years, R.A.B. has received research grants from Janssen Cilag, Roche and the governmental funding research agencies: CAPES, CNPq and FAPESP (2016/022465 and 2011/50740-5); has been a forum consultant for Janssen and Roche; and has participated in speaker bureaus for Ache, Janssen and Lundbeck, all outside the submitted work. The other authors have nothing to disclose.

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/s41398-020-0772-3).

Received: 14 October 2019 Revised: 14 February 2020 Accepted: 27 February 2020
Published online: 17 March 2020

References

1. Copeland, W. E. et al. Diagnostic transitions from childhood to adolescence to early adulthood. J. Child Psychol. Psychiatry Allied Discip. 54, 791–799 (2013).
2. Martel, M. M. et al. A general psychopathology factor (P factor) in children: structural model analysis and external validation through familial risk and child global executive function. J. Abnorm. Psychol. 126, 137–148 (2017).
3. 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).
4. Franke, B. et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. Mol. Psychiatry 17, 960–987 (2012).
