BMC Psychiatry

RESEARCH

Open Access

Disrupted genes and pathways in schizophrenia: a robust analysis of the brain and blood

Rebeca Rebouças[1,2,3,5], Mariana Araújo-Pereira[1,2,5], Tiago F. Mota[1,2,5], Henrique Q. S. Scoppetta[1,2,5], Moreno M. S. Rodrigues[1,2,4], Bruno B. Andrade[1,2,3,5], Eduardo R. Fukutani[1,2,5†] and Artur Trancoso Lopo de Queiroz[1,2,5*†]

Abstract

Background Schizophrenia (SCZ) is a neuropsychiatric disorder that is not yet fully understood, characterized by behavioral, emotional, and cognitive dysfunctions. In this study, we analyzed publicly available gene-expression data to identify SCZ-associated genes and pathways, offering deeper insights into its biological basis.

Methods The data collection aimed to retrieve gene expression databases featuring samples from the brain's prefrontal cortex and blood. After performing exclusion criteria and quality checks, 17 datasets were retrieved from six different sources: Brain biopsy ($n=6$), whole blood ($n=2$), peripheral blood mononuclear cells ($n=2$), leukocytes ($n=1$), lymphocytes ($n=1$) and isolated neurons ($n=5$). We used four brain datasets as the discovery set for the initial analysis. Differentially expressed genes (DEGs) were identified by comparing SCZ patients with controls and were subsequently used in enrichment analysis. Finally, we applied feature selection to pinpoint the most informative DEGs for SCZ and evaluated their accuracy using data from other tissues.

Results This analytical approach identified 532 DEGs. Feature selection revealed three genes—*HUWE1*, *PTGDS*, and *RPL31*—that effectively discriminated SCZ from control samples. The 3-gene model's performance was validated in other datasets, presenting accuracy (> 72%) in brain tissue, whole blood, PMBCs and leukocytes. Furthermore, the enrichment analysis revealed a potential linkage with neurodegenerative biological pathways.

Conclusion These insights open new avenues for exploring key genes for SCZ, which can lead to new therapeutic targets or tools for diagnosis, potentially transforming the management of SCZ and enhancing patient care.

Keywords Schizophrenia, Bioinformatics, IPD meta-analysis, Data mining, Neurodegenerative disorders

†Eduardo R. Fukutani and Artur Trancoso Lopo de Queiroz contributed equally to this work.

*Correspondence:
Artur Trancoso Lopo de Queiroz
artur.queiroz@fiocruz.br
[1]Instituto de Pesquisa Clínica e Translacional (IPCT), Faculdade Zarns, Clariens Educação, Salvador, Bahia, Brazil
[2]Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador, Brazil
[3]Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador 40290-150, Brazil
[4]Transplant / Oncology & Infectious Disease Program, School of Medicine, Johns Hopkins University, 2000 East Monument St, Baltimore, MD 21205, USA
[5]Laboratório de Pesquisa Clínica e Translacional (LPCT), Fundação Oswaldo Cruz, Salvador, Bahia 40296-710, Brazil

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Background

Schizophrenia (SCZ) is a multifactorial neuropsychiatric disorder characterized by behavioral, emotional, and cognitive dysfunctions with worldwide prevalence rates ranging between 0.33% and 0.75% [1, 2]. Although the widely used Diagnostic and Statistical Manual of Mental Disorders (DSM) provides classification criteria, diagnosing SCZ remains challenging, requiring extensive psychiatric assessment [3]. An additional diagnostic challenge involves distinguishing SCZ from other mental health conditions, as its symptoms overlap with psychiatric disorders including major depressive disorder, schizoaffective disorder, and bipolar disorder [4]. Treatment usually involves antipsychotics and psychotherapy and, in some cases, hospitalization. Despite established diagnostic and treatment criteria, the neurobiological mechanisms underlying SCZ remain poorly understood. This knowledge gap has driven increased research into SCZ-specific genetic expression patterns in recent years [5].

Recent years have seen intensified efforts to characterize the SCZ expression profile, driven by the need to improve diagnosis, enhance treatment, and elucidate its underlying mechanisms. A primary focus has been the prefrontal cortex (PFC), as this brain region supports cognitive functions frequently impaired in SCZ - including attention, working memory, and cognitive flexibility [6]. Furthermore, a previous study observed that patients with SCZ displayed altered activation of PFC during cognitive tests, suggesting that this region's disruption could be related to such deficits [7]. While brain biopsy-based biomarkers remain clinically impractical for diagnosis, numerous studies have examined this approach to identify potential biomarkers and elucidate key mechanisms in schizophrenia pathogenesis [8, 9].

Blood represents another crucial source for biomarker discovery due to its easy accessibility, with whole blood and peripheral blood mononuclear cells (PBMCs) being particularly valuable. Among several previously proposed blood based biomarkers [10], a microarray study identified six genes, including *BTG1*, *GSK3A*, *HLA-DRB1*, *HNRPA3*, *SELENBP1*, and *SFRS1*, which showed differential expression in brain and peripheral blood samples from individuals with SCZ. *SELENBP1* emerged as a particularly potential biomarker candidate, demonstrating consistent differential expression across all examined tissues. However, subsequent replication studies yielded inconsistent results, highlighting the necessity for additional validation [11].

Machine learning has emerged as a transformative approach for transcriptomic data analysis in several disease contexts, offering a capacity to uncover complex patterns beyond the reach of traditional methods [12–14]. Unlike conventional differential expression analyses, which rely on predefined case-control comparisons and emphasize statistical significance of individual genes, machine learning leverages multidimensional models that capture nonlinear relationships, spatial dependencies, and higher-order interactions within raw or normalized data [15, 16]. This enables not only more accurate classification of patients, but also the discovery of novel biomarkers and regulatory interactions with enhanced robustness [17–19]. In contrast to the lens of classical approaches, machine learning provides a multidimensional framework that is capable of identifying highly informative transcriptomic alterations within disease contexts, thereby advancing the identification of disrupted pathways and molecular mechanisms underpinning [15].

In this study, we employed a robust bioinformatics pipeline incorporating machine learning algorithms to identify transcripts significantly associated with SCZ pathophysiology. By analyzing publicly available postmortem PFC datasets, we applied this approach - previously validated in other disease contexts [12–14] - to elucidate key genes and pathways associated with SCZ. Additionally, we evaluated the diagnostic accuracy of identified biomarkers in more clinically accessible samples (e.g., blood). Our findings elucidate SCZ-associated genetic mechanisms and molecular pathways, advancing the understanding of disease pathophysiology. This work establishes a foundation for developing improved diagnostic tools and targeted therapies, ultimately aiming to enhance the quality of life for individuals with SCZ.

Methods

Search strategy and selection criteria

The 209 publicly available datasets were screened from the NCBI's Bioproject database (https://www.ncbi.nlm.nih.gov/bioproject/) on February 07, 2025, using the query "Schizophrenia[All Fields] OR Schizophrenic[All Fields] AND ("transcriptome gene expression"[Filter] AND "org human"[Filter])". Data collection was carried out independently by five researchers, each following the same standardized protocol. The first screening process aimed to exclude datasets not related to schizophrenia or not presenting the disease information on the features within their metadata ($n = 68$), had no control samples within ($n = 22$), did not feature human samples ($n = 1$), only measured miRNAs or circRNAs expression ($n = 5$) or had samples derived from induced pluripotent stem cells, cell lines or cells with modified expression in vitro ($n = 73$), excluding 169 datasets. The remaining 40 datasets were mostly of brain and blood samples, the second screening removed inadequate datasets which didn't had gene symbols nor included the selected genes for validation ($n = 8$), datasets which were unavailable to download ($n = 3$), brain datasets featuring other regions than prefrontal cortex ($n = 11$ and datasets featuring methylation

($n = 1$), a total of 23 removed datasets. After the screening, only datasets featuring samples from the brain's prefrontal cortex, PBMC, leukocytes, lymphocytes, and whole blood, with both SCZ and control samples, were included in the analysis.

The final sampling for this analysis resulted in 17 datasets, for microarray, 6 featuring samples of whole brain, 2 featuring brains' prefrontal cortex pyramidal neuron cells, 1 featuring brains' prefrontal cortex parvalbumin neuron cells, 2 featuring whole blood, and 2 featuring PBMCs. For RNA-seq, 4 datasets were included in this analysis, being 1 featuring peripheral blood, 1 featuring lymphocytes and 2 featuring the brain's PFC isolated neurons and oligodendrocytes. All patients with SCZ were diagnosed according to DSM-IV or DSM-III in their respective original studies, the great majority on treatment with antipsychotic medicines, such as haloperidol, risperidone, olanzapine, clozapine and quetiapine. The overall age and sex distribution among each dataset with available information was assessed using the *compareGroups* package (version 4.9.1) [15], statistical difference in age among the groups was tested using Mann-Whitney *U*, while the sex frequency was tested using Pearson's chi-square.

The microarray datasets were downloaded using the *GEOquery* package (version 2.72.0) [16], using the function *getGEO* with the parameter *GSEMatrix* set as *TRUE* to download the expression matrix and *FALSE* to download the metadata. Gene probes were collapsed using the *WGCNA* package (version 1.73) [17], using the function *collapseRows* with the expression data and annotation data as input, method parameter was set as "*absMaxMean*". Outlier samples' detection and removal were performed using the *arrayQualityMetrics* package (version 3.60.0), using the function outlier with the method set as "*KS*" [18]. For this analysis, only control and SCZ samples were included, even from datasets featuring other conditions such as bipolar disorder and PTSD. After the download, the datasets GSE17612 and GSE53987 were transformed into log2, due to the absence of previous normalization, datasets with previous normalization were maintained untransformed. On the other hand, the RNA-seq datasets had their processed count tables retrieved directly from the geoNCBI's download data platform. The download and all subsequent analyses were performed in *R* (version 4.4.2).

Data analysis

Analysis design and batch effect correction

The datasets were split between discovery and validation sets. All combinations of brain datasets were merged to assess the most efficient combination based on the number of genes and samples comprising the merged dataset. The merging was performed using the *dplyr* package's *join_all* function [19], merging by gene symbols in order to retrieve only common genes among the datasets, then the number of common genes and samples for each combination were assessed (Supplementary Material 1). The most efficient discovery set was composed of 4 datasets (GSE17612 [20], GSE12649 [21], GSE53987 [22], and GSE92538 [22]), with a total of 226 samples (124 control and 102 SCZ) and 10,829 genes. The remaining 13 datasets were kept for validation. Afterward, the batch effect was corrected using the Surrogate Variable Analysis, available at the *sva* package's (version 3.52.0) [23]. The batch effect correction was evaluated using Principal Component Analysis (PCA) before and after the ComBat, principal components were retrieved using the *stats* package (version 4.2.2) *prcomp* function, with the parameter scale set as *TRUE*.

Furthermore, the validation microarray sets were arranged and merged according to the datasets' tissue. The brain validation dataset was composed of the 2 remaining brain datasets (GSE21138 [24] GSE12654 [25]), featuring 84 samples (43 control and 41 SCZ). The whole blood validation dataset was composed of 2 datasets (GSE38484 and GSE38481 [26]), with 235 samples (115 control and 120 SCZ). The PBMCs validation dataset al.so was composed of 2 datasets (GSE18312 [27] and GSE27383 [27]), presenting 88 samples (34 control and 54 SCZ). Regarding isolated neurons, the Pyramidal neuron dataset was composed by 2 datasets (GSE93987 [28] and GSE87610 [28]) and presented 309 samples (178 control and 131 SCZ). The parvalbumin neuron dataset, GSE93577 [24], was maintained as an individual dataset and had 139 samples (69 control and 70 SCZ).

For the RNA-seq validation datasets, the lymphocyte dataset GSE165604 [29]was included individually, featuring 37 samples (18 control and 19 SCZ). Peripheral blood was also included as validation, with the dataset GSE263180, which featured 29 samples (20 control and 9 SCZ). Lastly, isolated neurons and oligodendrocytes were also assessed in validation, with 2 datasets (GSE112523 [30] and GSE107638 [31]) being joined, totaling 74 samples (44 control and 30 SCZ) after removing non-SCZ or control samples. All merged validation datasets for each tissue with 2 or more datasets also had their batch effects corrected and evaluated using the same packages and methods as the discovery set.

Feature selection: differential expression analysis and machine learning

After the batch effect correction, differential expression analysis was applied to the discovery set, using the *limma* package (version 3.60.6) [32]. This analysis used a design data created based on groups (SCZ or control) information and expression data as input, the statistical values were calculated by fitting a linear model with the function

lmFit, including contrasts based on groups with the function *contrasts.fit*, and applying empirical Bayes moderation of the standard errors with the function *eBayes*, all functions parameters were set as default, aside from the input expression and design data. The performed comparison was SCZ vs. control groups, set with SCZ - control at the *limma*'s *makeContrasts* function, differentially expressed genes (DEGs) were identified using thresholds of |logFC| > 1 and FDR-adjusted p-value < 0.05.

The identification of the most informative DEGs, their expression data was retrieved, and feature selection was performed using the random forest algorithm, implemented in the *randomForest* (version 4.7–1.2) [33] and *caret* (version 7.0–1) [34] packages. The random forest's parameters were *mtry* set as 24, *maxnodes* set as 6, and generating 400 trees. The most informative DEGs were selected based on their mean decrease gini, which is related to the tree's group homogeneity in its leaves, in order to retrieve the genes with highest classification potential. To assess these most informative genes' overall profile, a heatmap with unsupervised hierarchical clustering was created with these genes' z-scaled expression values, using the Manhattan distance and ward D2 clustering method. A k-fold cross-validation was also performed to evaluate the predictive model, using the caret package's confusion matrix to assess its overall performance, with 5-folds and 100 repetitions. The inputs for the confusion matrix were the real classes of each sample and predictions performed by the model, trained with the "rf" method. The *trControl* parameter was set using the method "*repeatedcv*", fold of 5, and repeat 100 times.

The validation step was performed to evaluate the most informative genes' performance in discriminating SCZ from control samples in other tissue types. For this, the samples' classification was performed in each validation dataset, classifying its samples into SCZ or control, the overall classification accuracy was assessed comparing the models' predicted sample classification and the real sample classification, its overall performance was assessed using ROC curves, with the package *pROC* (version 1.18.5) [35]. A total of eight validation sets were included, featuring samples of brain, whole blood, PBMCs, leukocytes, lymphocytes, brain PFC's pyramidal neurons, parvalbumin neurons and isolated neurons.

Identifying the most represented pathways: enrichment analysis

To identify which biological pathways were associated with the DEGs and random forest-selected genes, an enrichment analysis was performed with the DEGs' available Entrez IDs, retrieved using the *biomaRt* package (version 2.60.1) [36], and sorted fold change values as input, using the *clusterProfiler* [37] package (version 4.12.6) [37]. The main employed function was the *compareCluster*, with a maximum gene set size of 500, minimum gene set the size of 10, q value cutoff of 0.2, p-value cutoff of 0.05 and p adjust method "FDR" as parameters. This analysis was performed twice, using the parameter "fun" set as "enrichPathway" for using the Reactome [38] and "enrichKEGG" for KEGG [39] databases, aiming to provide further information regarding the DEGs and random forest genes' associated pathways. The list with all packages employed in this study is available as Supplementary Material (Supplementary Material 1).

Results

Study design and population description

From the 209 initially assessed datasets, 17 were retrieved for further analysis. The sampling details, including the exclusion criteria applied to each dataset, are available as supplementary material (supplementary material 2). The PFC, composed mainly of Brodmann Areas (BA) 10, 46 and 9, has been selected due to neurobiological evidence demonstrating its relevance to SCZ pathogenesis, as this region plays a central role in supporting cognitive functions well known to be impaired in this disease [6]. Furthermore, the PFC's particular prolonged maturation, prolonged through adolescence, aligns with the typical onset period of SCZ [40].

After assessing all combinations of brain's prefrontal cortex datasets, 4 datasets were selected as the Discovery dataset, while the remaining 2 brain datasets were used for validation, alongside 2 featuring PBMCs, 2 featuring whole blood, 2 featuring PFC's isolated pyramidal neurons, 1 featuring prefrontal cortex's isolated parvalbumin neurons, 1 featuring leukocytes, 1 featuring lymphocytes and two featuring PFC's isolated neurons (Fig. 1). All assessed combinations's datasets, number of genes and samples have been summarized (Supplementary material 3). Regarding the age and sex distribution among the datasets, several datasets had no age and sex related information available (GSE12649, GSE12654, GSE87610, GSE93577, GSE93987). Among the datasets with available age and sex information, the only significant difference was in sex distribution, observed in the Whole Blood set (comprising GSE38481 and GSE38484), where the SCZ group included more males than the control group (Supplementary material 4).

Differential expression and feature selection to characterize schizophrenia

The four datasets comprising the discovery set were retrieved, and batch effects were corrected (Figure S1). The final discovery dataset included 226 samples (124 controls and 102 SCZ cases) and 10,829 genes A total of 532 DEGs were identified when comparing the SCZ vs. control groups, of which 383 were upregulated and 149

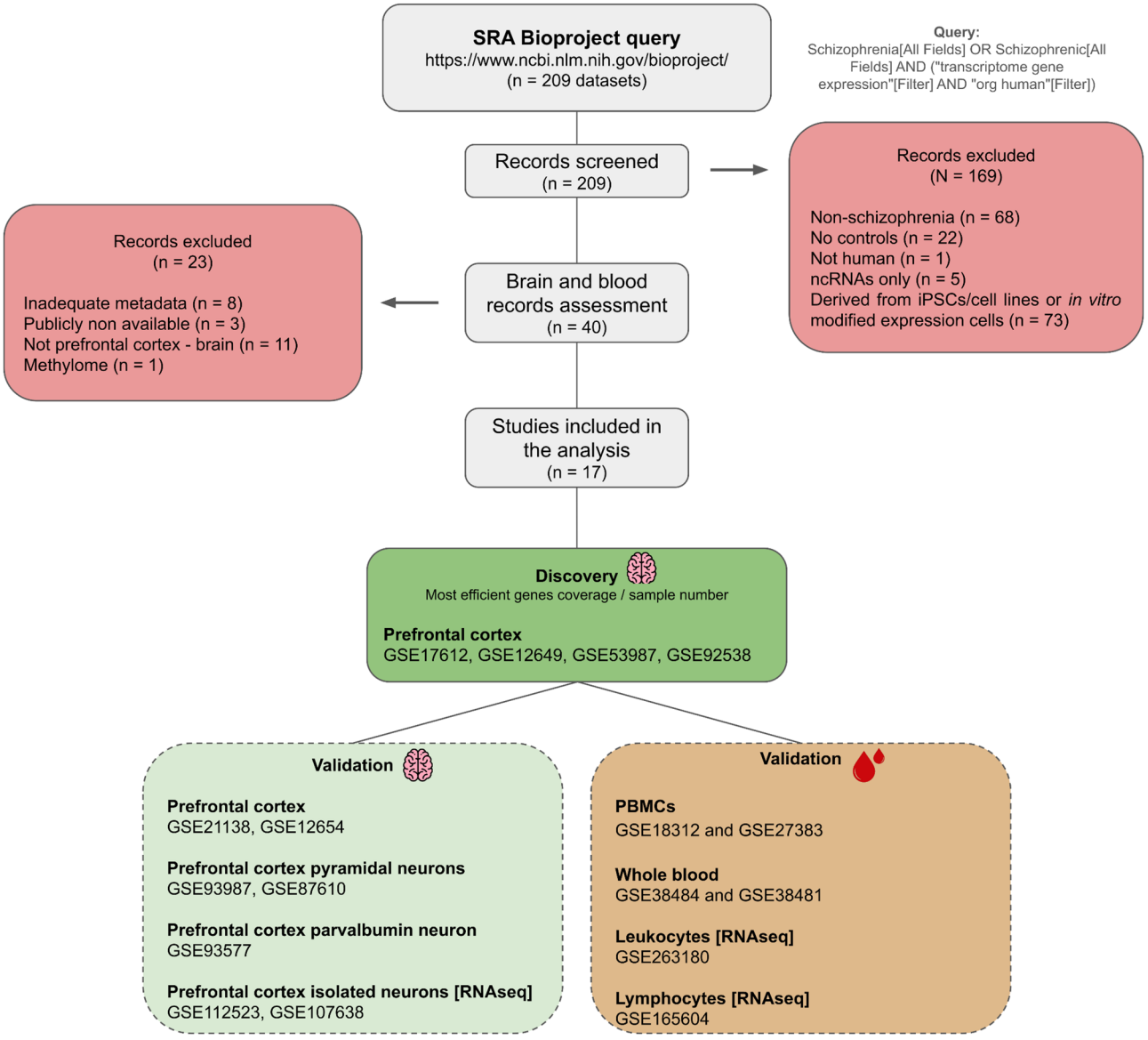

Fig. 1 Sampling flow chart. Flow chart depicting the sampling process for this Individual participant data meta-analysis, with details about exclusion criteria and numbers of excluded/included studies

were downregulated. A summary of all identified DEGs and their related statistics are available as Supplementary material (Supplementary material 5). To assess potential p-value inflation in our differential expression analysis, we generated a QQ-plot to visualize the observed versus expected p-value distributions (Figure S2). To identify the most informative differentially expressed genes (DEGs) distinguishing SCZ from control samples, we performed feature selection using the random forest algorithm. The top three genes—*HUWE1*, *PTGDS*, and *RPL31*—were selected based on the mean decrease in the Gini index, forming the most predictive model (Fig. 2A & Supplementary Material 6). To evaluate the expression profile of these genes, a heatmap with z-score-scaled expression data was employed (Fig. 2B). Two major clusters were observed on the heatmap, the first being composed of mainly SCZ samples (84.31%), while the second cluster was formed mainly by control samples (87.09%). Furthermore, all three genes were upregulated in SCZ samples: *RPL31*, *HUWE1*, and *PTGDS* had fold changes of 7.62, 20.84, and 26.90, respectively. The k-fold cross-validation analysis revealed a good discrimination performance in the discovery samples, with an accuracy of 1.0, 95% confidence interval (C.I.) [0.98, 1], no information rate of 0.55, kappa of 1, sensitivity of 1, specificity of 1, positive and negative predictive values of 1.

The validation sets were separated based on the tissue: brain, whole blood, PBMCs, leukocytes, parvalbumin

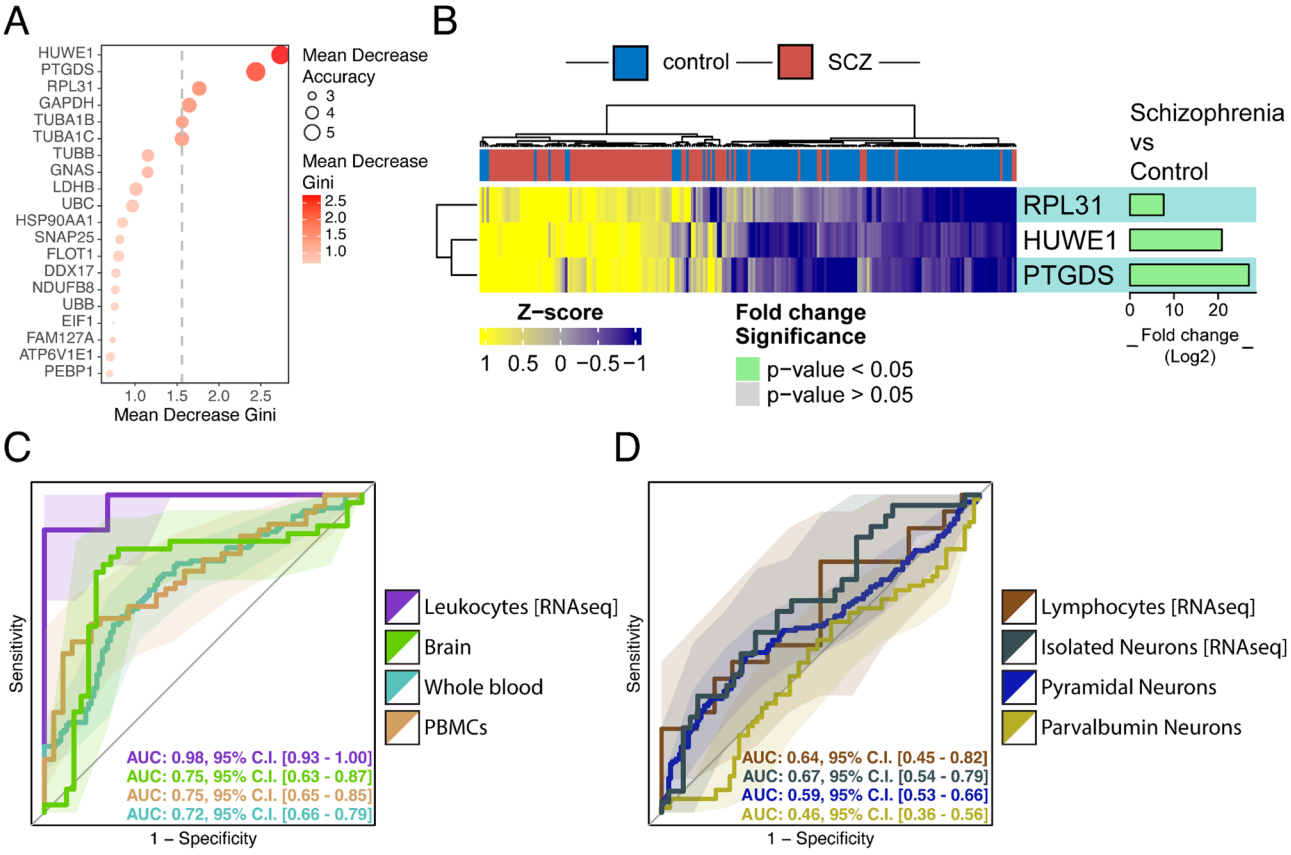

Fig. 2 Most informative genes identification and discriminative performance assessment. **A**. Dot plot displaying the random forest's importance values used to select the most informative genes to separate SCZ from control samples. *Mean decrease accuracy* is related to the models' accuracy drop when variables are removed, while *Mean decrease gini* is related to the tree branches' homogeneity variation when variables are removed. **B**. Heatmap with hierarchical clustering displaying the most informative genes' z-scaled expression data in the discovery set, topside bars represent sample groupings (red for SCZ and blue for control). The employed distance and clustering methods were Manhattan and ward D2, respectively. Barplots at the right side display fold change values calculated in the SCZ vs. control comparison performed using limma, green bars represent significant fold changes by P-value adjusted by False discovery rate. **C**. ROC curves displaying the most informative genes' performance when classifying SCZ and control samples, Y axis represent the sensitivity and X axis the 1-specificity. Curves are colored by sets (purple for leukocytes, green for brain, light blue for whole blood and light brown for PBMC). **D**. ROC curves displaying the most informative genes' performance when classifying SCZ and control samples. Curves are colored by sets (brown for lymphocytes, grey for isolated neurons, blue for pyramidal neurons and dark yellow for parvalbumin neurons)

neurons, pyramidal neurons, lymphocytes, and isolated neurons; each of them being composed of 1 to 3 different datasets. Each tissue composed of 2 or more datasets had its batch effect corrected before proceeding with the performance assessment (Figure S3). The validation data with brain cells was composed of 2 different datasets and presented 84 samples (43 control and 41 SCZ). The 3-genes in this set had their expression values retrieved and the model's predictive performance assessed using an ROC curve, presenting an AUC of 0.75, 95% C.I. [0.63, 0.87] (Fig. 2C). The same was done with the whole blood set, which was composed of 2 different datasets, presenting 235 samples (115 control and 120 SCZ). In this dataset, the ROC curve had an AUC of 0.72, 95% C.I. [0.66, 0.79] (Fig. 2C). The PBMC set was also composed of 2 datasets, had a total of 88 samples (34 control and 54 SCZ), and the presented AUC in this tissue was 0.75, 95% C.I. [0.65, 0.85] (Fig. 2C). The leukocyte RNA-seq dataset was composed of 1 dataset featuring 29 samples (20 control and 9 SCZ), the AUC when classifying these samples was 0.98, 95% C.I. [0.93–1.00] (Fig. 2C).

Next, the pyramidal neurons microarray set was assessed, composed of 2 datasets and with 309 samples (178 control and 131 SCZ). Unlike the other tissues, the 3-gene model showed reduced discriminative performance in this dataset, with an AUCs of 0.59, 95% C.I. [0.53, 0.66] (Fig. 2D). Following, the parvalbumin neurons set was composed of only 1 dataset, presenting 139 samples (69 control and 70 SCZ). The 3-gene model accuracy was also reduced in this set, with an AUC of 0.46, 95% C.I. [0.36, 0.56] (Fig. 2D). The lymphocyte classification analysis, comprising 1 RNA-seq dataset (37 total samples: 18 controls and 19 SCZ cases), demonstrated modest discriminative performance with an AUC of 0.64 (95% CI: 0.45–0.82) (Fig. 2D). Lastly, the RNA-seq set with isolated neuron samples was composed of 2 datasets

with 74 samples (44 control and 30 SCZ). The model's performance in this set achieved an AUC of 0.67, 95% C.I. [0.54–0.79] (Fig. 2D). The featured tissues, number of samples, calculated AUCs, type of data, and datasets composing each validation set were compiled in Table 1. These findings suggest that the model's discriminatory signal may not be detectable in isolated neurons or lymphocytes but could be transmitted across the blood-brain barrier and rise from other supporting cell types.

Pathways associated with schizophrenia's DEGs are also associated with other neuronal and psychiatric disorders

Following the identification of a SCZ-associated transcriptomic signature and validation of the model's accuracy across independent datasets (including multiple tissue types), we next investigated the biological pathways associated with these candidate genes. From the DEGs we identified 92 enriched pathways in the KEGG database and 432 in the Reactome database; the tables with all enriched pathways and input data are available as Supplementary material (Supplementary material 7). We then extracted the top 10 most enriched pathways (ranked by gene ratio) for both up and downregulated DEGs from each database. To ensure comprehensive coverage, we additionally included pathways that showed enrichment in both DEG groups' top 10, even if they didn't rank among the top 10 in one group. Finally, we highlighted the top 10 pathways associated with our random forest model's top 5 most important genes.

First, the KEGG database's top 10 pathways enriched by upregulated DEGs may suggest consistent associations between the identified genes with neuronal and psychiatric disorders. Pathways such as "Phagosome", "Alzheimer's disease", "Diabetic cardiomyopathy", "Amyotrophic lateral sclerosis", "Prion disease", "Huntington's disease", "Pathways of neurodegeneration – multiple diseases", and "Parkinson's disease" included one or more genes from the random forest model in its composition, potentially involving neuronal loss or inflammatory processes [41, 42]. (Fig. 3). Alongside them, the "Synaptic vesicle cycle and Oxidative phosphorylation" pathways were also enriched in this gene set (Fig. 4). On the other hand, its downregulated counterpart's top 10 enriched pathways were mostly associated with metabolic pathways, such as "Gastric acid secretion", "Aldosterone synthesis and secretion", "Carbohydrate digestion and absorption", "Growth hormone synthesis", secretion and action", "Thyroid hormone synthesis", "Ferroptosis", "Aldosterone – regulated sodium reabsorption", "Glutamatergic synapse", "cGMP – PKG signaling pathway" and "Mineral absorption" (Fig. 4).

Regarding the Reactome database's top 10 pathways enriched by upregulated DEGs, the pathways "Transmission across Chemical Synapses", "Neurotransmitter receptors, and postsynaptic signal transmission", "HSP90 chaperone cycle for steroid hormone receptors (SHR) in the presence of ligand", "Neuronal System", and "Cellular response to starvation" had random forest genes in their compositions (Fig. 3). Furthermore, the pathways "citric acid (TCA) cycle and respiratory electron transport", "Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins", "Transferrin endocytosis and recycling", "Iron uptake and transport" and "Respiratory electron transport" were also enriched by this set of genes (Fig. 4). The downregulated DEGs' enriched pathways were "Opioid Signalling", "Signaling by NTRK1 (TRKA)", "Signaling by NTRKs", "VEGFA – VEGFR2 Pathway", "CS/DS degradation", "CLEC7A (Dectin – 1) induces NFAT activation", "Golgi Associated Vesicle Biogenesis", "Fcgamma

Table 1 Characterization of validation datasets, exhibiting type of tissue, number of samples, overall performance of the most informative genes when discriminating SCZ and control samples, type of data (microarray or RNA-seq) and datasets that compose the validation set. The assessed brain's prefrontal cortex are mainly from brodmann's areas (BA) 10, 46 and 9

Tissue	Samples	AUC	Type of data	Datasets composing
Brain prefrontal cortex (BA 10, 46, 9)	*N* = 84 [SCZ: 41, Control: 43]	0.75, 95% C.I. [0.63, 0.87]	Microarray	GSE12654, GSE21138
PBMC	*N* = 88 [SCZ: 54, Control: 34]	0.75, 95% C.I. [0.65, 0.85]	Microarray	GSE18312, GSE27383
Whole Blood	*N* = 235 [SCZ: 120, Control: 115]	0.72, 95% C.I. [0.66, 0.79]	Microarray	GSE38484, GSE38481
Brain dorsolateral prefrontal cortex (pyramidal neurons)	*N* = 309 [SCZ: 131, Control: 178]	0.59, 95% C.I. [0.53, 0.66]	Microarray	GSE93987, GSE87610
Brain dorsolateral prefrontal cortex (parvalbumin neurons)	*N* = 139 [SCZ: 70, Control: 69]	0.46, 95% C.I. [0.36, 0.56]	Microarray	GSE93577
Leukocytes	*N* = 29 [SCZ: 9, Control: 20]	0.98, 95% C.I. [0.93–1.00]	RNA-seq	GSE263180
Lymphocytes	*N* = 37 [SCZ: 18, Control: 19]	0.64, 95% C.I. [0.45–0.82]	RNA-seq	GSE165604
Brain prefrontal cortex (neurons)	*N* = 139 [SCZ: 70, Control: 69]	0.67, 95% C.I. [0.54–0.79]	RNA-seq	GSE112523, GSE107638

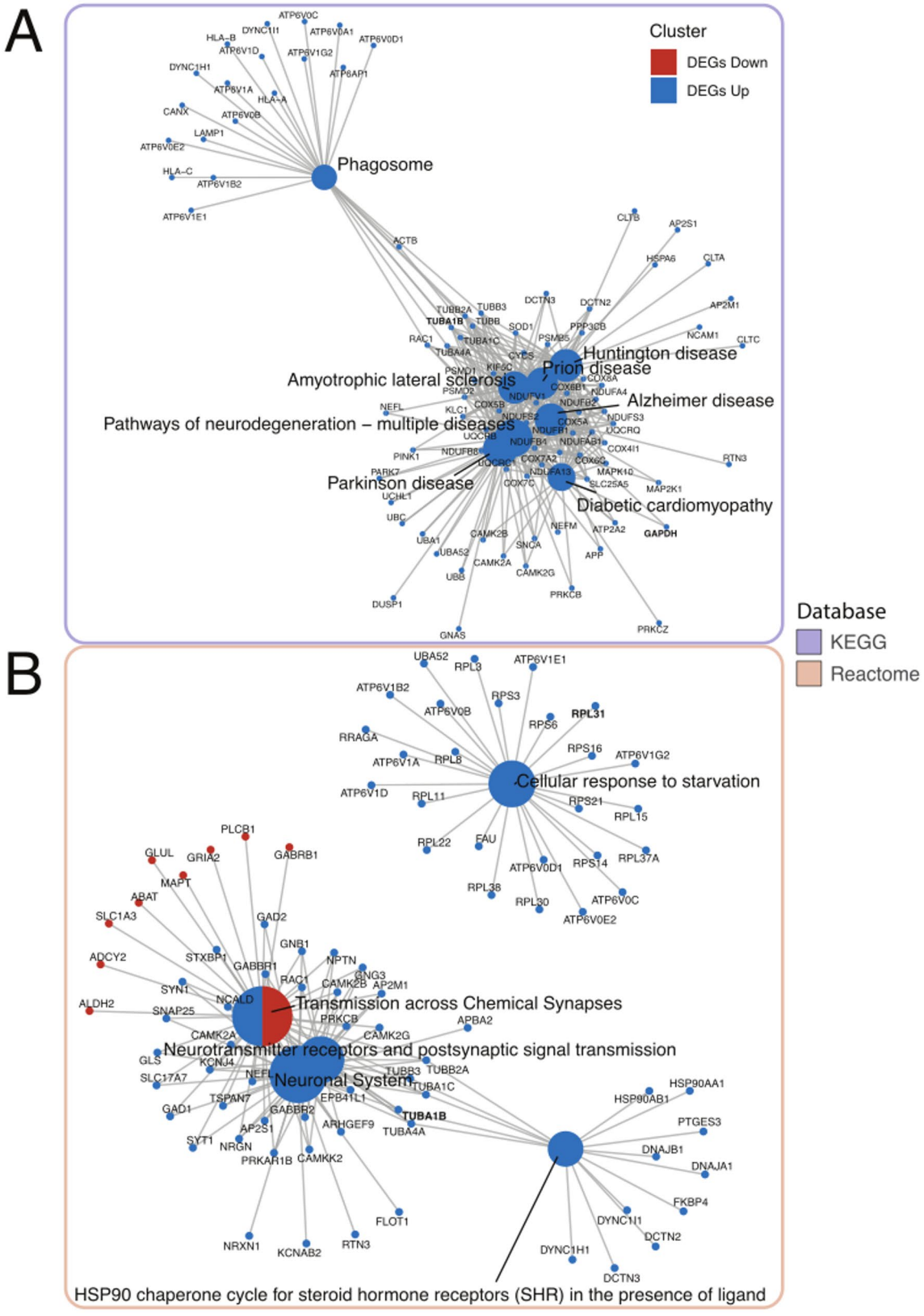

Fig. 3 Network of the most informative genes in the enrichment analysis. Enriched pathways in both KEGG and Reactome databases. The network displays the pathways that were enriched using the up and downregulated DEGs as input, highlighting the pathways containing genes selected by the random forest. Red and blue dots represent the up and down regulated DEGs, respectively. Dot size for pathways is related to the number of genes comprising the pathway

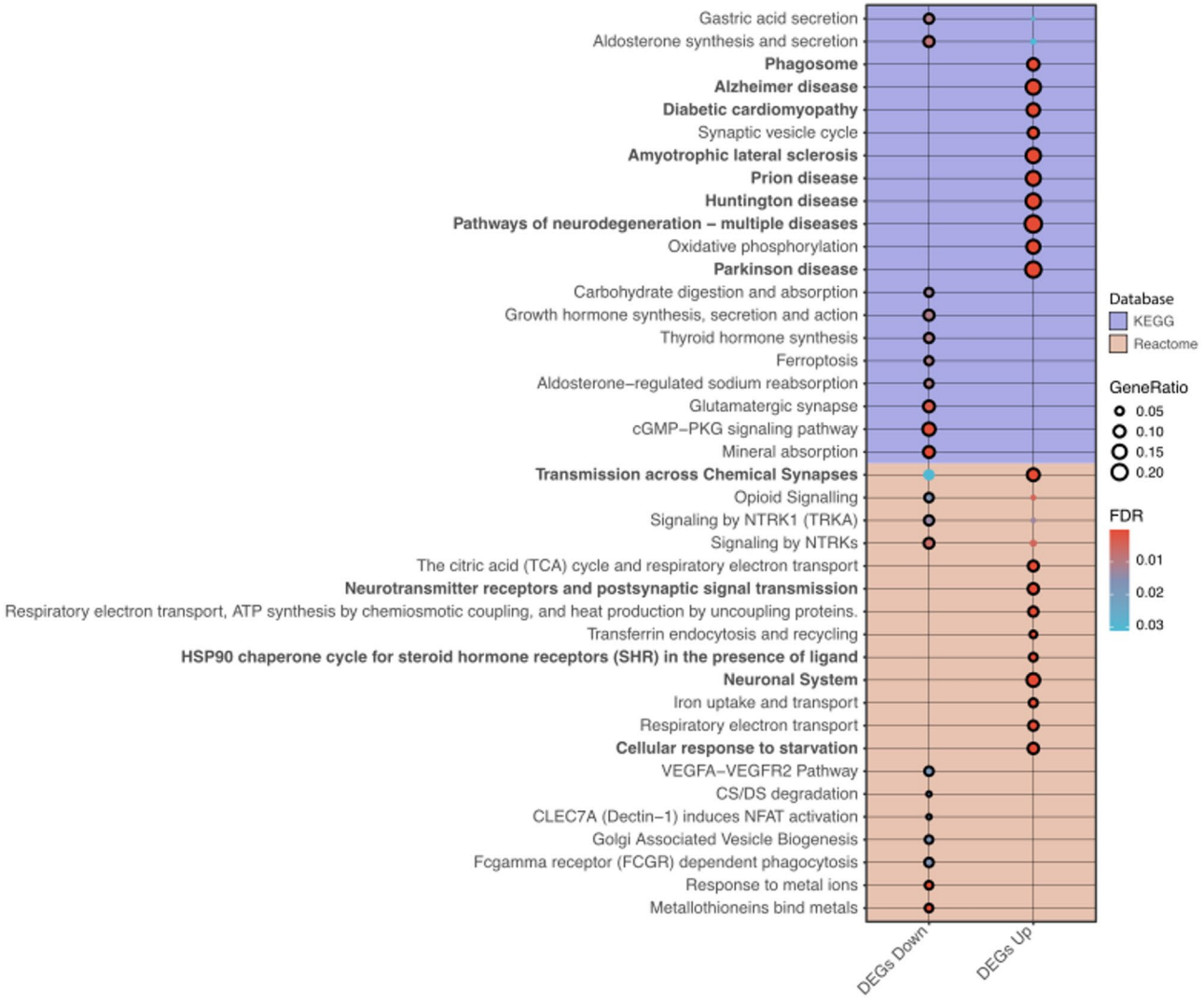

Fig. 4 Dot plot for the enrichment analysis top pathways. Enriched pathways in both KEGG and Reactome databases. The dot plot displays the pathways that were enriched using the up and downregulated DEGs as input. The dot size represents the pathways' gene ratio, while the color represents its P-value adjusted by False discovery rate. Background color represents the two different databases assessed for the enrichment, light blue represents KEGG pathways and light red represents reactome pathways. The Y and X axis displays the pathways and DEG set, separated by fold change direction (up or down compared with control), respectively. Pathway names written in bold letters have the 3 genes in its composition. Employed parameters in each enrichment were: Maximum gene set size of 500, minimum gene set size of 10, q-value cutoff of 0.2, p-value cutoff of 0.05, p-adjustment method "FDR", and fun was set as "enrichPathway" for Reactome and "enrichKEGG" for KEGG

receptor (FCGR) dependent phagocytosis", "Response to metal ions" and "Metallothioneins bind metals" (Fig. 4).

Discussion

This study employed a robust bioinformatics pipeline to identify consistent transcriptomic signatures of schizophrenia (SCZ) across multiple publicly available gene-expression datasets. This method of employing machine learning algorithms to identify transcriptomic biomarkers for characterizing diseases was used in previous studies with other diseases, such as Tuberculosis [12], HTLV-1 [13], and even Parkinson's disease [43]. In this study, the brain's prefrontal cortex's Brodmann areas 9, 10, and 46, were the tissues used as the discovery set, as these areas' dysfunction is a hallmark of SCZ [44, 45]. The predictive model based on random forest achieved AUCs > 0.70 in brain, blood, leukocytes and PBMC samples, exhibiting a consistently high accuracy even in more accessible types of tissue. Finally, this predictive model was composed of only three upregulated genes in SCZ: *HUWE1*, *PTGDS*, and *RPL31*.

The most informative gene, *HUWE1*, is suggested to play key roles as a conserved regulator in neuron progenitor cells' proliferation and differentiation [46]. Despite being located on the X-chromosome, this gene's expression levels are comparable between males and females

due to X-inactivation in females [47]In the SCZ context, *de novo* mutations in this gene were found in patients with this disease [48], which can be related to this gene's defective variations being associated with intellectual disability [49], a common symptom of SCZ. Unlike the other two genes, *RPL31* presents no previous associations with SCZ, although a previous study identified proteins coded by genes from the same family with lower abundance in SCZ compared to control in neural cells derived from olfactory neurosphere-derived (ONS) cells [50]. However, the results observed in the previous study cannot be directly compared with this analysis's results, due to methodological differences between both studies.

The third gene composing this model, *PTGDS*, may also present important associations with SCZ, being previously proposed as a cerebrospinal fluid biomarker for this disease and also for Alzheimer's [51]. This gene was selected as a biomarker in the brain datasets, despite being highly expressed in non-neuronal cell types [52], which suggests a strong influence of glial and other supporting cell types of cells present in the brain. Encoding for the protein prostaglandin D synthase, *PTGDS*, participates in prostaglandin metabolism, as it catalyzes the conversion of prostaglandin H2 to *PGD2*, a potent modulator of inflammation in the central nervous system. This molecule plays major roles in neuroinflammatory processes [53], being related to pathways enriched in KEGG, such as "Phagosome" and glial-associated pathways enriched in Reactome, including "Cellular response to starvation" and "Fcgamma receptor (FCGR)-dependent phagocytosis" [54].

The influence of glial and other supporting cell types was also observed in the lower performance validation analyses using isolated neurons, which achieved AUCs < 0.59. This suggests that isolated neurons lack the discriminative expression levels seen in whole-brain samples. In addition to *PTGDS*'s expression being more prominent in non-neuronal cell types, the reduced performance of our model in isolated neuron datasets may also reflect the absence of neuron-glia interactions. Neuron–glia communication plays a fundamental role in regulating neuronal activity and gene expression [55], and the lack of these interactions in isolated neuron samples could diminish the signal detected in bulk tissue analyses. The successful validation in whole blood, leukocytes, and PBMCs suggests that the biomarker's expression profile can cross the blood-brain barrier and be detected in peripheral circulation. Notably, the reduced classification performance in isolated lymphocytes implies the discriminative signal may not be lymphocyte-specific, but rather influenced by other leukocyte populations - a hypothesis supported by the stronger performance observed in leukocyte samples.

While most of the top informative genes we identified have been previously linked to SCZ, our enrichment analysis further supports their association with this disorder. Notably, several KEGG pathways—particularly those involving the random forest-selected genes—showed relevance to SCZ, with most being associated with neuronal disorders. Despite their dominant etiologic pathways and clinical presentations being different, both Alzheimer's disease and SCZ share the brain's white matter impairment and cognitive deficit as common features [56]. Furthermore, people with schizophrenia have 2–4 times higher risk of developing Alzheimer's than the general population [57]. Amyotrophic lateral sclerosis also has associations with SCZ, their common biological mechanisms being not well understood, however, a genetic loci correlation of 14.3% between both diseases has been observed [58]. Regarding Prion and Huntington's diseases, both diseases lack strong associations with SCZ, being mostly related due to the SCZ-like psychosis clinical feature [59, 60]. Lastly, SCZ is a risk factor for Parkinson's disease [61] and diabetes [62], although more studies are necessary to enlighten possible associations between these conditions.

Several enriched pathways containing random forest genes in the Reactome database may have associations with the SCZ's abnormal synaptic functions and neuronal plasticity [63], such as "Transmission across Chemical Synapses", "Neurotransmitter receptors and postsynaptic signal transmission", and "Neuronal System". Furthermore, the pathways enriched by downregulated DEGs related to Neurotrophins, such as signaling by NTRKs, are also related to survival, differentiation, and plasticity of neurons [64]. The NTRK1 (TRKA) is related to the maintenance of neuronal cells' integrity in adulthood [65]. The gene *HUWE1* may play a major role in these pathways, encoding for E3 ubiquitin ligase, as a regulator for synaptic protein turnover, neuronal differentiation, and neurotransmission, processes which dysregulation has been associated with SCZ [66].

Some enriched pathways containing upregulated DEGs were related to mitochondrial activity, such as "the citric acid (CA) cycle and respiratory electron transport", "Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins" and "Respiratory electron transport". The model's gene *RPL31* can also be related to these metabolic and energy-related pathways, as it encodes for a ribosomal protein, which participates in translational control and protein synthesis [67]. Furthermore, reduced number of mitochondria in several brain regions has been associated with the SCZ's pathophysiology before [68], perhaps the upregulated DEGs enriching these pathways reflect the lower number of mitochondria compensating for the lack of units with higher activity levels.

This study has some limitations; information regarding the samples' ethnicities is not available in the metadata

of several employed studies, possibly introducing a bias regarding ethnicity representation, which cannot be measured with the present data, once SCZ clinical presentation and risk vary across different ethnic and ancestral backgrounds [69, 70]. Covariates such as gender and age couldn't be included in the limma model used to identify the DEGs, as this model demands consistent metadata across all unavailable datasets. Another potential bias is related to the use of antipsychotics, as all datasets have samples from SCZ patients who use antipsychotics regularly; this probably creates a confounding factor as to whether the changes in their gene expression levels are related to the disease or the treatment. Another limitation is towards the results' causal interpretations, as this study's cross-sectional nature doesn't provide enough information to establish if the observed gene expression values are SCZ causes or consequences. To address this limitation, longitudinal studies that evaluate these gene expressions across time points would be necessary. Lastly, further studies are necessary to explore this set of genes' performance with other populations to further validate these genes and pathways' specific expression in SCZ.

Conclusions

The enrichment analysis revealed pathways with previous associations towards SCZ's pathophysiology in both KEGG and Reactome databases, which anchors this study's findings with the present knowledge towards SCZ. Our findings suggest that SCZ may involve systemic inflammatory processes rather than being centralized around neurons, which may potentially shift future research focus toward other non-neuron cell types. Although further studies are necessary to further validate this biomarker's performance and application, including in other populations, our findings pinpointed key genes and their associated pathways that are differentially expressed in SCZ, shedding light on potential disease's underlying mechanisms. This knowledge is crucial for potentially developing a new diagnostic test based on the expression levels of three genes, using easily accessible blood samples. This approach represents a significant improvement over current diagnostic strategies, enhancing the detection and management of SCZ.

Abbreviations

SCZ	Schizophrenia
DSM	Diagnostic and Statistical Manual of Mental Disorders
PFC	Prefrontal cortex
PBMC	Peripheral blood mononuclear cells
DEGs	differentially expressed genes
AUC	Area under the curve
C.I.	Confidence Interval

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12888-025-07524-6.

Supplementary Material 1

Acknowledgements

The authors would like to thank Adorielze Regina de Macedo Leite and Tatiana Maia for the support provided in this study.

Author contributions

R.R., M.A.P., E.R.F., B.B.A. and A.T.L.Q. performed the data curation, analysis, interpretation and draft of this manuscript's first version. T.F.M., H.Q.S.S., M.M.S.R. and M.A.P. performed the data interpretation, revising the manuscript critically for important intellectual content and final approval of the article.

Funding

This study was funded by the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) and Fundação Oswaldo Cruz (Fiocruz).

Data availability

All datasets present in this manuscript are publicly available at NCBI's GEO database (https://www.ncbi.nlm.nih.gov/geo/).

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 May 2025 / Accepted: 10 October 2025
Published online: 20 December 2025

References

1. Bighelli I, Rodolico A, García-Mieres H, Pitschel-Walz G, Hansen WP, Schneider-Thoma J, et al. Psychosocial and psychological interventions for relapse prevention in schizophrenia: a systematic review and network meta-analysis. Lancet Psychiatry. 2021;8(11):969–80.
2. Schizophrenia [Internet]. [cited 2024 Mar 5]. Available from: https://www.who.int/news-room/fact-sheets/detail/schizophrenia
3. Keshavan MS, Collin G, Guimond S, Kelly S, Prasad KM, Lizano P. Neuroimaging in schizophrenia. Neuroimaging Clin N Am. 2020;30(1):73–83.
4. Fischer BA, Carpenter WT Jr. Will the Kraepelinian dichotomy survive DSM-V? Neuropsychopharmacology. 2009;34(9):2081–7.
5. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. Nat Neurosci. 2016;19(11):1397–407.
6. Levin HS, Eisenberg HM, Benton AL. Frontal lobe function dysfunct. 1991. 460 p.
7. Karlsgodt KH, Sanz J, van Erp TGM, Bearden CE, Nuechterlein KH, Cannon TD. Re-evaluating dorsolateral prefrontal cortex activation during working memory in schizophrenia. Schizophr Res. 2009;108(1–3):143–50.
8. Miyahara K, Hino M, Shishido R, Nagaoka A, Izumi R, Hayashi H, et al. Identification of schizophrenia symptom-related gene modules by postmortem brain transcriptome analysis. Transl Psychiatry. 2023;13(1):144.
9. Song X, Liu Y, Pu J, Gui S, Zhong X, Chen X, et al. Transcriptomics analysis reveals shared pathways in peripheral blood mononuclear cells and brain tissues of patients with schizophrenia. Front Psychiatry. 2021;12:716722.
10. Lai CY, Scarr E, Udawela M, Everall I, Chen WJ, Dean B. Biomarkers in schizophrenia: A focus on blood based diagnostics and theranostics. World J Psychiatry. 2016;6(1):102–17.
11. Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sásik R, Khanlou N, et al. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci U S A. 2005;102(43):15533–8.

12. Rocha EF, Vinhaes CL, Araújo-Pereira M, Mota TF, Gupte AN, Kumar NP, et al. The sound of silent RNA in tuberculosis and the LncRNA role on infection. iScience. 2024;27(1):108662.
13. Fukutani ER, Ramos PIP, Kasprzykowski JI, Azevedo LG, Rodrigues MM, de Lima S, de OP JV, et al. Meta-Analysis of HTLV-1-Infected patients identifies CD40LG and GBP2 as markers of ATLL and HAM/TSP clinical status: two genes beat as one. Front Genet. 2019;10:1056.
14. Fukutani KF, Kasprzykowski JI, Paschoal AR, Gomes M, de Barral S, de Oliveira A. Meta-Analysis of Aedes aegypti expression datasets: comparing virus infection and Blood-Fed transcriptomes to identify markers of virus presence. Front Bioeng Biotechnol. 2017;5:84.
15. Subirana I, Sanz H, Vila J. Building bivariate tables: the comparegroups package for R. J Stat Softw. 2014;57:1–16.
16. Davis S, Meltzer PS. GEOquery: a Bridge between the gene expression omnibus (GEO) and bioconductor. Bioinformatics. 2007;23(14):1846–7.
17. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):559.
18. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics–a bioconductor package for quality assessment of microarray data. Bioinformatics. 2009;25(3):415–6.
19. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: a grammar of data manipulation [Internet]. CRAN: Contributed Packages. The R Foundation. 2014. Available from: https://cran.r-project.org/package=dplyr
20. Maycox PR, Kelly F, Taylor A, Bates S, Reid J, Logendra R, et al. Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. Mol Psychiatry. 2009;14(12):1083–94.
21. Hagenauer MH, Schulmann A, Li JZ, Vawter MP, Walsh DM, Thompson RC, et al. Inference of cell type content from human brain transcriptomic datasets illuminates the effects of age, manner of death, dissection, and psychiatric diagnosis. PLoS ONE. 2018;13(7):e0200003.
22. Lanz TA, Reinhart V, Sheehan MJ, Rizzo SJS, Bove SE, James LC, et al. Postmortem transcriptional profiling reveals widespread increase in inflammation in schizophrenia: a comparison of prefrontal cortex, striatum, and hippocampus among matched tetrads of controls with subjects diagnosed with schizophrenia, bipolar or major depressive disorder. Transl Psychiatry. 2019;9(1):151.
23. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The Sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics. 2012;28(6):882–3.
24. Tang B, Capitao C, Dean B, Thomas EA. Differential age- and disease-related effects on the expression of genes related to the arachidonic acid signaling pathway in schizophrenia. Psychiatry Res. 2012;196(2–3):201–6.
25. Abdolmaleky HM, Gower AC, Wong CK, Cox JW, Zhang X, Thiagalingam A, et al. Aberrant transcriptomes and DNA methylomes define pathways that drive pathogenesis and loss of brain laterality/asymmetry in schizophrenia and bipolar disorder. Am J Med Genet B Neuropsychiatr Genet. 2019;180(2):138–49.
26. de Jong S, Boks MPM, Fuller TF, Strengman E, Janson E, de Kovel CGF, et al. A gene co-expression network in whole blood of schizophrenia patients is independent of antipsychotic-use and enriched for brain-expressed genes. PLoS ONE. 2012;7(6):e39498.
27. van Beveren NJM, Buitendijk GHS, Swagemakers S, Krab LC, Röder C, de Haan L, et al. Marked reduction of AKT1 expression and deregulation of AKT1-associated pathways in peripheral blood mononuclear cells of schizophrenia patients. PLoS ONE. 2012;7(2):e32618.
28. Arion D, Huo Z, Enwright JF, Corradi JP, Tseng G, Lewis DA. Transcriptome alterations in prefrontal pyramidal cells distinguish schizophrenia from bipolar and major depressive disorders. Biol Psychiatry. 2017;82(8):594–600.
29. Gatta E, Saudagar V, Drnevich J, Forrest MP, Auta J, Clark LV, et al. Concordance of Immune-Related markers in lymphocytes and prefrontal cortex in schizophrenia. Schizophr Bull Open. 2021;2(1):sgab002.
30. Pai S, Li P, Killinger B, Marshall L, Jia P, Liao J, et al. Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. Nat Commun. 2019;10(1):2046.
31. Mendizabal I, Berto S, Usui N, Toriumi K, Chatterjee P, Douglas C, et al. Cell type-specific epigenetic links to schizophrenia risk in the brain. Genome Biol. 2019;20(1):135.
32. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
33. Liaw A, Wiener M. Classification and regression by randomforest [Internet]. 2002. Available from: https://CRAN.R-project.org/doc/Rnews/
34. Kuhn M. Building predictive models in R using the caret package. J Stat Softw. 2008;28:1–26.
35. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S + to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77.
36. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomart. Nat Protoc. 2009;4(8):1184–91.
37. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (Camb). 2021;2(3):100141.
38. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The reactome pathway knowledgebase 2024. Nucleic Acids Res. 2024;52(D1):D672–8.
39. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
40. Hoftman GD, Datta D, Lewis DA. Layer 3 excitatory and inhibitory circuitry in the prefrontal cortex: developmental trajectories and alterations in schizophrenia. Biol Psychiatry. 2017;81(10):862–73.
41. Jarskog LF, Glantz LA, Gilmore JH, Lieberman JA. Apoptotic mechanisms in the pathophysiology of schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry. 2005;29(5):846–58.
42. Gorman AM. Neuronal cell death in neurodegenerative diseases: recurring themes around protein handling. J Cell Mol Med. 2008;12(6A):2263–80.
43. Su C, Tong J, Wang F. Mining genetic and transcriptomic data using machine learning approaches in parkinson's disease. NPJ Parkinsons Dis. 2020;6:24.
44. Lewis DA, Glausier JR. Alterations in prefrontal cortical circuitry and cognitive dysfunction in schizophrenia. Nebr Symp Motiv. 2016;63:31–75.
45. Slifstein M, van de Giessen E, Van Snellenberg J, Thompson JL, Narendran R, Gil R, et al. Deficits in prefrontal cortical and extrastriatal dopamine release in schizophrenia: a positron emission tomographic functional magnetic resonance imaging study. JAMA Psychiatry. 2015;72(4):316–24.
46. Zhao X, Heng JIT, Guardavaccaro D, Jiang R, Pagano M, Guillemot F, et al. The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat Cell Biol. 2008;10(6):643–53.
47. Friez MJ, Brooks SS, Stevenson RE, Field M, Basehore MJ, Adès LC, et al. HUWE1 mutations in Juberg-Marsidi and Brooks syndromes: the results of an X-chromosome exome sequencing study. BMJ Open. 2016;6(4):e009537.
48. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De Novo mutations in schizophrenia implicate synaptic networks. Nature. 2014;506(7487):179–84.
49. Moortgat S, Berland S, Aukrust I, Maystadt I, Baker L, Benoit V, et al. HUWE1 variants cause dominant X-linked intellectual disability: a clinical study of 21 patients. Eur J Hum Genet. 2018;26(1):64–74.
50. English JA, Fan Y, Föcking M, Lopez LM, Hryniewiecka M, Wynne K, et al. Reduced protein synthesis in schizophrenia patient-derived olfactory cells. Transl Psychiatry. 2015;5(10):e663.
51. Davalieva K, Maleva Kostovska I, Dwork AJ. Proteomics research in schizophrenia. Front Cell Neurosci. 2016;10:18.
52. Dai Y, Hu R, Manuel AM, Liu A, Jia P, Zhao Z. CSEA-DB: an omnibus for human complex trait and cell type associations. Nucleic Acids Res. 2021;49(D1):D862–70.
53. Urade Y, Biochemical, Characteristics S, Regulation G. Physiological, pathological and clinical features of Lipocalin-Type prostaglandin D synthase as a multifunctional Lipocalin. Front Physiol. 2021;12:718002.
54. Fu R, Shen Q, Xu P, Luo JJ, Tang Y. Phagocytosis of microglia in the central nervous system diseases. Mol Neurobiol. 2014;49(3):1422–34.
55. Tremblay MÈ, Stevens B, Sierra A, Wake H, Bessis A, Nimmerjahn A. The role of microglia in the healthy brain. J Neurosci. 2011;31(45):16064–9.
56. Kochunov P, Zavaliangos-Petropulu A, Jahanshad N, Thompson PM, Ryan MC, Chiappelli J, et al. A white matter connection of schizophrenia and alzheimer's disease. Schizophr Bull. 2021;47(1):197–206.
57. Ribe AR, Laursen TM, Charles M, Katon W, Fenger-Grøn M, Davydow D, et al. Long-term risk of dementia in persons with schizophrenia: A Danish Population-Based cohort study. JAMA Psychiatry. 2015;72(11):1095–101.
58. McLaughlin RL, Schijven D, van Rheenen W, van Eijk KR, O'Brien M, Kahn RS, et al. Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun. 2017;8:14774.
59. Corrêa BB, Xavier M, Guimarães J. Association of huntington's disease and schizophrenia-like psychosis in a huntington's disease pedigree. Clin Pract Epidemiol Ment Health. 2006;2:1.

60. Thompson A, MacKay A, Rudge P, Lukic A, Porter MC, Lowe J, et al. Behavioral and psychiatric symptoms in prion disease. Am J Psychiatry. 2014;171(3):265–74.
61. Kuusimäki T, Al-Abdulrasul H, Kurki S, Hietala J, Hartikainen S, Koponen M, et al. Increased risk of parkinson's disease in patients with schizophrenia spectrum disorders. Mov Disord. 2021;36(6):1353–61.
62. Ali S, Santomauro D, Ferrari AJ, Charlson F. Schizophrenia as a risk factor for cardiovascular and metabolic health outcomes: a comparative risk assessment. Epidemiol Psychiatr Sci. 2023;32:e8.
63. Obi-Nagata K, Temma Y, Hayashi-Takagi A. Synaptic functions and their disruption in schizophrenia: from clinical evidence to synaptic optogenetics in an animal model. Proc Jpn Acad Ser B Phys Biol Sci. 2019;95(5):179–97.
64. Park H, Poo MM. Neurotrophin regulation of neural circuit development and function. Nat Rev Neurosci. 2013;14(1):7–23.
65. Marlin MC, Li G. Biogenesis and function of the NGF/TrkA signaling endosome. Int Rev Cell Mol Biol. 2015;314:239–57.
66. Luza S, Opazo CM, Bousman CA, Pantelis C, Bush AI, Everall IP. The ubiquitin proteasome system and schizophrenia. Lancet Psychiatry. 2020;7(6):528–37.
67. Surya A, Bolton BM, Rothe R, Mejia-Trujillo R, Leonita A, Zhao Q, et al. Differential impacts of ribosomal protein haploinsufficiency on mitochondrial function. J Cell Biol [Internet]. 2025;224(3). Available from: https://doi.org/10.1083/jcb.202404084
68. Roberts RC. Mitochondrial dysfunction in schizophrenia: with a focus on postmortem studies. Mitochondrion. 2021;56:91–101.
69. Wang D, Wang Y, Chen Y, Yu L, Fang X, Liu R, et al. Ethnicity-dependent effects of zinc finger 804A variant on schizophrenia: a systematic review and meta-analysis. Psychiatr Genet. 2021;31(1):21–8.
70. Miyahara K, Hino M, Shishido R, Izumi R, Nagaoka A, Hayashi H, et al. Ethnicity-dependent effect of rs1799971 polymorphism on OPRM1 expression in the postmortem brain and responsiveness to antipsychotics. J Psychiatr Res. 2023;166:10–6.

Publisher's note

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