



Mestrado Profissional em
Tecnologias em Saúde

ESCOLA BAHIANA DE MEDICINA E SAÚDE PÚBLICA
PROGRAMA DE PÓS-GRADUAÇÃO EM TECNOLOGIAS EM SAÚDE

REBECA REBOUÇAS DA CUNHA SILVA

**APLICAÇÃO DE *MACHINE LEARNING* NA CARACTERIZAÇÃO DE FATORES
TRANSCRICIONAIS E VIAS MOLECULARES ASSOCIADAS À ESQUIZOFRENIA**

DISSERTAÇÃO DE MESTRADO

Salvador

2025

REBECA REBOUÇAS DA CUNHA SILVA

APLICAÇÃO DE *MACHINE LEARNING* NA CARACTERIZAÇÃO DE FATORES
TRANSCRICIONAIS E VIAS MOLECULARES ASSOCIADAS À ESQUIZOFRENIA

Dissertação apresentada à Escola Bahiana de Medicina e Saúde Pública, como parte das exigências do Programa de Pós-Graduação em Tecnologias em Saúde, para obtenção do título de mestre.

Orientador: Dr. Bruno de Bezerril Andrade.
Coorientador: Dr. Artur Lopo Trancoso de Queiroz.

Salvador

2025

Ficha Catalográfica elaborada pelo Sistema Integrado de Bibliotecas

S586 Silva, Rebeca Rebouças da Cunha
Aplicação de machine learning na caracterização de fatores transcricionais e vias moleculares associadas à esquizofrenia. / Rebeca Rebouças da Cunha Silva – 2025. 68f.: 30cm.

Orientador: Prof.º Dr. Bruno de Bezerril Andrade
Coorientador: Prof.º Dr. Artur Lopo Trancoso de Queiróz

Dissertação (Programa de Pós-graduação em Tecnologias em Saúde) Escola Bahiana de Medicina e Saúde Pública – EBMSP, 2025.

Inclui bibliografia

1. Esquizofrenia. 2. Biomarcadores. 3. Doenças neurodegenerativas. I. Andrade, Bruno de Bezerril. II. Título.

CDU: 616.895.8

REBECA REBOUÇAS DA CUNHA SILVA

**"APLICAÇÃO DE MACHINE LEARNING NA CARACTERIZAÇÃO DE FATORES
TRANSCRICIONAIS E VIAS MOLECULARES ASSOCIADAS À
ESQUIZOFRENIA."**

Dissertação apresentada à Escola Bahiana
de Medicina e Saúde Pública, como
requisito parcial para a obtenção do Título
de Mestre em Tecnologias em Saúde.

Salvador, 11 de novembro de 2025.

BANCA EXAMINADORA

Dr. Cassio Dos Santos Lima
Doutorado em Psicologia
Universidade Federal da Bahia

Dra. Viviane Schuch
Doutorado em Educação e Saúde na Infância e Adolescência.
Universidade Federal de São Paulo

Dr. Ícaro Santos Lopes
Doutorado em Bioinformática
Universidade Federal de Minas Gerais

RESUMO

Introdução: A esquizofrenia é um transtorno neuropsiquiátrico de etiologia ainda pouco compreendida, caracterizado por disfunções comportamentais, emocionais e cognitivas. Embora avanços tenham sido feitos na identificação de fatores associados à doença, os mecanismos moleculares permanecem pouco definidos. Este trabalho caracterizou fatores transcricionais e vias moleculares associadas à esquizofrenia, com o objetivo de ampliar a compreensão de sua base biológica. **Métodos:** Foram analisados dados públicos de expressão genética, contendo amostras do córtex pré-frontal do cérebro (*post mortem*) e do sangue periférico. Após critérios de elegibilidade e verificações de qualidade, 17 conjuntos de dados foram incluídos, provenientes de seis tipos de amostra diferentes: tecido cerebral (n = 6), neurônios isolados (n = 5), sangue total (n = 2), células mononucleares do sangue periférico (n = 2), leucócitos (n = 1) e linfócitos (n = 1). Quatro conjuntos de dados cerebrais foram definidos como conjunto de descoberta. Genes diferencialmente expressos (DEGs) foram identificados comparando pacientes com esquizofrenia e controle, sendo posteriormente usados na análise de enriquecimento funcional. Por fim, foi aplicada técnicas de seleção de características para identificar os DEGs mais informativos para a classificação de esquizofrenia, e posteriormente foi avaliada sua performance em dados de outros tecidos. **Resultados:** Foram identificados 532 DEGs. A aplicação de *machine learning* revelou três genes (*HUWE1*, *PTGDS* e *RPL31*) como os mais informativos para discriminar eficazmente indivíduos com esquizofrenia e controle. O desempenho do modelo foi validado em outros tecidos, apresentando AUC (>72%) em cérebro, sangue total, PMBC e leucócitos. A análise de enriquecimento revelou um potencial ligação com vias biológicas neurodegenerativas. **Conclusão:** A integração de dados de expressão gênica com ferramentas de seleção de características permitiu a identificação de marcadores transcricionais consistentes da esquizofrenia em diferentes tipos de tecidos. Esses achados permitiram expandir o conhecimento da fisiopatologia da esquizofrenia e para a compreensão dos processos biológicos compartilhados entre distúrbios neuropsiquiátricos e neurodegenerativos.

Palavras-chave: Esquizofrenia. Biomarcadores. Doenças neurodegenerativas.

ABSTRACT

Introduction: Schizophrenia is a neuropsychiatric disorder of poorly understood etiology, characterized by behavioral, emotional, and cognitive dysfunction. Although advances have been made in identifying factors associated with the disease, the molecular mechanisms remain poorly defined. This work characterized transcriptional factors and molecular pathways associated with schizophrenia, aiming to broaden the understanding of its biological basis. **Methods:** Public gene expression data were analyzed, including samples from the prefrontal cortex of the brain (*postmortem*) and peripheral blood. After eligibility criteria and quality checks, 17 datasets were included, originating from six different sample types: brain tissue (n = 6), isolated neurons (n = 5), whole blood (n = 2), peripheral blood mononuclear cells (n = 2), leukocytes (n = 1), and lymphocytes (n = 1). Four brain datasets were defined as the discovery set. Differentially expressed genes (DEGs) were identified by comparing schizophrenia patients and controls and were subsequently used in functional enrichment analysis. Finally, feature selection techniques were applied to identify the most informative DEGs for schizophrenia classification, and their performance was subsequently evaluated in data from other tissues. **Results:** 532 DEGs were identified. Machine learning revealed three genes (*HUWE1*, *PTGDS*, and *RPL31*) as the most informative for effectively discriminating individuals with schizophrenia from controls. The model's performance was validated in other tissues, showing AUC (>72%) in the brain, whole blood, PMBC, and leukocytes. Enrichment analysis revealed a potential link with neurodegenerative biological pathways. **Conclusion:** The integration of gene expression data with feature selection tools allowed the identification of consistent transcriptional markers of schizophrenia in different tissue types. These findings have expanded our knowledge of the pathophysiology of schizophrenia and the understanding of the biological processes shared between neuropsychiatric and neurodegenerative disorders.

Keywords: Schizophrenia. Biomarkers. Neurodegenerative diseases.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma de amostragem	23
Figura 2 – Componentes principais do conjunto de descobertas antes e depois da mitigação do efeito de lote	29
Figura 3 – Identificação dos genes mais informativos e avaliação do desempenho discriminativo	31
Figura 4 – Componentes principais de cada conjunto de validação composto por mais de um conjunto de dados antes e depois da mitigação do efeito de lote	32
Figura 5 – Diagrama de pontos para as principais vias da análise de enriquecimento	35

LISTA DE TABELAS

Tabela 1 – Caracterização dos conjuntos de dados de validação	33
--	----

LISTA DE ABREVIATURAS E SIGLAS

CA – Ácido Cítrico

CPF – Córtex Pré-Frontal

CPs – Componentes Principais

DEGs – Genes Diferencialmente Expressos

DSM – Manual Diagnóstico e Estatístico de Transtornos Mentais

FCGR – Receptor Fc gama

HTLV-1 – Vírus Linfotrópico de Células T Humanas

IC – Intervalo de Confiança

KEGG – *Kyoto Encyclopedia of Genes and Genomes*

LAI – Antipsicótico de Longa Ação

NCBI – *National Center for Biotechnology Information*

ONS – *Human Olfactory Neurosphere-derived*

PBMCs – Células Mononucleares do Sangue Periférico

PCA – Análise de Componentes Principais

ROC – *Receiver Operating Characteristic*

SCZ – Esquizofrenia

SHR – Receptores de Hormônios Esteróides

TCA – Ciclo do ácido tri carboxílico

TEPT – Transtorno de Estresse Pós-Traumático

SUMÁRIO

1 INTRODUÇÃO	9
2 REFERENCIAL TEÓRICO	10
2.1 Esquizofrenia: uma visão geral.....	10
2.2 Etimologia	11
2.3 Diagnóstico clínico da esquizofrenia	12
2.3.1 Alterações históricas no diagnóstico clínico	16
2.4 Tratamento	17
3 JUSTIFICATIVA	20
4 OBJETIVOS	21
4.1 Objetivo geral	21
4.2 Objetivos específicos.....	21
5 METODOLOGIA.....	22
5.1 Estratégia de pesquisa e critérios de seleção	22
5.2 Análise dos dados.....	24
5.2.1 Desenho da análise e correção do efeito de lote	24
5.2.2 Seleção de recursos: análise de expressão diferencial e machine learning	25
5.2.3 Identificação das vias biológicas mais representadas: análise de enriquecimento.....	26
6 RESULTADOS	28
6.1 Desenho do estudo e descrição da população	28
6.2 Expressão diferencial e seleção de características para representar a esquizofrenia	28
6.3 As vias associadas aos DEGs da esquizofrenia também estão associadas a outros distúrbios neuronais e psíquicos	33
7 DISCUSSÃO	36
8 CONCLUSÃO.....	40
REFERÊNCIAS.....	41
ANEXO.....	56

1 INTRODUÇÃO

A esquizofrenia (SCZ) é um transtorno neuropsiquiátrico grave, multifatorial, hereditário, e de etiologia desconhecida. Este transtorno afeta o desempenho psicossocial de aproximadamente 24 milhões de pessoas em todo o mundo, com uma prevalência global estimada entre 0,3% e 0,7% da população (1–3).

O início dos sintomas psicóticos ocorre comumente no final da adolescência e início da vida adulta (3). Em homens, os primeiros episódios geralmente acontecem entre os 15 e 25 anos, enquanto nas mulheres tende se apresentar um pouco mais tarde, por volta dos 25 a 30 anos (4,5). Os primeiros sinais da esquizofrenia frequentemente se desenvolvem de maneira progressiva e incluem isolamento social, alterações afetivas e comprometimento cognitivo. Em cerca de metade dos casos, também são observados sintomas depressivos, podendo ser um agravamento no prognóstico (3).

A esquizofrenia representa uma das 20 principais causas de incapacidade no mundo, demonstrando a complexidade das implicações socioeconômicas do transtorno e o impacto direto na vida dos indivíduos acometidos (6,7). Além disso, observa-se uma redução na expectativa de vida, estimada entre 15 a 20 anos a menos em relação à população geral, o que reforça a relevância do transtorno como um desafio de saúde pública (3,8–13).

Não existem exames laboratoriais ou radiológicos específicos para seu diagnóstico, que segue critérios clínicos (3,14). O tratamento é baseado em intervenções farmacológicas, sobretudo antipsicóticos, e intervenções psicossociais, como a Terapia Cognitivo-Comportamental (TCC), que juntas buscam controle sintomático, prevenção de recaídas e promoção da funcionalidade (15–24).

2 REFERENCIAL TEÓRICO

2.1 Esquizofrenia: uma visão geral

Os primeiros sinais da esquizofrenia frequentemente se desenvolvem de maneira progressiva e incluem isolamento social, alterações afetivas e comprometimento cognitivo. Em cerca de metade dos casos, também são observados sintomas depressivos, podendo ser um agravamento no prognóstico (25).

A prevalência da esquizofrenia por sexo varia de acordo com os critérios diagnósticos adotados (25). Homens tendem a apresentar sintomas negativos, episódios mais prolongados e prognóstico desfavorável. No entanto, definições diagnósticas mais amplas, que incluem sintomas afetivos e apresentações de início breve, tendem a indicar riscos proporcionais entre os sexos (25). As características relativas ao sexo e ao gênero também se manifestam em aspectos hormonais e neurobiológicos. Nas mulheres, os sintomas psicóticos podem oscilar em função das variações hormonais ao longo do ciclo menstrual, por exemplo, tendendo a piorar no período pré-menstrual e no puerpério, quando os níveis de estrogênio estão mais baixos. Em contrapartida, durante a gestação, os níveis elevados desse hormônio parecem exercer um efeito protetor (25).

Embora o curso da esquizofrenia varie significativamente entre cada indivíduo, é comum a persistência de sintomas ao longo da vida (4). Em alguns casos, é possível apresentar remissão parcial ou completa, enquanto outros casos seguem com quadros crônicos, marcados por recaídas e perdas funcionais frequentes (4).

O comportamento suicida representa uma preocupação significativa no manejo da esquizofrenia (26–28). Aproximadamente 5% a 6% dos indivíduos diagnosticados com esquizofrenia chegam a cometer suicídio, enquanto 20% tentam ou idealizam ao longo da vida (25,27,29). Fatores de riscos incluem depressão comórbida, desesperança, desemprego, uso de substâncias, múltiplas hospitalizações, início tardio da doença e histórico familiar de transtornos psiquiátricos (27,28,30). Em homens mais jovens, especialmente aqueles com elevada funcionalidade pré-mórbida, o risco de suicídio ainda é mais acentuado (28,30). Acrescente-se, ainda, que o comportamento suicida pode também estar diretamente relacionado a delírios ou alucinações com conteúdo autodestrutivo (25,28).

Além dos prejuízos cognitivos e emocionais, é importante destacar os prejuízos na escolarização, na inserção no mercado de trabalho e na manutenção das relações

sociais (31–33). Há uma elevada prevalência por uso de substâncias, sendo o tabagismo presente em mais da metade dos indivíduos com esquizofrenia (10,12,26). No âmbito da saúde física, destaca-se o aumento da incidência de obesidade (26,34–36), diabetes tipo 2 (26,37–39), hipertensão (26), síndrome metabólica e doenças cardiovasculares (40–42). Essas comorbidades decorrem de múltiplos fatores, incluindo uso de antipsicóticos, estilo de vida sedentário, alimentação inadequada, além do tabagismo mencionado anteriormente (42). A integralidade desses elementos resulta em uma expectativa de vida significativamente reduzida (9,26).

Estudos de neuroimagem, neuropatologia e neurofisiologia têm demonstrado diferenças consistentes em diversas regiões cerebrais entre indivíduos com esquizofrenia e indivíduos sem o transtorno, como alterações celular, na conectividade da substância branca e no volume da substância cinzenta em áreas como os córtex pré-frontal e temporal. Observa-se, ainda, uma redução do volume cerebral, a qual se acentua com o envelhecimento de forma mais destacada em indivíduos com esquizofrenia quando comparadas em indivíduos sem esquizofrenia (14,25,43,44). Além dessas alterações estruturais e funcionais, também foram identificadas diferenças em índices eletrofisiológicos e em medidas de rastreamento ocular, além de prejuízos na coordenação motora, na percepção do próprio corpo e do meio e na execução de movimentos complexos (25).

2.2 Etimologia

Desde as primeiras tentativas de descrever estados mentais patológicos até as formulações mais elaboradas do século XX, o entendimento da esquizofrenia passou por grandes transformações. Em 1853, o psiquiatra franco-austríaco Benedict Morel introduziu pela primeira vez o termo *démence précoce* (demência precoce) para descrever um conjunto de sintomas caracterizados por alterações de pensamentos, emoções e comportamentos desorganizados ou inadequados, com início precoce, comumente na adolescência. Morel associava tais quadros a uma degeneração hereditária progressiva e inevitável. Sua concepção estava inserida no contexto do pensamento degeneracionista do século XIX, influenciado por teorias biológicas e morais sobre o declínio das funções mentais (45,46).

No final do século 19, o psiquiatra alemão Emil Kraepelin refinou e consolidou o entendimento de “demência precoce” ao classificá-la como um quadro psicopatológico distinto, de início precoce e evolução crônica e com prognóstico

geralmente desfavorável. Por meio de observações longitudinais de seus pacientes, Kraepelin estabeleceu uma diferenciação entre a demência precoce e a psicose maníaco-depressiva, esta última com curso episódico e melhor prognóstico. A contribuição de Kraepelin foi fundamental para a consolidação do modelo médico descritivo de diagnóstico em psiquiatria, baseado não apenas na sintomatologia aguda, mas também na evolução da doença (32,47–49).

Já no início do século seguinte, o psiquiatra suíço Eugen Bleuler propôs uma mudança de paradigma ao introduzir o termo “esquizofrenia” (do grego, *schizo*=divisão; *phrene*=mente), para descrever a dissociação entre pensamento, emoção e comportamento. Para Bleuler, a esquizofrenia não deveria ser encarada com uma forma de demência no sentido literal, mas sim como uma desordem complexa e heterogênea da psique, cuja evolução da doença variava significativamente de indivíduo para indivíduo. Bleuler também foi responsável por classificar os sintomas fundamentais da esquizofrenia que, segundo o psiquiatra, estariam presentes mesmo na ausência de delírios ou alucinações. Essa nova abordagem permitiu uma ampliação na compreensão do diagnóstico e prognóstico da doença, abrindo um novo caminho para uma visão integrativa, menos determinista e mais sensível às particularidades de cada caso (32,45,49,50).

2.3 Diagnóstico clínico da esquizofrenia

O diagnóstico da esquizofrenia envolve um processo longo e cuidadoso por parte dos profissionais de saúde mental. A exclusão de outras condições psiquiátricas é um desafio por si só, visto que seus sintomas são comuns em outros transtornos, como depressão maior, transtorno obsessivo-compulsivo, transtorno de estresse pós-traumático e transtorno bipolar (25,51). Esse processo envolve uma abordagem abrangente, incluindo uma avaliação psiquiátrica e psicológica detalhada pautada em sistemas classificatórios padronizados, especialmente o *Diagnostic and Statistical Manual of Mental Disorders* (DSM), atualmente em sua 5ª edição revisada (DSM-5-TR), e a Classificação Internacional de Doenças, em sua décima revisão (CID-10), que permanece vigente no Brasil até a implementação definitiva da CID-11, prevista para 2027 (25,52).

De acordo com o DSM-5-TR, a esquizofrenia é classificada como um transtorno psicótico e é necessário que, durante um período mínimo de um mês, o indivíduo

apresente pelo menos dois sintomas positivos (delírios, alucinações, discurso desorganizado) (25). Além desses, também são considerados o comportamento disfuncional ou catatônico e os sintomas negativos, como a diminuição da expressão emocional, empobrecimento do repertório verbal (alogia) e a perda de motivação ou iniciativa (avolia) (32). Os delírios, por exemplo, podem envolver desde ideias persecutórias até crenças de grandeza (4,25,32,53). Já as alucinações ocorrem mais frequentemente de forma auditiva, com relatos típicos de vozes conversando em terceira pessoa (4,33,54). O discurso desorganizado pode se manifestar por meio de respostas incoerentes ou associações fracas entre as ideias, dificultando a compreensão (4). Vale destacar que, mesmo nos casos em que há melhora clínica ou remissão parcial dos sintomas da fase ativa, o diagnóstico do transtorno permanece válido se os critérios forem atendidos previamente, desde que os sintomas tenham persistido na ausência de intervenções terapêuticas eficazes. Essa característica reforça o caráter episódico e muitas vezes crônico da esquizofrenia, ressaltando a importância de um acompanhamento longitudinal no intuito de alcançar uma abordagem diagnóstica e terapêutica adequada (25).

Além da manifestação dos sintomas positivos descritos, o diagnóstico da esquizofrenia também requer, conforme os critérios de prejuízo social e ocupacional do DSM-5-TR, a identificação de danos significativos na funcionalidade (25). Tais prejuízos podem refletir em diferentes áreas da vida do indivíduo, como o desempenho acadêmico, a produtividade no trabalho ou na qualidade das relações interpessoais (25,31,33). Quando o transtorno se inicia na infância ou adolescência, muitas vezes é possível observar uma interrupção ou regressão nos níveis esperados de desenvolvimento, sendo comum que o indivíduo apresente dificuldades em acompanhar o rendimento escolar ou estabelecer vínculos sociais compatíveis com a sua faixa etária (25,32,55). A permanência dessas desordens durante o curso da doença pode ser agravada pela presença de sintomas negativos, como o comprometimento na capacidade de iniciar e manter comportamentos dirigidos a metas, prejudicando a execução de atividades cotidianas, como manter uma rotina de estudos ou comparecimento no trabalho (33). Além disso, as implicações cognitivas, sobretudo nas áreas de atenção, memória e funções executivas, contribui de forma relevante no declínio funcional. Essas limitações, somadas ao impacto dos sintomas positivos e negativos, tornam a esquizofrenia um transtorno que implica de maneira crônica a autonomia e a qualidade de vida dos indivíduos acometidos (25).

Adicionalmente, é necessário que alguns sintomas estejam presentes de forma contínua por um período mínimo de seis meses, incluindo pelo menos uma fase ativa com sintomas psicóticos. Frequentemente, a fase ativa é antecedida por um estágio conhecido como fase prodrômica (25,56). Nessa fase são comuns sintomas como retraimento social, queda no desempenho escolar ou profissional, além de alterações no sono e no humor (4,32).

Sintomas prodrômicos também podem incluir o surgimento de ideias de referência, em que o indivíduo começa a interpretar eventos cotidianos, como comentários alheios ou programas de televisão, como tendo um significado especial ou direcionado (54). Outras manifestações incluem o desenvolvimento de crenças incomuns ou pensamentos mágicos, como acreditar que consegue influenciar o pensamento de outras pessoas apenas com a mente ou interpretar coincidências banais como sinais com significado especial (25,56).

Além disso, podem ocorrer experiências perceptivas disfuncionais, como ouvir sons desconhecidos, sentir a presença de alguém invisível ou perceber alterações nos ambientes sem que haja uma causa objetiva (57). Durante essa fase, o discurso do indivíduo pode se tornar vago, com dificuldades de manter o foco e clareza nas palavras. O comportamento, embora não necessariamente de forma grosseira, pode parecer excêntrico ou inadequado ao contexto, como rir sem motivo aparente ou falar sozinho em público (57).

Após a fase ativa da doença, muitos indivíduos evoluem para uma fase residual, caracterizada pela persistência de sintomas em grau subclínico, como traços leves de delírios ou alucinações, além dos sintomas negativos (57,58). É comum que indivíduos antes socialmente ativos, apresentem um declínio contínuo em suas interações e nas atividades que antes desempenhavam normalmente, tornando o isolamento e a perda de interesse pelas atividades corriqueiras um dos primeiros sinais observáveis do transtorno (25).

A Classificação Internacional de Doenças (CID-10), por sua vez, oferece um ponto de vista clínico mais fenomenológico centrado nas observações dos sintomas e na descrição dos diferentes modos de expressão da esquizofrenia. Sob o código F20, a esquizofrenia é definida como um grupo de transtornos caracterizados por distorções do pensamento e da percepção, além de afetos inapropriados ou embotados. Geralmente, a consciência e a capacidade intelectual permanecem relativamente preservadas, embora prejuízos cognitivos possam agravar ao longo do

tempo. Entre os sintomas psicopatológicos mais característicos, a CID-10 destaca o eco do pensamento, a imposição ou roubo do pensamento, bem como vozes alucinatórias que comentam ou discutem em terceira pessoa. Também são observados transtornos do pensamento e sintomas negativos, como diminuição da expressão emocional, alogia e avolia (59).

O curso clínico da esquizofrenia, segundo a CID-10, pode ser contínuo, episódico com prejuízos progressivos ou estáveis, ou apresentar um ou mais episódios seguidos de remissão completa ou parcial. É importante destacar que para a CID-10, o diagnóstico não deve ser definido quando há presença de sintomas depressivos ou maníacos, a menos que os sintomas esquizofrênicos tenham claramente antecedido a mudança afetiva. Nesse sentido, também não se deve diagnosticar a esquizofrenia quando os sintomas decorrem de uma doença cerebral ou intoxicação por substâncias psicoativas (60).

A CID-10, diferente do DSM-5-TR, propõe uma subdivisão da esquizofrenia em diferentes formas clínicas de acordo com a predominância dos sintomas. A esquizofrenia paranoide (F20.0) é representada por delírios e alucinações persistentes, geralmente auditivas, que tendem permanecer ao longo do tempo, com menor comprometimento do afeto e da organização do pensamento. Já a esquizofrenia hebefrênica (F20.1) apresenta um quadro distinto, caracterizado pela presença de afeto superficial ou inadequado, comportamento imprevisível e discurso incoerente, refletindo uma desorganização do funcionamento psíquico e social. A esquizofrenia catatônica (F20.2) é definida pelo predomínio de alterações psicomotoras, que podem variar entre estupor e hiperinesia, incluindo episódios de agitação intensa, rigidez postural e negativismo. Quando o quadro clínico reúne características de diferentes subtipos sem um predomínio claro, classifica-se como esquizofrenia indiferenciada (F20.3), que abrange apresentações mistas ou atípicas dentro do espectro esquizofrênico (61).

Além dessas formas clássicas, a CID-10 descreve subtipos que refletem fases evolutivas da doença. A depressão pós-esquizofrênica (F20.4) ocorre quando um episódio depressivo importante se instala após uma fase esquizofrênica ativa, mantendo ainda sintomas residuais do transtorno. Já a esquizofrenia residual (F20.5) representa uma fase crônica, em que os sintomas negativos persistem. E, por fim, a esquizofrenia simples (F20.6) marcada por um declínio gradual e progressivo do

funcionamento social, afetivo e ocupacional, sem que o indivíduo apresente episódios psicóticos claramente definidos (61).

Ainda que o DSM-5-TR e a CID-10 apresentem abordagens diferentes, ambos complementam na prática clínica. Enquanto o Manual Diagnóstico e Estatístico de Transtornos Mentais (DSM) contribui com uma lógica dimensional e funcional, além da compreensão teórica e padronização internacional da pesquisa, a Classificação Internacional de Doenças (CID) oferece uma perspectiva descritiva e classificatória da esquizofrenia, orientando oficialmente os registros diagnósticos e epidemiológicos. Essa complementaridade entre as duas diretrizes é fundamental, sobretudo no contexto clínico, em que a CID-10 permanece como referência normativa.

Compreender ambos os sistemas classificatórios permitem não apenas alcançar maior precisão diagnóstica, mas também desenvolver uma perspectiva mais ampla e humanizada da experiência psicótica, articulando o conhecimento técnico às singularidades e ao contexto cultural no qual o indivíduo está inserido (62).

2.3.1 Alterações históricas no diagnóstico clínico

Embora os critérios de diagnóstico clínico da esquizofrenia mencionados acima são os estabelecidos nas edições do DSM-5-TR e a CID-10, vale destacar que os dados públicos usados nas análises deste trabalho foram originalmente coletados de estudos anteriores, nos quais o diagnóstico dos participantes com esquizofrenia foi estabelecido segundo as versões anteriores do manual diagnóstico, o DSM-3 e DSM-4. A compreensão da evolução desses critérios é fundamental para contextualizar as condições sob as quais esses participantes foram diagnosticados, e assim, compreender as características clínicas representadas nos dados analisados.

O DSM-3, publicado em 1980, representou um marco importante na psiquiatria ao propor critérios diagnósticos mais padronizados, buscando maior validade e confiabilidade entre os profissionais de saúde mental (60). Nesta edição, o diagnóstico da esquizofrenia se baseava na presença de sintomas psicóticos característicos, declínio funcional e critérios bem definidos de duração e exclusão diagnóstica. Inicialmente, era exigida a apresentação de pelo menos um sintoma da fase ativa do transtorno, como delírios, alucinações auditivas ou alteração de pensamento, como incoerência ou empobrecimento verbal, desde que associados à delírios, comportamento desorganizado e paritímia.

Além disso, era necessário que o indivíduo apresentasse um comprometimento funcional significativo nas áreas ocupacional, social e de autocuidado, quando comparado ao seu nível de funcionalidade anterior ao surgimento dos primeiros sintomas. O quadro clínico deveria se estender por no mínimo seis meses, incluindo obrigatoriamente uma fase ativa com sintomas psicóticos, podendo abranger fases prodrômicas e/ou residuais, marcadas por isolamento social, prejuízo no desempenho de atividades, afeto embotado e experiências perceptivas atípicas. Importante destacar que outros critérios exigidos incluíam início dos sintomas antes dos 45 anos e exclusão de transtornos afetivos primários, transtornos mentais orgânicos ou deficiência intelectual (61). Assim, o DSM-3 consolidou uma abordagem diagnóstica baseada na combinação de critérios sintomatológicos, funcionais e temporais.

Posteriormente, o DSM-4, publicado em 1994, manteve a estrutura diagnóstica introduzida pelo manual anterior, mas com aprimoramentos conceituais. Para o diagnóstico da esquizofrenia, o DSM-4 exigia a presença de pelo menos dois (ou mais) dos seguintes sintomas da fase ativa, sendo cada um presente durante um período de um mês: delírios, alucinações, discurso e comportamento desorganizado ou catatônico, e sintomas negativos, como embotamento afetivo, alogia e avolia.

Contudo, se delírios fossem considerados bizarros ou se houvesse alucinações consistindo em vozes que comentam ou conversam entre si, a presença de apenas um desses sintomas já era suficiente. Assim como no DSM-III, era necessário que o indivíduo apresentasse um comprometimento funcional significativo nas áreas ocupacional e social, no entanto, com duração mínima de seis meses de perturbação, incluindo ao menos um mês de sintomas ativos. O DSM-4 também reafirmava a necessidade de excluir transtornos de humor com sintomas psicóticos relevantes, transtorno esquizoafetivo e condições médicas ou uso de substâncias que pudessem justificar o quadro. Além disso, houve um cuidado maior na diferenciação entre esquizofrenia e transtornos do espectro autista, especialmente em casos de início precoce, exigindo alucinações ou delírios por pelo menos um mês para que o diagnóstico fosse considerado (62,63).

2.4 Tratamento

O tratamento da esquizofrenia permanece como um dos maiores desafios em saúde mental, tanto pela complexidade clínica da doença quanto pelas limitações terapêuticas disponíveis, sendo a farmacoterapia, principalmente os antipsicóticos, o

pilar central do manejo clínico (15,64). Desde a introdução do primeiro antipsicótico na década de 50, a clorpromazina, essa classe de fármacos se consolidou como a principal intervenção farmacológica, transformando o manejo do transtorno e contribuindo para uma redução significativa das internações psiquiátricas, além de possibilitar que muitos indivíduos retornassem ao convívio social com maior autonomia (17,64). Apesar disso, a esquizofrenia continua sendo uma doença crônica que exige manejo contínuo, sem perspectiva de cura, mas com possibilidade de controle sintomático e prevenção de recaídas (18,19).

Atualmente, diretrizes terapêuticas recomendam o uso de antipsicóticos, escolhidos conforme o perfil de segurança e tolerabilidade do indivíduo (6,65). Em casos refratários, após tentativa com pelo menos dois antipsicóticos em dose e tempo adequados, a clozapina é indicada, sendo considerada o fármaco mais eficaz para casos resistentes, embora necessite monitoramento rigoroso devido ao risco de agranulocitose e outros efeitos adversos (66).

Antipsicóticos típicos, como haloperidol, atuam como antagonistas dos receptores dopaminérgicos D2 e demonstram eficácia no controle dos sintomas positivos, como delírios e alucinações, porém apresentam elevada incidência de efeitos colaterais, como sintomas extrapiramidais, que podem ser capazes de comprometer a adesão ao tratamento (20). Já os antipsicóticos atípicos, como risperidona e olanzapina, além de atuarem nos receptores dopaminérgicos, interagem com receptores serotoninérgicos, apresentando menor risco de efeitos extrapiramidais e maior tolerabilidade, embora estejam associados a riscos metabólicos, como ganho ponderal de peso, dislipidemias e resistência à insulina (20).

Apesar dos avanços, a adesão ao tratamento farmacológico da esquizofrenia permanece sendo influenciado por fatores como os efeitos adversos, estigma, falta de “*insights*” e aspectos socioculturais. Nesse sentido, a não adesão terapêutica está diretamente relacionada ao aumento do risco de recaídas, re-hospitalização, maior custo para os sistemas de saúde e piora no prognóstico funcional e cognitivo (67–69).

Os antipsicóticos injetáveis de longa duração (LAI) representam uma abordagem promissora no manejo farmacológico da esquizofrenia, principalmente quando se refere à adesão (70). Os LAI, administrados por via intramuscular, foram desenvolvidos com o intuito de promover a liberação prolongada do fármaco no organismo, seja ele de primeira ou segunda geração, assegurando níveis confiáveis e permitindo reduzir a frequência de administração de antipsicóticos, enquanto

mantém um controle satisfatório dos sintomas da esquizofrenia, com eficácia e segurança comparáveis aos antipsicóticos orais. Além de influenciar positivamente à adesão terapêutica, alguns estudos demonstram que o uso de LAI contribui na redução de recaídas e hospitalização, no menor risco de mortalidade e na melhora da funcionalidade e qualidade de vida desses indivíduos (21,70,71). Nesse sentido, o uso do LAI se torna uma ferramenta importante dentro das estratégias de cuidado, especialmente em contextos nos quais a adesão do tratamento se apresenta como obstáculo.

Embora os antipsicóticos representem a base do tratamento farmacológico da esquizofrenia, estratégias terapêuticas complementares têm se destacado, especialmente as abordagens psicossociais, como a Terapia Cognitivo-Comportamental (TCC) (22,72,73). Baseada no modelo cognitivo desenvolvido por Aaron Beck, a TCC parte do pressuposto que distorções cognitivas influenciam diretamente emoções e comportamentos, e propõe intervenções que visam à reorganização sobre os padrões de pensamentos desadaptativos e das crenças disfuncionais (23).

Por se tratar de uma abordagem estruturada, psicoeducativa e orientada por objetivos, a TCC é capaz de oferecer aos indivíduos com esquizofrenia a reestruturação cognitiva, o desenvolvimento de habilidades de enfrentamento, bem como, uma melhora expressiva na adesão ao tratamento e redução do sofrimento associado aos sintomas psicóticos (44,74). Essa abordagem tem se mostrado favorável principalmente na atenuação de sintomas positivos, como delírios e alucinações, ao oferecer ferramentas para que os indivíduos com esquizofrenia reavaliem crenças persecutórias e enfrentam experiências psicóticas com menor sofrimento subjetivo (74).

Uma metanálise recente sugere que a TCC tem o potencial de reduzir alguns aspectos dos sintomas negativos como apatia, retraimento social e baixa motivação, sobretudo, quando combinada a outras intervenções psicossociais, como o treinamento de habilidades sociais (24). Há um consenso na literatura de que a TCC, quando bem conduzida e adjuvante ao tratamento farmacológico, contribui significativamente para a melhora da qualidade de vida, da autonomia e da recuperação funcional dos indivíduos com esquizofrenia, especialmente ao abordar dimensões clínicas não suficientemente controladas pela farmacoterapia isolada (22,72).

3 JUSTIFICATIVA

Diante desse panorama clínico, histórico e terapêutico, observa-se que, apesar dos avanços na caracterização e manejo da esquizofrenia, persistem importantes lacunas quanto à compreensão de seus mecanismos moleculares e biológicos do transtorno. Essas limitações sustentam a necessidade de abordagens complementares, capazes de integrar dados genômicos e ferramentas computacionais avançadas.

A investigação molecular representa uma abordagem fundamental para o aprofundamento da compreensão da fisiopatologia de doenças complexas, permitindo a identificação de seus mecanismos intrínsecos, como depressão maior, transtorno bipolar, autismo e esquizofrenia (75–77). A análise de expressão genética tem permitido identificar padrões de regulação associados a processos celulares e vias biológicas alteradas, fornecendo subsídios para potenciais biomarcadores diagnósticos, prognósticos e terapêuticos (78–82).

O uso de algoritmos de aprendizado de máquina ou *machine learning* (ML) tem ampliado o potencial das análises em dados transcriptômicos, permitindo detectar assinaturas moleculares complexas e classificar fenótipos com maior precisão (83–86). A disponibilidade crescente de bases públicas de expressão gênica fortalece a robustez das descobertas e possibilita a integração de diferentes tecidos e populações (78,85–87).

Apesar dos avanços no diagnóstico e tratamento, ainda há um desconhecimento relevante sobre os mecanismos neurobiológicos da esquizofrenia, o que justifica a intensificação das pesquisas em expressão gênica, especialmente com o uso de ferramentas bioinformáticas robustas (88). Assim, este trabalho busca caracterizar genes diferencialmente expressos e suas vias moleculares associadas à esquizofrenia, utilizando conjuntos de dados cerebrais *post-mortem* do córtex pré-frontal disponíveis publicamente, com uma abordagem bioinformática robusta e apoio de algoritmos de *machine learning* (89–91).

A elucidação da base genética e das vias moleculares associadas à esquizofrenia pode fornecer insights essenciais para aprofundar a compreensão de sua fisiopatologia, além de abrir perspectivas promissoras para o desenvolvimento de ferramentas inovadoras que aprimorem tanto o diagnóstico quanto as estratégias terapêuticas da doença.

4 OBJETIVOS

4.1 Objetivo geral

Identificar e caracterizar genes e vias moleculares associadas à esquizofrenia em amostras de cérebro e sangue periférico humano.

4.2 Objetivos específicos

- I. Identificar genes diferencialmente expressos capazes de distinguir indivíduos com e sem esquizofrenia em amostras cerebrais.
- II. Caracterizar vias moleculares associadas à esquizofrenia.
- III. Validar a assinatura transcricional encontrada em cérebro e amostras de sangue periférico humano.

5 METODOLOGIA

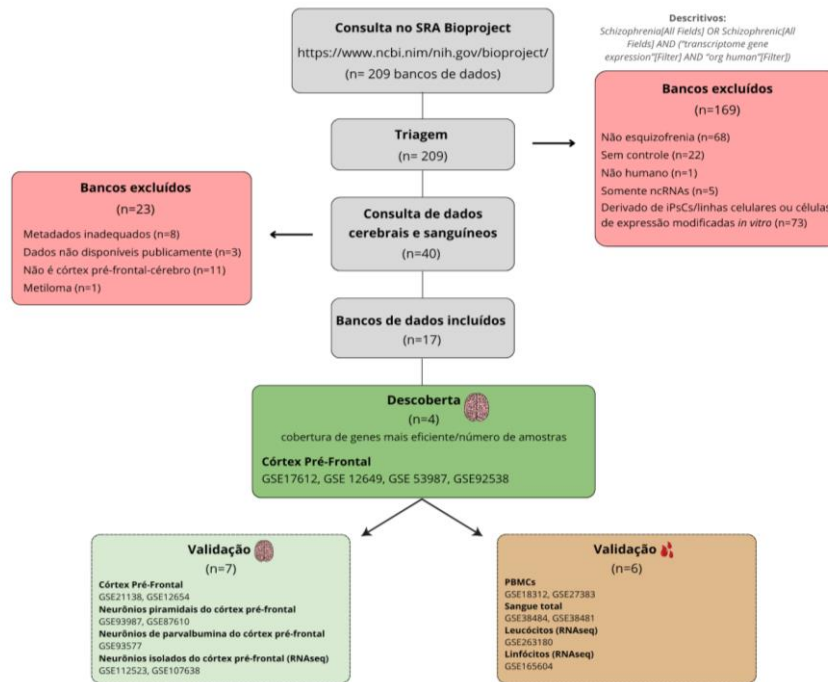
5.1 Estratégia de pesquisa e critérios de seleção

Os 209 conjuntos de dados disponíveis publicamente foram selecionados no banco de dados *Bioproject* do NCBI (<https://www.ncbi.nlm.nih.gov/bioproject/>) em 7 de fevereiro de 2025, usando a consulta "*Schizophrenia*(All Fields) OR *Schizophrenic*(All Fields) AND ("*transcriptome gene expression*"(Filter) AND "*org human*"(Filter))". O primeiro processo de seleção teve como objetivo excluir conjuntos de dados não relacionados à esquizofrenia ou que não apresentassem informações sobre a doença nas características contidas em seus metadados (n=68), não tivessem amostras de controle (n=22), não apresentassem amostras humanas (n=1), medissem apenas a expressão de RNAs não codificantes (*miRNAs* ou *circRNAs*) (n=5) ou tivessem amostras derivadas de células-tronco pluripotentes induzidas, linhagens celulares ou células com expressão modificada *in vitro* (n=73), excluindo 169 conjuntos de dados (**Figura 1**).

Os 40 conjuntos de dados restantes eram principalmente de amostras de cérebro e sangue, a segunda triagem removeu conjuntos de dados inadequados que não tinham símbolos genéticos nem incluíam os genes selecionados para validação (n=8), conjuntos de dados que não estavam disponíveis para *download* (n=3), conjuntos de dados cerebrais apresentando outras regiões além do córtex pré-frontal (n=11) e conjuntos de dados apresentando metilação (n=1), um total de 23 conjuntos de dados removidos. Após a triagem, apenas conjuntos de dados apresentando amostras do córtex pré-frontal do cérebro, PBMC, leucócitos, linfócitos e sangue total, com amostras de esquizofrenia e controle, foram incluídos na análise (**Figura 1**).

A amostragem final para esta análise resultou em 17 conjuntos de dados para *microarray*, sendo 6 conjuntos apresentando amostras de cérebro inteiro, 2 conjuntos apresentando células de neurônios piramidais do córtex pré-frontal, 1 conjunto apresentando células de neurônios parvalbumina do córtex pré-frontal, 2 conjuntos apresentando sangue total e, por fim, 2 conjuntos apresentando PBMCs. Para *RNAseq*, 4 conjuntos de dados foram incluídos nesta análise, sendo 1 conjunto apresentando sangue periférico, 1 conjunto apresentando linfócitos e 2 conjuntos apresentando neurônios isolados do córtex pré-frontal do cérebro e oligodendrócitos (**Figura 1**).

Figura 1. Fluxograma de amostragem



Fluxograma representando o processo de amostragem para esta meta-análise de dados de participantes individuais, com detalhes sobre os critérios de exclusão e o número de estudos excluídos/incluídos.

Os conjuntos de dados de *microarray* foram baixados usando o pacote *GEOquery* (versão 2.72.0) (92), usando a função *getGEO* com o parâmetro *GSEMatrix* definido como *TRUE* para baixar a matriz de expressão e *FALSE* para baixar os metadados. As sondas genéticas foram colapsadas usando o pacote *WGCNA* (versão 1.73) (93), usando a função *collapseRows* com os dados de expressão e dados de anotação como entrada, o parâmetro do método foi definido como "*absMaxMean*". A detecção e remoção de amostras discrepantes foram realizadas usando o pacote *arrayQualityMetrics* (versão 3.60.0), usando a função *outlier* com o método definido como "*KS*" (94). Para esta análise, apenas amostras de controle e esquizofrenia foram incluídas, mesmo de conjuntos de dados que apresentavam outras condições, como transtorno bipolar e TEPT. Após o *download*, os conjuntos de dados GSE17612 e GSE53987 foram transformados em *log2*. Devido à ausência de normalização prévia, os conjuntos de dados com normalização prévia foram mantidos sem transformação. Por outro lado, as tabelas de contagem processadas dos conjuntos de dados de *RNA-seq* foram recuperadas diretamente da

plataforma de *download* de dados do *geoNCBI*. O *download* e todas as análises subsequentes foram realizados em R (versão 4.4.2).

Todos os participantes com esquizofrenia foram diagnosticados de acordo com o DSM-3 ou DSM-4 em seus respectivos estudos originais, a grande maioria em tratamento com medicamentos antipsicóticos, como haloperidol, risperidona, olanzapina, clozapina e quetiapina. A distribuição geral de idade e sexo entre cada conjunto de dados com informações disponíveis foi avaliada usando o pacote *compareGroups* (versão 4.9.1) (95). A diferença estatística de idade entre os grupos foi testada usando o teste U de *Mann-Whitney*, enquanto a frequência de sexo foi testada usando o qui-quadrado de *Pearson*.

5.2 Análise dos dados

5.2.1 Desenho da análise e correção do efeito de lote

Os conjuntos de dados foram divididos entre conjuntos de descoberta e validação. Todos os conjuntos de dados cerebrais foram mesclados para avaliar a combinação mais eficiente com base no número de genes e amostras que compõem o conjunto de dados mesclado. O conjunto de descoberta mais eficiente foi composto por 4 conjuntos de dados (GSE17612 (96), GSE12649 (96,97), GSE53987 (98) e GSE92538 (97)), com um total de 226 amostras (124 controles e 102 esquizofrenia) e 10.829 genes. Os 10 conjuntos de dados restantes foram mantidos para validação. Posteriormente, o efeito do lote foi corrigido usando a Análise de Variável *Surrogate*, disponível no pacote *sva* (versão 3.52.0) (99). A correção do efeito do lote foi avaliada usando a Análise de Componentes Principais (PCA) antes e depois do *ComBat*, os componentes principais foram recuperados usando a função *prcomp* do pacote *stats* (versão 4.2.2), com o parâmetro *scale* definido como *TRUE*.

Além disso, os conjuntos de *microarrays* de validação foram organizados e mesclados de acordo com o tecido dos conjuntos de dados. O conjunto de dados de validação cerebral foi composto pelos 2 conjuntos de dados cerebrais restantes (GSE21138 (100) e GSE12654 (101)), apresentando 84 amostras (43 controles e 41 esquizofrenia). O conjunto de dados de validação de sangue total foi composto por 2 conjuntos de dados (GSE38484 e GSE38481 (102)), com 235 amostras (115 controles e 120 esquizofrenia). O conjunto de dados de validação de PBMCs também foi composto por 2 conjuntos de dados (GSE18312 e GSE27383 (103)), apresentando

88 amostras (34 controles e 54 esquizofrenia). Em relação aos neurônios isolados, o conjunto de dados do neurônio piramidal foi composto por 2 conjuntos de dados (GSE93987 e GSE87610 (104)) e apresentou 309 amostras (178 controles e 131 esquizofrenia). O conjunto de dados do neurônio parvalbumina, GSE93577 (100), foi mantido como 1 conjunto de dados individual e continha 139 amostras (69 controles e 70 esquizofrenia).

Para os conjuntos de dados de validação de *RNA-seq*, o conjunto de dados de linfócitos GSE165604 (105) foi incluído individualmente, apresentando 37 amostras (18 controles e 19 esquizofrenia). Dados de sangue periférico também foram incluídos como validação, com o conjunto de dados GSE263180 que apresentou 29 amostras (20 controles e 9 esquizofrenia). Por fim, neurônios e oligodendrócitos isolados também foram avaliados na validação, com 2 conjuntos de dados (GSE112523 (105,106) e GSE107638 (107)) sendo combinados, totalizando 74 amostras (44 controles e 30 esquizofrenia) após a remoção de amostras não esquizofrenia ou controle. Todos os conjuntos de dados de validação mesclados para cada tecido com 2 ou mais conjuntos de dados também tiveram seus efeitos de lote corrigidos e avaliados usando os mesmos pacotes e métodos do conjunto de descoberta.

5.2.2 Seleção de recursos: análise de expressão diferencial e machine learning

Após a correção do efeito do lote, a análise de expressão diferencial foi aplicada ao conjunto de descoberta, usando o pacote *limma* (versão 3.60.6) (108). Essa análise usou dados de *design* criados com base em informações de grupos (esquizofrenia ou controle) e dados de expressão como entrada, os valores estatísticos foram calculados ajustando um modelo linear com a função *lmFit*, incluindo contrastes com base em grupos com a função *contrasts.fit*, e aplicando moderação empírica de Bayes dos erros padrão com a função *eBayes* todos os parâmetros das funções foram definidos como padrão, além da expressão de entrada e dos dados de design. A comparação realizada foi entre grupos esquizofrenia e controle, definida com esquizofrenia - controle na função *makeContrasts* do *limma*, genes diferencialmente expressos (DEGs) foram identificados usando limiares de $|\log FC| > 1$ e valor de p ajustado por $FDR < 0,05$.

Na identificação dos DEGs mais informativos, seus dados de expressão foram recuperados e a seleção de características foi realizada usando o algoritmo de floresta aleatória, implementado nos pacotes *randomForest* (versão 4.7-1.2) (109) e *caret*

(versão 7.0-1) (110). Os parâmetros da floresta aleatória foram *mtry* definido como 24, *maxnodes* definido como 6 e gerando 400 árvores. Os DEGs mais informativos foram selecionados com base em seu índice de decréscimo médio de *Gini*, que está relacionado à homogeneidade do grupo da árvore em suas folhas, a fim de recuperar os genes com maior potencial de classificação. Para avaliar o perfil geral desses genes mais informativos, um mapa de calor com agrupamento hierárquico não supervisionado foi criado com os valores de expressão em escala z desses genes, usando a distância de *Manhattan* e o método de agrupamento *ward D2*. Uma validação cruzada *k-fold* também foi realizada para avaliar o modelo preditivo, usando a matriz de confusão do pacote *caret* para avaliar seu desempenho geral, com 5 dobras e 100 repetições. As entradas para a matriz de confusão foram as classes reais de cada amostra e as previsões realizadas pelo modelo, treinado com o método "*rF*". O parâmetro *trControl* foi definido usando o método "*repeatedcv*", multiplicado por 5 e repetido 100 vezes.

A etapa de validação foi realizada para avaliar o desempenho dos genes mais informativos na discriminação de esquizofrenia de amostras de controle em outros tipos de tecido. Para isso, a classificação das amostras foi realizada em cada conjunto de dados de validação, classificando suas amostras em esquizofrenia ou controle, a precisão geral do modelo foi avaliada comparando o agrupamento previsto pelos modelos e o agrupamento real das amostras, seu desempenho geral foi avaliado usando curvas ROC, com o pacote *pROC* (versão 1.18.5) (111). Um total de 8 conjuntos de validação foram incluídos, apresentando amostras de cérebro, sangue total, PBMCs, leucócitos, linfócitos, neurônios piramidais do córtex pré-frontal, neurônios de parvalbumina e neurônios isolados.

5.2.3 Identificação das vias biológicas mais representadas: análise de enriquecimento

Para identificar quais vias biológicas estavam associadas aos DEGs e aos genes selecionados na floresta aleatória, uma análise de enriquecimento foi realizada com os *IDs Entrez* disponíveis dos DEGs, recuperados usando o pacote *biomaRt* (versão 2.60.1) (112), e valores de *Fold change* classificados como entrada, usando o pacote *clusterProfiler* (versão 4.12.6) (113). A principal função empregada foi o *compareCluster*, com um tamanho máximo de conjunto de genes de 500, um tamanho mínimo de conjunto de genes de 10, um ponto de corte de valor q de 0,2, um ponto de corte de valor de p de 0,05 e o método de ajuste de p "*FDR*" como parâmetros.

Essa análise foi realizada duas vezes, usando o parâmetro "*fun*" definido como "*enrichPathway*" para usar os bancos de dados *Reactome* (114) e "*enrichKEGG*" para *KEGG* (115), com o objetivo de fornecer mais informações sobre as vias associadas aos DEGs e aos genes na floresta aleatória.

6 RESULTADOS

6.1 Desenho do estudo e descrição da população

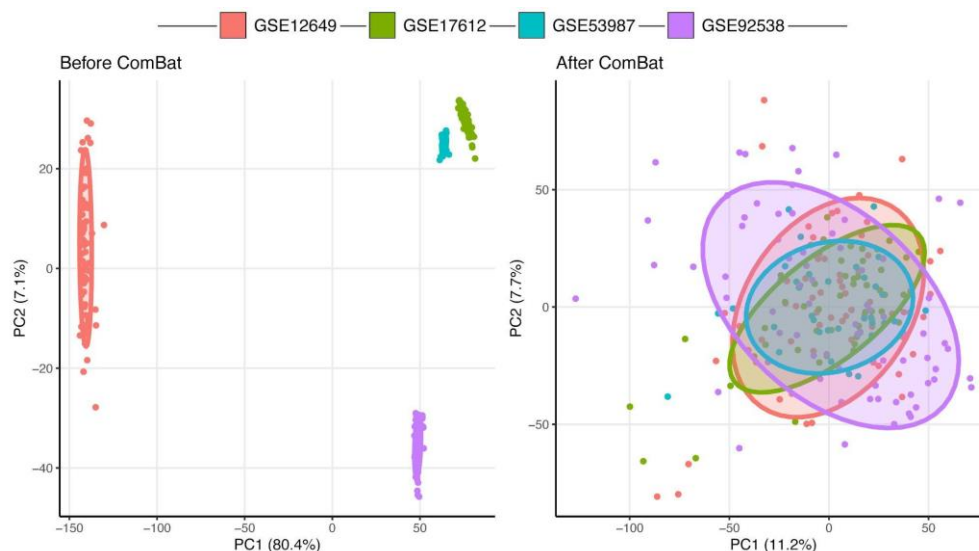
Dos 209 conjuntos de dados inicialmente avaliados, 17 foram recuperados para análise posterior. O córtex pré-frontal foi selecionado devido a evidências neurobiológicas que demonstram sua relevância para a patogênese da esquizofrenia, visto que essa região desempenha um papel central no suporte de funções cognitivas e executivas sabidamente prejudicadas no transtorno (116). Além disso, a maturação prolongada específica do córtex pré-frontal, que se estende até a adolescência, alinha-se ao período típico de início da esquizofrenia (117).

Após avaliar todas as combinações de conjuntos de dados do córtex pré-frontal cerebral, 4 conjuntos de dados foram selecionados como o conjunto de *descoberta*, enquanto os 2 conjuntos de dados cerebrais restantes foram usados para validação, juntamente com 2 conjuntos de dados contendo PBMCs, 2 conjuntos de dados contendo sangue total, 2 conjuntos de dados contendo neurônios piramidais isolados do córtex pré-frontal, 1 conjunto de dados contendo neurônios parvalbumina isolados do córtex pré-frontal, 1 conjunto de dados contendo leucócitos, 1 contendo linfócitos e 2 contendo neurônios isolados do córtex pré-frontal (**Figura 1**). Em relação à distribuição por idade e sexo entre os conjuntos de dados, vários conjuntos de dados não tinham informações disponíveis sobre idade e sexo (GSE12649, GSE12654, GSE87610, GSE93577, GSE93987). Entre os conjuntos de dados com informações disponíveis sobre idade e sexo, a única diferença significativa foi na distribuição por sexo, observada no conjunto de sangue total (compreendendo GSE38481 e GSE38484), onde o grupo esquizofrenia incluiu mais homens do que o grupo controle.

6.2 Expressão diferencial e seleção de características para representar a esquizofrenia

Os 4 conjuntos de dados que compõem o conjunto de descoberta foram recuperados e os efeitos de lote foram corrigidos (**Figura 2**). O conjunto de dados de descoberta final incluiu 226 amostras (124 controles e 102 esquizofrenia) e 10.829 genes. Um total de 532 DEGs foram identificados ao comparar os grupos esquizofrenia e controle, dos quais 383 foram regulados positivamente e 149 foram regulados negativamente.

Figura 2. Componentes principais do conjunto de descobertas antes e depois da mitigação do efeito de lote



Diagramas de pontos exibindo o primeiro e o segundo componentes principais (CPs) mais informativos do conjunto de descobertas antes e depois da mitigação do efeito de lote com Análise de Variável Substituta. As cores dos pontos representam os diferentes lotes e as elipses coloridas ilustram a área geral do gráfico que um lote ocupa.

Para identificar os genes diferencialmente expressos (DEGs) mais informativos que distinguem esquizofrenia de amostras de controle, realizamos a seleção de características usando o algoritmo de floresta aleatória. Os 3 principais genes — *HUWE1*, *PTGDS* e *RPL31* — foram selecionados com base na diminuição média no índice de Gini, formando o modelo mais preditivo (**Figura 3A**). Para avaliar o perfil de expressão desses genes, um mapa de calor com dados de expressão em escala de escore z foi empregado (**Figura 3B**). Dois *clusters* principais foram observados no mapa de calor, o primeiro sendo composto principalmente por amostras de esquizofrenia (84,31%), enquanto o segundo *cluster* foi formado principalmente por amostras de controle (87,09%). Além disso, todos os 3 genes foram regulados positivamente nas amostras de esquizofrenia: *RPL31*, *HUWE1* e *PTGDS* apresentaram alterações de 7,62, 20,84 e 26,90 vezes, respectivamente. A análise de validação cruzada de *k-fold* revelou um bom desempenho de discriminação nas amostras de descoberta, com precisão de 1,0, intervalo de confiança (IC) de 95% (0,98-1), taxa de ausência de informação de 0,55, *kappa* de 1, sensibilidade de 1, especificidade de 1 e valores preditivos positivo e negativo de 1.

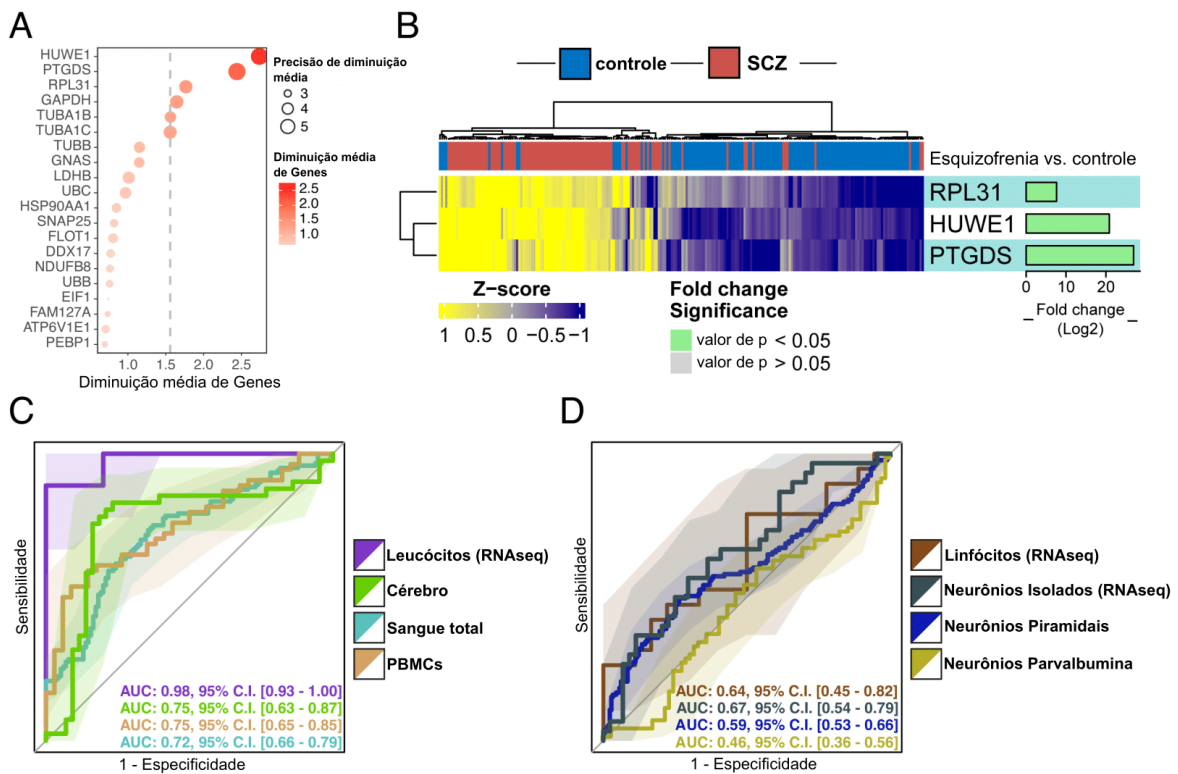
Os conjuntos de validação foram separados com base no tecido: cérebro, sangue total, PBMCs, leucócitos, neurônios de parvalbumina, neurônios piramidais,

linfócitos e neurônios isolados, cada um deles sendo composto de 1 a 3 conjuntos de dados diferentes. Cada tecido composto de dois ou mais conjuntos de dados teve seu efeito de lote corrigido antes de prosseguir com a avaliação de desempenho (**Figura 4**). Os dados de validação com células cerebrais foram compostos de 2 conjuntos de dados diferentes e apresentaram 84 amostras (43 controles e 41 esquizofrenia). Os 3 genes neste conjunto tiveram seus valores de expressão recuperados e o desempenho preditivo do modelo avaliado usando uma curva ROC, apresentando uma AUC de 0,75, IC de 95% (0,63-0,87) (**Figura 3C**). O mesmo foi feito com o conjunto de sangue total, que foi composto de dois conjuntos de dados diferentes, apresentando 235 amostras (115 controles e 120 esquizofrenia). Neste conjunto de dados, a curva ROC apresentou uma AUC de 0,72, IC de 95% (0,66-0,79) (**Figura 3C**). O conjunto PBMC também foi composto por dois conjuntos de dados, com um total de 88 amostras (34 controles e 54 esquizofrenia), e a AUC apresentada neste tecido foi de 0,75, IC de 95% (0,65-0,85) (**Figura 3C**). O conjunto de dados de *RNAseq* de leucócitos foi composto por 1 conjunto de dados com 29 amostras (20 controles e 9 esquizofrenia), a AUC ao classificar essas amostras foi de 0,98, IC de 95% (0,93-1,00) (**Figura 3C**).

Em seguida, o conjunto de *microarrays* de neurônios piramidais foi avaliado, composto por 2 conjuntos de dados e com 309 amostras (178 controles e 131 esquizofrenia). Diferentemente dos outros tecidos, o modelo de 3 genes apresentou desempenho discriminativo reduzido neste conjunto de dados, com AUCs de 0,59, IC de 95% (0,53-0,66) (**Figura 3D**). Em seguida, o conjunto de neurônios de parvalbumina foi composto por apenas um conjunto de dados, apresentando 139 amostras (69 controles e 70 esquizofrenia). A precisão do modelo de 3 genes também foi reduzida neste conjunto, com AUCs de 0,46, IC de 95% (0,36-0,56) (**Figura 3D**). A análise de classificação de linfócitos, compreendendo 1 conjunto de dados de *RNAseq* (37 amostras no total: 18 controles e 19 esquizofrenia), demonstrou desempenho discriminativo modesto com uma AUC de 0,64 (IC de 95%: 0,45-0,82) (**Figura 3D**). Por fim, o conjunto de *RNAseq* com amostras de neurônios isolados foi composto por 2 conjuntos de dados com 74 amostras (44 controles e 30 esquizofrenia). O desempenho do modelo neste conjunto atingiu uma AUC de 0,67, IC de 95% (0,54-0,79) (**Figura 3D**). Os tecidos apresentados, número de amostras, AUCs calculadas, tipo de dados e conjuntos de dados que compõem cada conjunto de validação foram compilados na **Tabela 1**. Essas descobertas sugerem que o sinal discriminatório do

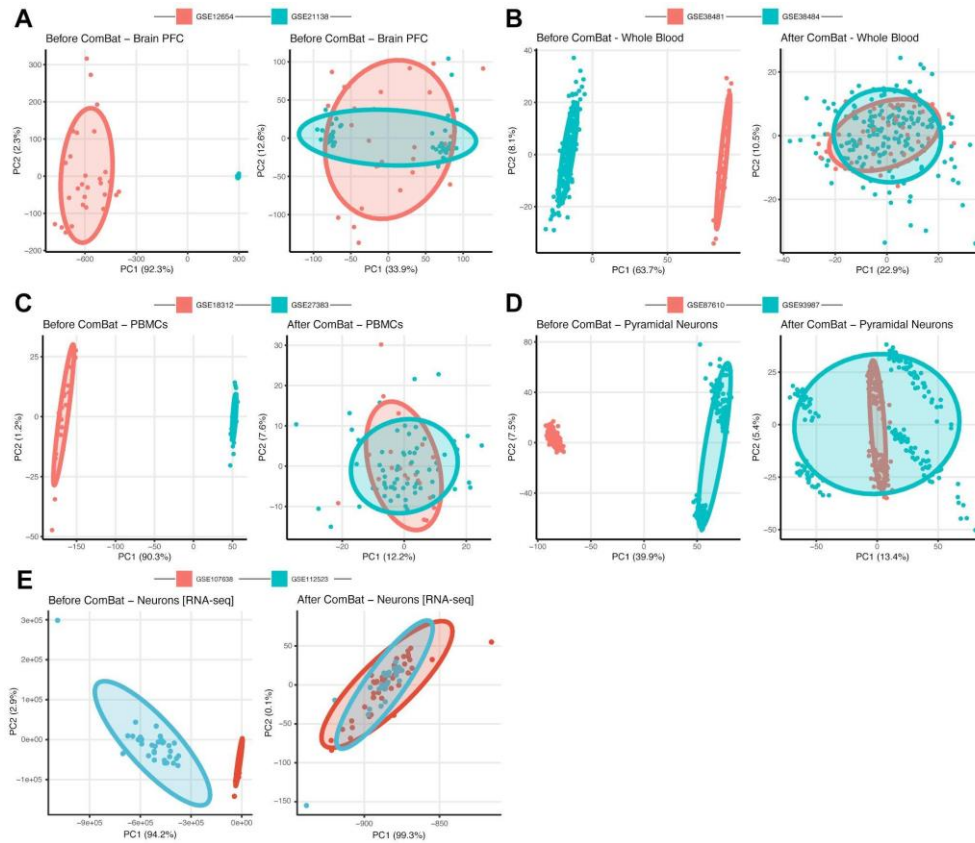
modelo pode não ser detectável em neurônios ou linfócitos isolados, mas pode ser transmitido através da barreira hematoencefálica e subindo de outros tipos de células de suporte.

Figura 3. Identificação dos genes mais informativos e avaliação do desempenho discriminativo



A. Diagrama de pontos exibindo os valores de importância da floresta aleatória usados para selecionar os genes mais informativos para separar amostras de esquizofrenia de controle. **B.** Mapa de calor com agrupamento hierárquico exibindo os dados de expressão em escala z dos genes mais informativos no conjunto de descoberta, as barras superiores representam os agrupamentos de amostras (vermelho para SCZ e azul para controle). Os diagramas de barras no lado direito exibem os valores de *Fold Change* calculados na comparação de SCZ vs controle, as barras verdes representam *Fold Change* significativos. **C.** Curvas ROC exibindo o desempenho dos genes mais informativos ao classificar amostras de SCZ e controle, o eixo Y representa a sensibilidade e o eixo X a especificidade 1. As curvas são coloridas por conjuntos (roxo para leucócitos, verde para cérebro, azul claro para sangue total e marrom claro para PBMC). **D.** Curvas ROC exibindo o desempenho dos genes mais informativos na classificação de amostras de SCZ e controle. As curvas são coloridas por conjuntos (marrom para linfócitos, cinza para neurônios isolados, azul para neurônios piramidais e amarelo escuro para neurônios de parvalbumina).

Figura 4. Componentes principais de cada conjunto de validação composto por mais de um conjunto de dados antes e depois da mitigação do efeito de lote



Diagramas de pontos exibindo o primeiro e o segundo componentes principais (CPs) mais informativos de cada conjunto usado para validação. Os CPs foram calculados usando a função *prcomp*, com a escala definida como *TRUE*. As cores dos pontos representam os diferentes lotes e as elipses coloridas ilustram a área geral do gráfico que um lote ocupa. **A.** Conjunto de validação do córtex pré-frontal cerebral. **B.** Conjunto de validação de sangue total. **C.** Conjunto de validação de células mononucleares do sangue periférico. **D.** Conjunto de validação de neurônios piramidais. **E.** Conjunto de validação de *RNAseq* de neurônios isolados.

Tabela 1. Caracterização dos conjuntos de dados de validação

Tecidos	Amostras	AUC	Tipos de dados	Conjuntos de dados
Córtex pré-frontal (BA 10, 46, 9)	N=84 (SCZ: 41, Controle: 43)	0.75, 95% C.I. (0.63 – 0.87)	Microarray	GSE12654, GSE21138
PBMC	N=88 (SCZ: 54, Controle: 34)	0.75, 95% C.I. (0.65 – 0.85)	Microarray	GSE18312, GSE27383
Sangue total	N=235 (SCZ: 120, Controle: 115)	0.72, 95% C.I. (0.66 – 0.79)	Microarray	GSE38484, GSE38481
Córtex pré-frontal dorsolateral (neurônios piramidais)	N=309 (SCZ: 131, Controle: 178)	0.59, 95% C.I. (0.53 – 0.66)	Microarray	GSE93987, GSE87610
Córtex pré-frontal dorsolateral cerebral (neurônios parvalbumina)	N=139 (SCZ: 70, Controle: 69)	0.46, 95% C.I. (0.36 – 0.56)	Microarray	GSE93577
Leucócitos	N=29 (SCZ: 9, Controle: 20)	0.98, 95% C.I. (0.93 – 1.00)	RNAseq	GSE263180
Linfócitos	N=37 (SCZ: 18, Controle: 19)	0.64, 95% C.I. (0.45 – 0.82)	RNAseq	GSE165604
Córtex pré-frontal (neurônios)	N=139 (SCZ: 70, Controle: 69)	0.67, 95% C.I. (0.54 – 0.79)	RNAseq	GSE112523, GSE107638

Exibindo tipo de tecido, número de amostras, desempenho geral dos genes mais informativos ao discriminar amostras SCZ e controle, tipos de dados (*microarray* ou *RNAseq*) e conjuntos de dados que compõem o conjunto de validação.

6.3 As vias associadas aos DEGs da esquizofrenia também estão associadas a outros distúrbios neuronais e psíquicos

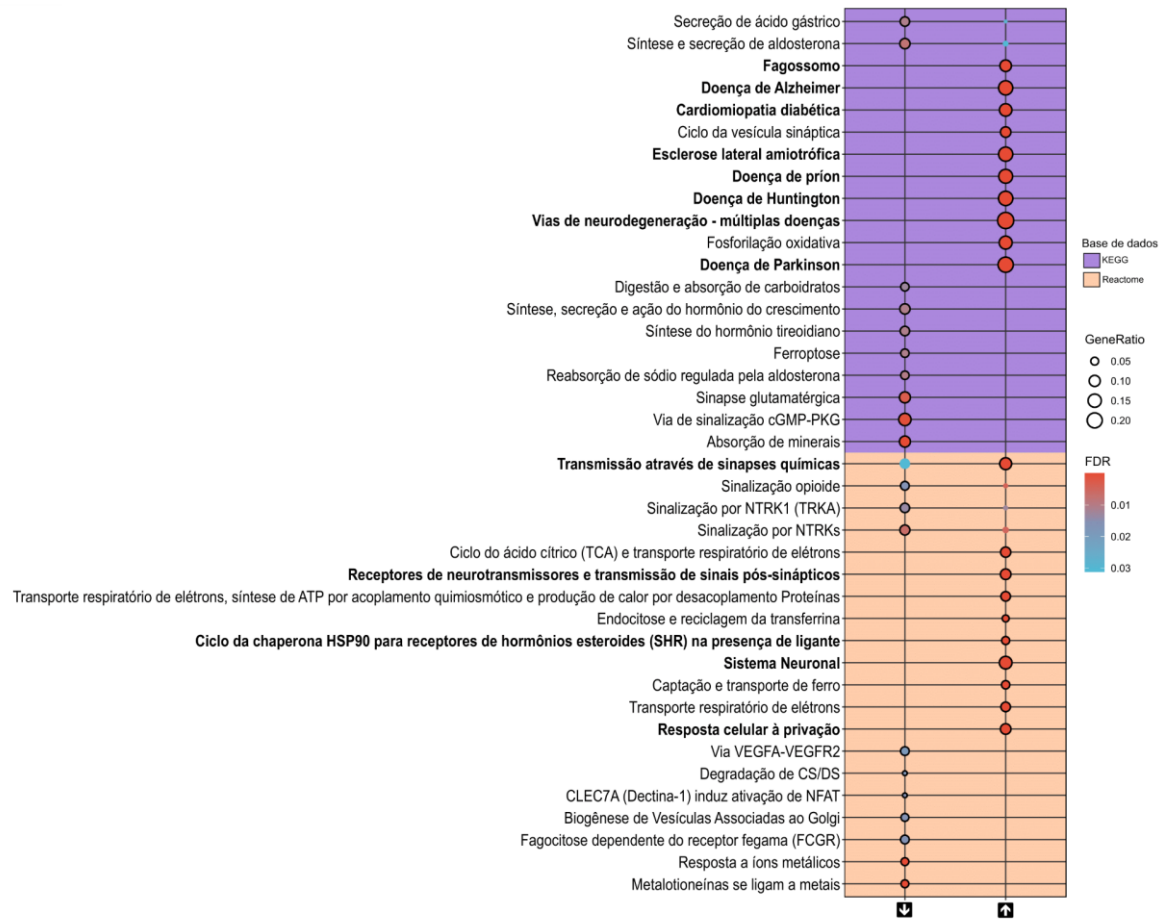
Após a identificação de uma assinatura transcriptômica associada à esquizofrenia e a validação da precisão do modelo em conjuntos de dados independentes (incluindo múltiplos tipos de tecidos), foram investigadas as vias biológicas associadas a esses genes candidatos. A partir dos DEGs, identificamos 92 vias enriquecidas no banco de dados *KEGG* e 432 no banco de dados *Reactome*. Em seguida, foram extraídas as 10 vias mais enriquecidas (classificadas por proporção gênica) para DEGs com regulação positiva e negativa de cada banco de dados. Para garantir uma cobertura abrangente, foram incluídas adicionalmente vias que apresentaram enriquecimento entre as 10 principais de ambos os grupos de DEGs, mesmo que não estivessem entre as 10 principais em um grupo. Por fim, foram

destacadas as 10 principais vias associadas aos 5 genes mais importantes do nosso modelo de floresta aleatória.

Em primeiro lugar, as 10 principais vias do banco de dados *KEGG* enriquecidas por DEGs superexpressos podem sugerir associações consistentes entre os genes identificados e distúrbios neuronais e psíquicos. Vias como Fagossomo, Doença de *Alzheimer*, Cardiomiopatia Diabética, Esclerose Lateral Amiotrófica, Doença de Prion, Doença de *Huntington*, Vias de Neurodegeneração - Doenças Múltiplas e Doença de *Parkinson* incluíram um ou mais genes do modelo de floresta aleatória em sua composição, potencialmente envolvendo perda neuronal ou processos inflamatórios (118,119) (**Figura 5**). Além delas, as vias do Ciclo da Vesícula Sináptica e da Fosforilação Oxidativa também foram enriquecidas neste conjunto de genes (**Figura 5**). Por outro lado, as 10 principais vias enriquecidas pelos genes sub expressos em esquizofrenia foram associados principalmente a vias metabólicas, como secreção de ácido gástrico, síntese e secreção de aldosterona, digestão e absorção de carboidratos, síntese, secreção e ação do hormônio do crescimento, síntese do hormônio tireoidiano, ferroptose, reabsorção de sódio regulada pela aldosterona, sinapse glutamatérgica, via de sinalização cGMP-PKG e absorção de minerais (**Figura 4**).

Em relação as 10 principais vias do banco de dados *Reactome* enriquecidas por DEGs super expressos na esquizofrenia, as vias Transmissão através de sinapses químicas, Receptores de neurotransmissores e transmissão de sinal pós-sináptico, Ciclo da chaperona *HSP90* para receptores de hormônios esteroides (SHR) na presença de ligante, Sistema neuronal e Resposta celular à inanição apresentaram genes de floresta aleatória em suas composições. Além disso, as vias ciclo do ácido cítrico (TCA) e transporte respiratório de elétrons, síntese de ATP por acoplamento quimiosmótico e produção de calor por proteínas desacopladoras, endocitose e reciclagem de transferrina, captação e transporte de ferro e transporte respiratório de elétrons também foram enriquecidas por esse conjunto de genes (**Figura 5**). As vias enriquecidas dos DEGs sub expressos foram sinalização de opioides, sinalização por *NTRK1* (*TRKA*), sinalização por *NTRKs*, via *VEGFA-VEGFR2*, degradação de *CS/DS*, *CLECTA* (Dectina-1) induz ativação de *NFAT*, biogênese de vesículas associadas ao sistema Golgi, fagocitose dependente do receptor Fcγ (FCGR), resposta a íons metálicos e metalotioneínas ligam-se a metais (**Figura 5**).

Figura 5. Diagrama de pontos para as principais vias da análise de enriquecimento. Vias enriquecidas nos bancos de dados *KEGG* e *Reactome*



O diagrama de pontos exibe as vias que foram enriquecidas usando os DEGs regulados para cima e para baixo como entrada. O tamanho do ponto representa a razão de genes das vias, enquanto a cor representa seu valor de p ajustado pela taxa de descoberta falsa. A cor de fundo representa os dois bancos de dados diferentes avaliados para o enriquecimento: o azul claro representa as vias *KEGG* e o vermelho claro representa as vias do *Reactome*. Os eixos Y e X exibem as vias e o conjunto DEG, separados pela direção da *Fold Change*.

7 DISCUSSÃO

Este trabalho empregou uma abordagem de bioinformática robusta para identificar assinaturas transcriptômicas consistentes da esquizofrenia em múltiplos conjuntos de dados de expressão genética disponíveis publicamente. Esta metodologia de algoritmos de *machine learning* para identificar biomarcadores transcriptômicos para caracterizar doenças foi usado em estudos anteriores com outras doenças, como tuberculose (89), HTLV-1 (90) e até mesmo doença de *Parkinson* (120). Neste estudo, as áreas de *Brodmann* 9, 10 e 46 do córtex pré-frontal do cérebro foram os tecidos usados como conjunto de descoberta, já que a disfunção dessas áreas é uma característica da esquizofrenia (121,122). O modelo preditivo baseado em floresta aleatória alcançou AUCs > 0,70 em amostras de cérebro, sangue, leucócitos e PBMC, exibindo uma precisão consistentemente alta mesmo em tipos de tecido mais acessíveis. Portanto, este modelo preditivo foi composto por apenas 3 genes regulados positivamente em esquizofrenia: *HUWE1*, *PTGDS* e *RPL31*.

O gene mais informativo, *HUWE1*, é sugerido como tendo um papel central enquanto um regulador conservado na proliferação e diferenciação de células progenitoras neuronais (123). Apesar de estar localizado no cromossomo X, os níveis de expressão desse gene são comparáveis entre homens e mulheres devido à inativação do cromossomo X em mulheres (123,124). No contexto da esquizofrenia, mutações espontâneas nesse gene foram encontradas em indivíduos com essa doença (125), o que pode estar relacionado às variações irregulares desse gene estarem associadas à deficiência intelectual (126), uma característica presente na esquizofrenia. Ao contrário dos outros dois genes, o *RPL31* não apresenta associações anteriores com a esquizofrenia, embora um estudo anterior tenha identificado proteínas codificadas por genes da mesma família com menor abundância na esquizofrenia em comparação ao controle em células neurais derivadas de células da neuroesfera olfatória (ONS) (127). No entanto, os resultados observados no estudo anterior não podem ser comparados diretamente com os resultados desta análise, devido às diferenças metodológicas entre os dois estudos.

O terceiro gene que compõe este modelo, *PTGDS*, também pode apresentar associações importantes com esquizofrenia, tendo sido proposto anteriormente como um biomarcador do líquido cefalorraquidiano para este transtorno e para *Alzheimer*

(128). Este gene foi selecionado como um biomarcador nos conjuntos de dados cerebrais, apesar de ser altamente expresso em tipos de células não neuronais (129), sugerindo uma forte influência da glia e de outros tipos de células de suporte presentes no cérebro. A codificação da proteína prostaglandina D sintase, *PTGDS*, participa do metabolismo da prostaglandina, pois catalisa a conversão da prostaglandina H2 em *PGD2*, um potente modulador da inflamação no sistema nervoso central. Esta molécula desempenha papéis importantes em processos neuro inflamatórios (130), estando relacionada a vias enriquecidas em *KEGG*, como o Fagossomo e vias associadas à glia, enriquecidas no *Reactome*, incluindo a Resposta celular à privação e a Fagocitose dependente do receptor *Fcgama* (FCGR) (131).

A influência da glia e de outros tipos de células de suporte também foi observada nas análises de validação de desempenho inferior em neurônios isolados, que alcançaram AUCs < 0,59. Isso sugere que neurônios isolados não apresentaram os níveis de expressão discriminativa observados em amostras de cérebro inteiro. Além da expressão de *PTGDS* ser mais proeminente em tipos de células não neuronais, o desempenho reduzido do nosso modelo em conjuntos de dados de neurônios isolados também pode refletir a ausência de interações neurônio-glia. A comunicação neurônio-glia desempenha um papel fundamental na regulação da atividade neuronal e da expressão gênica (132), e a ausência dessas interações em amostras de neurônios isolados pode diminuir o sinal detectado em análises de tecido em massa.

A validação bem-sucedida em sangue total, leucócitos e PBMCs sugere que o perfil de expressão do biomarcador pode atravessar a barreira hematoencefálica e ser detectado na circulação periférica. Notavelmente, o desempenho reduzido da classificação em linfócitos isolados sugere que o sinal discriminativo pode não ser específico do linfócito, mas sim influenciado por outras populações de leucócitos, hipótese reforçada pelo melhor desempenho observado em amostras de leucócitos.

Embora a maioria dos principais genes informativos que identificamos tenham sido previamente associados à esquizofrenia, a análise de enriquecimento reforça ainda mais sua associação com esse transtorno. Diversas vias *KEGG*, particularmente aquelas que envolvem os genes selecionados pela floresta aleatória, mostraram relevância para a esquizofrenia, com a maioria sendo associada a distúrbios neuronais. Apesar de suas vias etiológicas dominantes e apresentações clínicas serem diferentes, tanto a doença de *Alzheimer* quanto a esquizofrenia compartilham

o comprometimento da substância branca cerebral e o *déficit* cognitivo como características comuns (133). Além disso, pessoas com esquizofrenia têm risco 2 a 4 vezes maior de desenvolver *Alzheimer* do que a população em geral (134). A esclerose lateral amiotrófica também apresenta associações com a esquizofrenia, embora seus mecanismos biológicos comuns não sejam bem compreendidos. No entanto, foi observada uma correlação de *loci* genéticos de 14,3% entre ambas as doenças (135). Em relação às doenças de *Prion* e *Huntington*, ambas as doenças não apresentam fortes associações com a esquizofrenia, sendo relacionadas principalmente devido à característica clínica de psicose semelhante à esquizofrenia (136,137). Por fim, estudos apontam uma associação entre a esquizofrenia e maior risco para a doença de Parkinson (138,139) e diabetes (37,38,140).

Várias vias enriquecidas no banco de dados *Reactome* contendo genes da floresta aleatória podem ter associações com funções sinápticas e plasticidade neuronal anormais observados na esquizofrenia (141), como Transmissão através de Sinapses Químicas, Receptores de Neurotransmissores e transmissão de sinal pós-sináptico e Sistema Neuronal. Além disso, as vias enriquecidas por DEGs regulados negativamente relacionadas a Neurotrofinas, como a sinalização por *NTRKs*, também estão relacionadas à sobrevivência, diferenciação e plasticidade de neurônios (142), sendo a *NTRK1* (*TRKA*). O *NTRK1* (*TRKA*) está relacionado à manutenção da integridade das células neuronais na idade adulta (143). O gene *HUWE1* pode desempenhar um papel importante nessas vias, codificando a ubiquitina ligase E3, como regulador do *turnover* proteico sináptico, diferenciação neuronal e neurotransmissão, processos cuja desregulação tem sido associada à esquizofrenia (144).

Algumas vias enriquecidas contendo DEGs regulados positivamente foram relacionadas à atividade mitocondrial, como o ciclo do ácido cítrico (CA) e o transporte respiratório de elétrons, transporte respiratório de elétrons, síntese de ATP por acoplamento quimiosmótico e produção de calor por proteínas desacopladoras e Transporte respiratório de elétrons. O gene *RPL31* do modelo também pode estar relacionado a essas vias metabólicas e energéticas, pois codifica uma proteína ribossomal, que participa do controle translacional e da síntese proteica (145). Além disso, o número reduzido de mitocôndrias em diversas regiões cerebrais já havia sido associado à fisiopatologia da esquizofrenia (146). Possivelmente os DEGs super-

expressos que enriquecem essas vias reflitam o menor número de mitocôndrias, compensando a falta de unidades com níveis de atividade mais elevados.

Este trabalho apresenta algumas limitações, como informações sobre as etnias das amostras não estão disponíveis nos metadados de vários estudos, podendo introduzir um viés em relação à representação étnica, que não pode ser avaliado com os dados atuais, uma vez que a apresentação clínica e o risco da esquizofrenia variam entre diferentes origens étnicas e ancestrais (147,148). Covariáveis como gênero e idade não puderam ser incluídas no modelo *limma* usado para identificar os DEGs, pois este modelo exige metadados consistentes em todos os conjuntos de dados, os quais não estavam disponíveis. Outro potencial viés está relacionado ao uso de antipsicóticos, já que todos os conjuntos de dados contêm amostras de pacientes com esquizofrenia que usam antipsicóticos regularmente. Isso provavelmente cria um fator de confusão sobre se as alterações nos níveis de expressão gênica estão relacionadas à doença ou ao tratamento. Outra limitação diz respeito às interpretações causais dos resultados, visto que a natureza transversal deste estudo não fornece informações suficientes para estabelecer se os valores de expressão genética observados são causas ou consequências da esquizofrenia. Para abordar essa limitação, seriam necessários estudos longitudinais que avaliassem essas expressões gênicas ao longo do tempo. Por fim, estudos adicionais são necessários para avaliar o desempenho desse conjunto de genes em diferentes populações, a fim de fortalecer a validação da expressão desses genes e vias associadas à esquizofrenia.

8 CONCLUSÃO

A análise de enriquecimento revelou vias com associações prévias à fisiopatologia da esquizofrenia nos bancos de dados *KEGG* e *Reactome*, o que apoia os achados deste estudo com o conhecimento atual sobre o transtorno. Esses achados sugerem que a esquizofrenia pode envolver processos inflamatórios sistêmicos em vez de ser centralizada em neurônios, o que pode potencialmente desviar o foco de pesquisas futuras para outros tipos de células não neuronais. Além disso, a identificação de vias relacionadas a neurodegeneração, como *Alzheimer*, *Parkinson* e ELA, reforça o entendimento que a esquizofrenia compartilha vias patofisiológicas com essas doenças, abrindo um caminho para o desenvolvimento de novas estratégias terapêuticas combinadas.

Embora mais estudos sejam necessários para validar o desempenho destes biomarcadores em populações estratégicas, como por exemplo populações miscigenadas da América Latina, nossos achados identificaram genes-chave e suas vias associadas que são diferencialmente expressas na esquizofrenia, esclarecendo os potenciais mecanismos subjacentes à doença. Esse conhecimento é crucial para o potencial desenvolvimento de um novo teste diagnóstico baseado nos níveis de expressão de três genes, utilizando amostras de sangue de fácil acesso. Essa abordagem representa uma melhoria significativa em relação às estratégias diagnósticas e terapêuticas atuais, aprimorando a detecção e o manejo da esquizofrenia.

REFERÊNCIAS

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* [Internet]. novembro de 2021;8(11):969–80. Disponível em: [http://dx.doi.org/10.1016/S2215-0366\(21\)00243-1](http://dx.doi.org/10.1016/S2215-0366(21)00243-1)
2. Schizophrenia [Internet]. [citado 29 de setembro de 2025]. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/schizophrenia>
3. American Psychiatric Association. DSM Library. [citado 29 de setembro de 2025]. *Diagnostic and Statistical Manual of Mental Disorders*. Disponível em: <https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425787>
4. Queirós T, Coelho F, Linhares L, Telles-Correia D. [Schizophrenia: What Non-Psychiatrist Physicians Need to Know]. *Acta medica portuguesa* [Internet]. 1º de fevereiro de 2019 [citado 29 de setembro de 2025];32(1). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30753806/>
5. Fonseca VC, de Freitas Rios NV. Linguagem e comunicação na perspectiva do sujeito com esquizofrenia. *Distúrb Comun* [Internet]. 2 de dezembro de 2021 [citado 29 de setembro de 2025];33(4):615–26. Disponível em: <https://revistas.pucsp.br/index.php/dic/article/view/52122>
6. Keepers GA, Fochtmann LJ, Anzia JM, Benjamin S, Lyness JM, Mojtabai R, et al. The American Psychiatric Association Practice Guideline for the Treatment of Patients With Schizophrenia. *American Journal of Psychiatry* [Internet]. 1º de setembro de 2020 [citado 29 de setembro de 2025]; Disponível em: <https://psychiatryonline.org/doi/10.1176/appi.ajp.2020.177901>
7. Disease G 2017, Incidence I, Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)* [Internet]. 10 de novembro de 2018 [citado 29 de setembro de 2025];392(10159):1789. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6227754/>
8. Correll CU, Solmi M, Croatto G, Schneider LK, Christy Rohani-Montez S, Fairley L, et al. Mortality in people with schizophrenia: a systematic review and meta-analysis of relative risk and aggravating or attenuating factors. *World Psychiatry* [Internet]. 1º de junho de 2022 [citado 29 de setembro de 2025];21(2):248–71. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/wps.20994>
9. Veeneman RR, Vermeulen JM, Abdellaoui A, Sanderson E, Wootton RE, Tadros R, et al. Exploring the Relationship Between Schizophrenia and Cardiovascular Disease: A Genetic Correlation and Multivariable Mendelian Randomization Study. *Schizophr Bull* [Internet]. 1º de março de 2022;48(2):463–73. Disponível em: <http://dx.doi.org/10.1093/schbul/sbab132>
10. Dickerson F, Khan S, Origoni A, Rowe K, Katsafanas E, Harvin A, et al. Risk

Factors for Natural Cause Mortality in Schizophrenia. *JAMA Netw Open* [Internet]. 3 de setembro de 2024;7(9):e2432401. Disponível em: <http://dx.doi.org/10.1001/jamanetworkopen.2024.32401>

11. Factors and their weight in reducing life expectancy in schizophrenia. *Schizophrenia Research* [Internet]. 1º de dezembro de 2022 [citado 29 de setembro de 2025];250:67–75. Disponível em: <http://dx.doi.org/10.1016/j.schres.2022.10.019>

12. Peritogiannis V, Ninou A, Samakouri M. Mortality in Schizophrenia-Spectrum Disorders: Recent Advances in Understanding and Management. *Healthcare (Basel)* [Internet]. 25 de novembro de 2022;10(12). Disponível em: <http://dx.doi.org/10.3390/healthcare10122366>

13. Aymerich C, Salazar de Pablo G, Pacho M, Pérez-Rodríguez V, Bilbao A, Andrés L, et al. All-cause mortality risk in long-acting injectable versus oral antipsychotics in schizophrenia: a systematic review and meta-analysis. *Mol Psychiatry* [Internet]. janeiro de 2025;30(1):263–71. Disponível em: <http://dx.doi.org/10.1038/s41380-024-02694-3>

14. Keshavan MS, Collin G, Guimond S, Kelly S, Prasad KM, Lizano P. Neuroimaging in Schizophrenia. *Neuroimaging Clin N Am* [Internet]. fevereiro de 2020;30(1):73–83. Disponível em: <http://dx.doi.org/10.1016/j.nic.2019.09.007>

15. Lisoway AJ, Chen CC, Zai CC, Tiwari AK, Kennedy JL. Toward personalized medicine in schizophrenia: Genetics and epigenetics of antipsychotic treatment. *Schizophr Res* [Internet]. junho de 2021;232:112–24. Disponível em: <http://dx.doi.org/10.1016/j.schres.2021.05.010>

16. Elkis H, Louzã MR. Novos antipsicóticos para o tratamento da esquizofrenia. *Arch Clin Psychiatry (Impr)* [Internet]. 2007 [citado 29 de setembro de 2025];193–7. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-60832007000800009

17. Leucht S, Priller J, Davis JM. Antipsychotic Drugs: A Concise Review of History, Classification, Indications, Mechanism, Efficacy, Side Effects, Dosing, and Clinical Application. *Am J Psychiatry* [Internet]. 1º de outubro de 2024;181(10):865–78. Disponível em: <http://dx.doi.org/10.1176/appi.ajp.20240738>

18. Shimomura Y, Kikuchi Y, Suzuki T, Uchida H, Mimura M, Takeuchi H. Antipsychotic treatment in the maintenance phase of schizophrenia: An updated systematic review of the guidelines and algorithms. *Schizophr Res* [Internet]. janeiro de 2020;215:8–16. Disponível em: <http://dx.doi.org/10.1016/j.schres.2019.09.013>

19. Cruz LO, Dolabela MF. Tratamento medicamentoso de portadores da esquizofrenia: adesão, interações medicamentosas e reações adversas. *RSD* [Internet]. 3 de março de 2021 [citado 29 de setembro de 2025];10(3):e2010313087–e2010313087. Disponível em: <https://rsdjournal.org/rsd/article/view/13087>

20. Chokhawala K, Stevens L. Antipsychotic Medications. Em: *StatPearls* [Internet] [Internet]. StatPearls Publishing; 2023 [citado 29 de setembro de 2025]. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK519503/>

21. Brissos S, Veguilla MR, Taylor D, Balanzá-Martinez V. The role of long-acting injectable antipsychotics in schizophrenia: a critical appraisal. *Therapeutic Advances in Psychopharmacology* [Internet]. outubro de 2014 [citado 29 de setembro de 2025];4(5):198. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4212490/>
22. Xu F, Zhang H. The application of cognitive behavioral therapy in patients with schizophrenia: A review. *Medicine* [Internet]. 11 de agosto de 2023 [citado 29 de setembro de 2025];102(32):e34827. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10419479/>
23. Guaiana G, Abbatecola M, Aali G, Tarantino F, Ebuenyi ID, Lucarini V, et al. Cognitive behavioural therapy (group) for schizophrenia. *The Cochrane Database of Systematic Reviews* [Internet]. 12 de julho de 2022 [citado 29 de setembro de 2025];2022(7):CD009608. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9308944/>
24. Hong Y, Chen Y, Bai Y, Tan W. Cognitive-behavioral therapy for the improvement of negative symptoms and functioning in schizophrenia: A systematic review and meta-analysis of randomized controlled trials. *PLOS One* [Internet]. 20 de maio de 2025 [citado 29 de setembro de 2025];20(5):e0324685. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12091889/>
25. American Psychiatric Association. DSM Library. [citado 28 de maio de 2025]. *Diagnostic and Statistical Manual of Mental Disorders*. Disponível em: <https://psychiatryonline.org/doi/book/10.1176/appi.books.9780890425787>
26. Wildgust HJ, Hodgson R, Beary M. The paradox of premature mortality in schizophrenia: new research questions. *Journal of Psychopharmacology* [Internet]. novembro de 2010 [citado 28 de junho de 2025]; Disponível em: <https://journals.sagepub.com/doi/10.1177/1359786810382149>
27. De Sousa A, Shah B, Shrivastava A. Suicide and Schizophrenia: an Interplay of Factors. *Current Psychiatry Reports* [Internet]. 8 de outubro de 2020 [citado 14 de julho de 2025];22(12):1–5. Disponível em: <https://link.springer.com/article/10.1007/s11920-020-01188-7>
28. Sher L, Kahn RS. Suicide in Schizophrenia: An Educational Overview. *Medicina* [Internet]. 10 de julho de 2019 [citado 14 de julho de 2025];55(7):361. Disponível em: <https://www.mdpi.com/1648-9144/55/7/361>
29. Berardelli I, Rogante E, Sarubbi S, Erbutto D, Lester D, Pompili M. The Importance of Suicide Risk Formulation in Schizophrenia. *Front Psychiatry* [Internet]. 16 de dezembro de 2021 [citado 14 de julho de 2025];12:779684. Disponível em: <https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsy.2021.779684/pdf>
30. Aydın M, İlhan BÇ, Tekdemir R, Çökünlü Y, Erbasan V, Altınbaş K. Suicide attempts and related factors in schizophrenia patients. *Saudi Medical Journal* [Internet]. 1º de maio de 2019 [citado 14 de julho de 2025];40(5):475–82. Disponível em: <https://smj.org.sa/content/40/5/475.abstract>
31. Almeida MM de, Schal VT, Martins AM, Modena CM. A sobrecarga de cuidadores de pacientes com esquizofrenia. *Rev psiquiatr Rio Gd Sul* [Internet]. 2010

[citado 27 de junho de 2025];32(3):73–9. Disponível em:

<https://www.scielo.br/j/rprs/a/D9J4Z69Tf8tvdS84Z59RfjH/?lang=pt&format=pdf>

32. Tostes JG, De Vasconcelos ACB, Tostes CB da S, De Brito KAND, De Souza TF, De Freitas RL. Esquizofrenia e Cognição. *Rev Psicol Em Pesqui* [Internet]. 24 de outubro de 2020;14(4):102–19. Disponível em:

<https://periodicos.ufjf.br/index.php/psicologiaempesquisa/article/view/30284>

33. Oliveira RM, Facina PCBR, Siqueira Júnior AC. A realidade do viver com esquizofrenia. *Rev Bras Enferm* [Internet]. abril de 2012 [citado 27 de junho de 2025];65(2):309–16. Disponível em:

<https://www.scielo.br/j/reben/a/xCB7BQk3xcCnccx89pqRRpz/?lang=pt&format=pdf>

34. McWhinney SR, Brosch K, Calhoun VD, Crespo-Facorro B, Crossley NA, Dannowski U, et al. Obesity and brain structure in schizophrenia - ENIGMA study in 3021 individuals. *Mol Psychiatry* [Internet]. setembro de 2022;27(9):3731–7. Disponível em: <http://dx.doi.org/10.1038/s41380-022-01616-5>

35. Chen J, Perera G, Shetty H, Broadbent M, Xu Y, Stewart R. Body mass index and mortality in patients with schizophrenia spectrum disorders: a cohort study in a South London catchment area. *Gen Psychiatr* [Internet]. 4 de novembro de 2022;35(5):e100819. Disponível em: <http://dx.doi.org/10.1136/gpsych-2022-100819>

36. Liang J, Cai Y, Xue X, Li X, Li Z, Xu C, et al. Does Schizophrenia Itself Cause Obesity? *Front Psychiatry* [Internet]. 23 de junho de 2022 [citado 28 de junho de 2025];13:934384. Disponível em:

<https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsy.2022.934384/pdf>

37. Suvisaari J, Keinänen J, Eskelinen S, Mantere O. Diabetes and Schizophrenia. *Current Diabetes Reports* [Internet]. 23 de janeiro de 2016 [citado 28 de junho de 2025];16(2):1–10. Disponível em:

<https://link.springer.com/article/10.1007/s11892-015-0704-4>

38. Dong K, Wang S, Qu C, Zheng K, Sun P. Schizophrenia and type 2 diabetes risk: a systematic review and meta-analysis. *Front Endocrinol (Lausanne)* [Internet]. 11 de setembro de 2024;15:1395771. Disponível em:

<http://dx.doi.org/10.3389/fendo.2024.1395771>

39. Annamalai A, Kosir U, Tek C. Prevalence of obesity and diabetes in patients with schizophrenia. *World J Diabetes* [Internet]. 15 de agosto de 2017;8(8):390–6. Disponível em: <http://dx.doi.org/10.4239/wjd.v8.i8.390>

40. de Hert M, Schreurs V, Vancampfort D, Van Winkel R. Metabolic syndrome in people with schizophrenia: a review. *World Psychiatry* [Internet]. 1º de fevereiro de 2009 [citado 28 de junho de 2025];8(1):15–22. Disponível em:

<https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2051-5545.2009.tb00199.x>

41. Goh KK, Chen CYA, Wu TH, Chen CH, Lu ML. Crosstalk between Schizophrenia and Metabolic Syndrome: The Role of Oxytocinergic Dysfunction. *Int J Mol Sci* [Internet]. 25 de junho de 2022;23(13). Disponível em:

<http://dx.doi.org/10.3390/ijms23137092>

42. Manta A, Georganta A, Roumpou A, Zoumpourlis V, Spandidos DA, Rizos E, et al. Metabolic syndrome in patients with schizophrenia: Underlying mechanisms and therapeutic approaches (Review). *Mol Med Rep* [Internet]. maio de 2025;31(5). Disponível em: <http://dx.doi.org/10.3892/mmr.2025.13479>
43. Howes OD, Cummings C, Chapman GE, Shatalina E. Neuroimaging in schizophrenia: an overview of findings and their implications for synaptic changes. *Neuropsychopharmacology* [Internet]. 2 de setembro de 2022 [citado 14 de julho de 2025];48(1):151–67. Disponível em: <https://www.nature.com/articles/s41386-022-01426-x>
44. Dabiri M, Dehghani Firouzabadi F, Yang K, Barker PB, Lee RR, Yousem DM. Neuroimaging in schizophrenia: A review article. *Front Neurosci* [Internet]. 15 de novembro de 2022 [citado 14 de julho de 2025];16:1042814. Disponível em: <https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2022.1042814/pdf>
45. Gadelha A, Nardi AE, da Silva AG. Esquizofrenia - 2.ed.: Teoria e clínica [Internet]. Artmed Editora; 2020. 226 p. Disponível em: <https://play.google.com/store/books/details?id=FYoBEAAAQBAJ>
46. Pereira MEC. Morel e a questão da degenerescência. *Rev latinoam psicopatol fundam* [Internet]. setembro de 2008 [citado 25 de junho de 2025];11(3):490–6. Disponível em: <https://www.scielo.br/j/rlpf/a/z7jd4V4RsPPDddy5KzHzCJb/?lang=pt&format=pdf>
47. Ebert A, Bär KJ. Emil Kraepelin: A pioneer of scientific understanding of psychiatry and psychopharmacology. *Indian J Psychiatry* [Internet]. abril de 2010;52(2):191–2. Disponível em: <http://dx.doi.org/10.4103/0019-5545.64591>
48. Géraud M. [Emil Kraepelin: a pioneer of modern psychiatry. On the occasion of the hundred and fiftieth anniversary of his birth]. *Encephale* [Internet]. setembro de 2007;33(4 Pt 1):561–7. Disponível em: [http://dx.doi.org/10.1016/s0013-7006\(07\)92053-9](http://dx.doi.org/10.1016/s0013-7006(07)92053-9)
49. Silva RCB da. Esquizofrenia: uma revisão. *Psicol USP* [Internet]. 2006 [citado 9 de junho de 2025];17(4):263–85. Disponível em: <https://www.scielo.br/j/pusp/a/Vt9jGsLzGs535fdrsXKHxz/?lang=pt&format=pdf>
50. George EL. MentalHealth.com. 2024 [citado 5 de junho de 2025]. Historical And Contemporary Understandings Of Schizophrenia. Disponível em: <https://www.mentalhealth.com/library/historical-contemporary-understandings-schizophrenia>
51. ICD-11 for Mortality and Morbidity Statistics [Internet]. [citado 30 de setembro de 2025]. Disponível em: <https://icd.who.int/browse/2025-01/mms/en#1683919430>
52. ICD-10 Version:2019 [Internet]. Disponível em: <https://icd.who.int/browse10/2019/en#/F20>
53. Vista do Linguagem e comunicação na perspectiva do sujeito com esquizofrenia [Internet]. [citado 28 de junho de 2025]. Disponível em:

<https://revistas.pucsp.br/index.php/dic/article/view/52122/38107>

54. de Sousa CC. ESQUIZOFRENIA ESQUIZOAFETIVA E A DINÂMICA DOS ESTADOS PRODRÔMICO, EPISÓDICO E RESIDUAL, NO CASO CLÍNICO C.B.F.S. Rev Acad Online [Internet]. 1º de julho de 2019 [citado 29 de junho de 2025];5(27). Disponível em:

<https://revistaacademicaonline.com/index.php/rao/article/view/648>

55. Cannon M, Caspi A, Moffitt TE, Harrington H, Taylor A, Murray RM, et al. Evidence for early-childhood, pan-developmental impairment specific to schizophreniform disorder: results from a longitudinal birth cohort. Arch Gen Psychiatry [Internet]. maio de 2002;59(5):449–56. Disponível em:

<http://dx.doi.org/10.1001/archpsyc.59.5.449>

56. Ogden TH, Muszkat S. Sobre três formas de pensar: o pensamento mágico, o pensamento onírico e o pensamento transformativo. Rev bras psicanál [Internet]. 2012 [citado 30 de junho de 2025];46(2):193–214. Disponível em:

http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S0486-641X2012000200016&lng=pt&nrm=iso&tlng=pt

57. Oliveira CW de M, Roma TM, Moreira LG, Corsso C del, Sousa AFB de, Silva JCP da, et al. Esquizofrenia: prevalência, diagnóstico e avanços no tratamento. Rev Cad Pedagóg [Internet]. 20 de janeiro de 2025;22(1):e13486. Disponível em:

<https://ojs.studiespublicacoes.com.br/ojs/index.php/cadped/article/view/13486>

58. Vista do PROCESSOS NEUROLÓGICOS ENVOLVIDOS NA ESQUIZOFRENIA [Internet]. [citado 30 de junho de 2025]. Disponível em:

<https://themaetscientia.fag.edu.br/index.php/RTES/article/view/738/832>

59. ICD-10 Version:2019 [Internet]. Disponível em:

<https://icd.who.int/browse10/2019/en#/F20>

60. ICD-10 Version:2019 [Internet]. Disponível em:

<https://icd.who.int/browse10/2019/en#/F20>

61. ICD-10 Version:2019 [Internet]. Disponível em:

<https://icd.who.int/browse10/2019/en#/F20>

62. ICD-10 Version:2019 [Internet]. Disponível em:

<https://icd.who.int/browse10/2019/en#/F20>

63. Russo J, Venâncio ATA. Classificando as pessoas e suas perturbações: a “revolução terminológica” do DSM III. Rev latinoam psicopatol fundam [Internet]. 2006 [citado 24 de julho de 2025];9(3):460–83. Disponível em:

<https://www.scielo.br/j/rlpf/a/56hWVSkWPtnWTGrjVB6qWGS/?lang=pt&format=pdf>

64. Kendell RE. Diagnostic and Statistical Manual of Mental Disorders, 3rd ed., revised (DSM-III-R). American Journal of Psychiatry [Internet]. 1º de abril de 2006 [citado 22 de julho de 2025]; Disponível em:

<https://psychiatryonline.org/doi/10.1176/ajp.145.10.1301>

65. Guze SB. Diagnostic and Statistical Manual of Mental Disorders, 4th ed.

(DSM-IV). American Journal of Psychiatry [Internet]. 1º de abril de 2006 [citado 23 de julho de 2025]; Disponível em:

<https://psychiatryonline.org/doi/10.1176/ajp.152.8.1228>

66. Abuse S. Table 3.22, DSM-IV to DSM-5 Schizophrenia Comparison. junho de 2016 [citado 24 de julho de 2025]; Disponível em:

<https://www.ncbi.nlm.nih.gov/books/NBK519704/table/ch3.t22/>

67. Elkis H, Louzã MR. Novos antipsicóticos para o tratamento da esquizofrenia. Arch Clin Psychiatry (São Paulo) [Internet]. 2007 [citado 3 de julho de 2025];34:193–7. Disponível em:

<https://www.scielo.br/j/rpc/a/4vRtPYz8DqYPnrc4bBHpsGc/?lang=pt&format=pdf>

68. Bertoni RA, Leal FM. Uma revisão do tratamento da esquizofrenia: monoterapia vs associação de antipsicóticos. Debates em Psiquiatria [Internet]. 3 de fevereiro de 2023 [citado 3 de julho de 2025];13:1–20. Disponível em:

<https://revistardp.org.br/revista/article/view/414>

69. Rubio JM, Kane JM, Tanskanen A, Tiihonen J, Taipale H. Long-term persistence of the risk of agranulocytosis with clozapine compared with other antipsychotics: a nationwide cohort and case-control study in Finland. Lancet Psychiatry [Internet]. junho de 2024;11(6):443–50. Disponível em:

[http://dx.doi.org/10.1016/S2215-0366\(24\)00097-X](http://dx.doi.org/10.1016/S2215-0366(24)00097-X)

70. Loots E, Goossens E, Vanwesemael T, Morrens M, Van Rompaey B, Dilles T. Interventions to Improve Medication Adherence in Patients with Schizophrenia or Bipolar Disorders: A Systematic Review and Meta-Analysis. International Journal of Environmental Research and Public Health [Internet]. 28 de setembro de 2021 [citado 15 de julho de 2025];18(19):10213. Disponível em:

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8508496/>

71. García S, Martínez-Cengotitabengoa M, López-Zurbano S, Zorrilla I, López P, Vieta E, et al. Adherence to Antipsychotic Medication in Bipolar Disorder and Schizophrenic Patients: A Systematic Review. Journal of Clinical Psychopharmacology [Internet]. 14 de junho de 2016 [citado 15 de julho de 2025];36(4):355. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4932152/>

72. Lauriello J. Prevalence and Impact of Relapse in Patients With Schizophrenia. Psychiatrist.com [Internet]. 24 de março de 2020 [citado 15 de julho de 2025]; Disponível em: <https://www.psychiatrist.com/jcp/prevalence-and-impact-of-schizophrenia-relapse/>

73. Cipolla S, Delli Carpini F, Catapano P, De Santis V, Volpicelli A, Perris F, et al. Two Long-Acting Antipsychotics in a Patient with Treatment-Resistant Schizophrenia: A Case Report. Clinics and Practice [Internet]. 10 de março de 2025 [citado 11 de julho de 2025];15(3):55. Disponível em: <https://www.mdpi.com/2039-7283/15/3/55>

74. Hamina A, Taipale H, Lieslehto J, Lähteenvuo M, Tanskanen A, Mittendorfer-Rutz E, et al. Comparative Effectiveness of Antipsychotics in Patients With Schizophrenia Spectrum Disorder. JAMA Netw Open [Internet]. 9 de outubro de 2024 [citado 11 de julho de 2025];7(10):e2438358–e2438358. Disponível em:

https://jamanetwork.com/journals/jamanetworkopen/articlepdf/2824666/hamina_2024_oj_241113_1727811696.63881.pdf

75. Hasan A, Falkai P, Lehmann I, Gaebel W. Schizophrenia. *Deutsches Ärzteblatt International* [Internet]. 12 de junho de 2020 [citado 14 de julho de 2025];117(24):412. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC7477695/>
76. Kumari V, Fannon D, Peters ER, Ffytche DH, Sumich AL, Premkumar P, et al. Neural changes following cognitive behaviour therapy for psychosis: a longitudinal study. *Brain* [Internet]. 11 de julho de 2011 [citado 15 de julho de 2025];134(8):2396. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC3155705/>
77. Xu F, Xu S. Cognitive-behavioral therapy for negative symptoms of schizophrenia: A systematic review and meta-analysis. *Medicine* [Internet]. 6 de setembro de 2024 [citado 14 de julho de 2025];103(36):e39572. Disponível em: https://journals.lww.com/md-journal/abstract/2024/09060/cognitive_behavioral_therapy_for_negative_symptoms.89.aspx
78. Moreira-Rodrigues M. Editorial: Molecular mechanisms of neuropsychiatric diseases, volume II. *Frontiers in Molecular Neuroscience* [Internet]. 13 de novembro de 2024 [citado 20 de agosto de 2025];17:1517196. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC11599241/>
79. Nakamura T, Takata A. The molecular pathology of schizophrenia: an overview of existing knowledge and new directions for future research. *Molecular Psychiatry* [Internet]. 6 de março de 2023 [citado 20 de agosto de 2025];28(5):1868. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC10575785/>
80. Sestan N, Vaccarino F, Gerstein M, Weissman S, Pochareddy S, Matthew State, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* [Internet]. 9 de fevereiro de 2018 [citado 22 de agosto de 2025]; Disponível em: <https://www.science.org/doi/10.1126/science.aad6469>
81. Priyanka, Kumar R, Kumar V, Kumar A, Rana SS. Deciphering transcriptomic signatures in schizophrenia, bipolar disorder, and major depressive disorder. *Front Psychiatry* [Internet]. 14 de julho de 2025;16:1574458. Disponível em: <http://dx.doi.org/10.3389/fpsy.2025.1574458>
82. Hess JL, Tylee DS, Barve R, de Jong S, Ophoff RA, Kumarasinghe N, et al. Transcriptomic abnormalities in peripheral blood in bipolar disorder, and discrimination of the major psychoses. *Schizophrenia research* [Internet]. 4 de agosto de 2019 [citado 29 de setembro de 2025];217:124. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC6997041/>
83. (Differential) Co-Expression Analysis of Gene Expression: A Survey of Best Practices [Internet]. [citado 29 de setembro de 2025]. Disponível em: <https://ieeexplore.ieee.org/document/8613814>
84. Garcia JPT, Tayo LL. Codes between Poles: Linking Transcriptomic Insights

into the Neurobiology of Bipolar Disorder. *Biology* [Internet]. 30 de setembro de 2024 [citado 29 de setembro de 2025];13(10):787. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11505342/>

85. Kusonmano K. Gene Expression Analysis Through Network Biology: Bioinformatics Approaches. *Adv Biochem Eng Biotechnol* [Internet]. 2017;160:15–32. Disponível em: http://dx.doi.org/10.1007/10_2016_44

86. Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances* [Internet]. 1º de julho de 2021 [citado 29 de setembro de 2025];49:107739. Disponível em: <http://dx.doi.org/10.1016/j.biotechadv.2021.107739>

87. Website [Internet]. Disponível em: <https://doi.org/10.3390/genes16030244>

88. Bhuvaneshwar K, Gusev Y. Translational bioinformatics and data science for biomarker discovery in mental health: an analytical review. *Briefings in Bioinformatics* [Internet]. 16 de março de 2024 [citado 29 de setembro de 2025];25(2):bbae098. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC10944574/>

89. van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in Bioinformatics* [Internet]. 10 de janeiro de 2017 [citado 29 de setembro de 2025];19(4):575. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6054162/>

90. Feng H, Mancuso N, Gusev A, Majumdar A, Major M, Pasaniuc B, et al. Leveraging expression from multiple tissues using sparse canonical correlation analysis and aggregate tests improves the power of transcriptome-wide association studies. *PLoS Genetics* [Internet]. 8 de abril de 2021 [citado 29 de setembro de 2025];17(4):e1008973. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8057593/>

91. Gandal MJ, Leppa V, Won H, Parikshak NN, Geschwind DH. The road to precision psychiatry: translating genetics into disease mechanisms. *Nature neuroscience* [Internet]. 26 de outubro de 2016 [citado 29 de setembro de 2025];19(11):1397. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC9012265/>

92. 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* [Internet]. 19 de janeiro de 2024;27(1):108662. Disponível em: <http://dx.doi.org/10.1016/j.isci.2023.108662>

93. Fukutani ER, Ramos PIP, Kasprzykowski JI, Azevedo LG, Rodrigues MM de S, Lima JV de OP, 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* [Internet]. 8 de novembro de 2019 [citado 29 de setembro de 2025];10:477865. Disponível em: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.01056/pdf>

94. Fukutani KF, Kasprzykowski JI, Paschoal AR, Gomes M de S, Barral A, de Oliveira CI, et al. Meta-Analysis of *Aedes aegypti* Expression Datasets: Comparing Virus Infection and Blood-Fed Transcriptomes to Identify Markers of Virus Presence.

Front Bioeng Biotechnol [Internet]. 11 de janeiro de 2018 [citado 29 de setembro de 2025];5:323624. Disponível em: <https://www.frontiersin.org/journals/bioengineering-and-biotechnology/articles/10.3389/fbioe.2017.00084/pdf>

95. Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* [Internet]. 15 de julho de 2007;23(14):1846–7. Disponível em: <http://dx.doi.org/10.1093/bioinformatics/btm254>

96. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* [Internet]. 29 de dezembro de 2008 [citado 27 de maio de 2025];9(1):1–13. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-559>

97. Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics--a bioconductor package for quality assessment of microarray data. *Bioinformatics* [Internet]. 1º de fevereiro de 2009;25(3):415–6. Disponível em: <http://dx.doi.org/10.1093/bioinformatics/btn647>

98. Subirana I, Sanz H, Vila J. Building Bivariate Tables: ThecompareGroupsPackage for R. *J Stat Softw* [Internet]. 2014;57(12). Disponível em: <http://www.jstatsoft.org/v57/i12/>

99. 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* [Internet]. dezembro de 2009;14(12):1083–94. Disponível em: <http://dx.doi.org/10.1038/mp.2009.18>

100. 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* [Internet]. 17 de julho de 2018;13(7):e0200003. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0200003>

101. 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* [Internet]. 23 de maio de 2019;9(1):151. Disponível em: <http://dx.doi.org/10.1038/s41398-019-0492-8>

102. 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* [Internet]. 15 de março de 2012;28(6):882–3. Disponível em: <http://dx.doi.org/10.1093/bioinformatics/bts034>

103. Differential age- and disease-related effects on the expression of genes related to the arachidonic acid signaling pathway in schizophrenia. *Psychiatry Research* [Internet]. 30 de abril de 2012 [citado 27 de maio de 2025];196(2-3):201–6. Disponível em: <http://dx.doi.org/10.1016/j.psychres.2011.09.026>

104. 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* [Internet]. março de 2019;180(2):138–49. Disponível em: <http://dx.doi.org/10.1002/ajmg.b.32691>

105. 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* [Internet]. 27 de junho de 2012;7(6):e39498. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0039498>

106. 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* [Internet]. 29 de fevereiro de 2012;7(2):e32618. Disponível em: <http://dx.doi.org/10.1371/journal.pone.0032618>

107. 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* [Internet]. 15 de outubro de 2017;82(8):594–600. Disponível em: <http://dx.doi.org/10.1016/j.biopsych.2017.03.018>

108. 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* [Internet]. janeiro de 2021;2(1):sgab002. Disponível em: <http://dx.doi.org/10.1093/schizbullopen/sgab002>

109. 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* [Internet]. 3 de maio de 2019;10(1):2046. Disponível em: <http://dx.doi.org/10.1038/s41467-019-09786-7>

110. 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* [Internet]. 9 de julho de 2019;20(1):135. Disponível em: <http://dx.doi.org/10.1186/s13059-019-1747-7>

111. 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* [Internet]. 20 de abril de 2015;43(7):e47. Disponível em: <http://dx.doi.org/10.1093/nar/gkv007>

112. Liaw A, Wiener M. Classification and regression by randomForest. *R News* [Internet]. 1º de dezembro de 2002 [citado 27 de maio de 2025];2(3):18–22. Disponível em: <https://journal.r-project.org/articles/RN-2002-022/>

113. Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw* [Internet]. 2008;28(5). Disponível em: <http://www.jstatsoft.org/v28/i05/>

114. 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* [Internet]. 17 de março de 2011;12:77. Disponível em: <http://dx.doi.org/10.1186/1471-2105-12-77>

115. Durinck S, Spellman PT, Birney E, Huber W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* [Internet]. 23 de julho de 2009;4(8):1184–91. Disponível em: <http://dx.doi.org/10.1038/nprot.2009.97>
116. 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. *Innovation (Camb)* [Internet]. 28 de agosto de 2021;2(3):100141. Disponível em: <http://dx.doi.org/10.1016/j.xinn.2021.100141>
117. Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome Pathway Knowledgebase 2024. *Nucleic Acids Res* [Internet]. 5 de janeiro de 2024;52(D1):D672–8. Disponível em: <http://dx.doi.org/10.1093/nar/gkad1025>
118. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* [Internet]. 1º de janeiro de 2000;28(1):27–30. Disponível em: <http://dx.doi.org/10.1093/nar/28.1.27>
119. Levin HS, Eisenberg HM, Benton AL. *Frontal Lobe Function and Dysfunction* [Internet]. 1991. 460 p. Disponível em: https://books.google.com/books/about/Frontal_Lobe_Function_and_Dysfunction.html?hl=&id=9b1htO0V0rwC
120. Hoftman GD, Datta D, Lewis DA. Layer 3 Excitatory and Inhibitory Circuitry in the Prefrontal Cortex: Developmental Trajectories and Alterations in Schizophrenia. *Biol Psychiatry* [Internet]. 15 de maio de 2017;81(10):862–73. Disponível em: <http://dx.doi.org/10.1016/j.biopsych.2016.05.022>
121. Jarskog LF, Glantz LA, Gilmore JH, Lieberman JA. Apoptotic mechanisms in the pathophysiology of schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* [Internet]. junho de 2005;29(5):846–58. Disponível em: <http://dx.doi.org/10.1016/j.pnpbp.2005.03.010>
122. Gorman AM. Neuronal cell death in neurodegenerative diseases: recurring themes around protein handling. *J Cell Mol Med* [Internet]. dezembro de 2008;12(6A):2263–80. Disponível em: <http://dx.doi.org/10.1111/j.1582-4934.2008.00402.x>
123. Su C, Tong J, Wang F. Mining genetic and transcriptomic data using machine learning approaches in Parkinson’s disease. *npj Parkinson’s Disease* [Internet]. 9 de setembro de 2020 [citado 27 de maio de 2025];6(1):1–10. Disponível em: <https://www.nature.com/articles/s41531-020-00127-w>
124. Lewis DA, Glausier JR. Alterations in Prefrontal Cortical Circuitry and Cognitive Dysfunction in Schizophrenia. *Nebr Symp Motiv* [Internet]. 2016;63:31–75. Disponível em: http://dx.doi.org/10.1007/978-3-319-30596-7_3
125. 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* [Internet]. abril de 2015;72(4):316–24. Disponível em: <http://dx.doi.org/10.1001/jamapsychiatry.2014.2414>

126. 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* [Internet]. junho de 2008;10(6):643–53. Disponível em: <http://dx.doi.org/10.1038/ncb1727>

127. 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* [Internet]. 29 de abril de 2016;6(4):e009537. Disponível em: <http://dx.doi.org/10.1136/bmjopen-2015-009537>

128. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. *Nature* [Internet]. 13 de fevereiro de 2014;506(7487):179–84. Disponível em: <http://dx.doi.org/10.1038/nature12929>

129. 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* [Internet]. janeiro de 2018;26(1):64–74. Disponível em: <http://dx.doi.org/10.1038/s41431-017-0038-6>

130. 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* [Internet]. 20 de outubro de 2015;5(10):e663. Disponível em: <http://dx.doi.org/10.1038/tp.2015.119>

131. Davaliev K, Maleva Kostovska I, Dwork AJ. Proteomics Research in Schizophrenia. *Front Cell Neurosci* [Internet]. 16 de fevereiro de 2016;10:18. Disponível em: <http://dx.doi.org/10.3389/fncel.2016.00018>

132. 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* [Internet]. 8 de janeiro de 2021;49(D1):D862–70. Disponível em: <http://dx.doi.org/10.1093/nar/gkaa1064>

133. Urade Y. Biochemical and Structural Characteristics, Gene Regulation, Physiological, Pathological and Clinical Features of Lipocalin-Type Prostaglandin D2 Synthase as a Multifunctional Lipocalin. *Front Physiol* [Internet]. 22 de outubro de 2021 [citado 29 de setembro de 2025];12:718002. Disponível em: <https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.718002/pdf>

134. Fu R, Shen Q, Xu P, Luo JJ, Tang Y. Phagocytosis of Microglia in the Central Nervous System Diseases. *Molecular Neurobiology* [Internet]. 7 de janeiro de 2014 [citado 29 de setembro de 2025];49(3):1422–34. Disponível em: <https://link.springer.com/article/10.1007/s12035-013-8620-6>

135. Tremblay MÈ, Stevens B, Sierra A, Wake H, Bessis A, Nimmerjahn A. The Role of Microglia in the Healthy Brain. *J Neurosci* [Internet]. 9 de novembro de 2011 [citado 29 de setembro de 2025];31(45):16064–9. Disponível em: <https://www.jneurosci.org/content/31/45/16064.abstract>

136. Wang J, Kochunov P, Sampath H, Hatch KS, Ryan MC, Xue F, et al. White

matter brain aging in relationship to schizophrenia and its cognitive deficit. *Schizophr Res* [Internet]. abril de 2021;230:9–16. Disponível em: <http://dx.doi.org/10.1016/j.schres.2021.02.003>

137. 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* [Internet]. novembro de 2015;72(11):1095–101. Disponível em: <http://dx.doi.org/10.1001/jamapsychiatry.2015.1546>

138. 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. *Nature Communications* [Internet]. 21 de março de 2017 [citado 29 de setembro de 2025];8:14774. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC5364411/>

139. Corrêa BB, Xavier M, Guimarães J. Association of Huntington's disease and schizophrenia-like psychosis in a Huntington's disease pedigree. *Clinical Practice and Epidemiology in Mental Health : CP & EMH* [Internet]. 15 de fevereiro de 2006 [citado 29 de setembro de 2025];2:1. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC1386660/>

140. Thompson A, MacKay A, Rudge P, Lukic A, Porter MC, Lowe J, et al. Behavioral and Psychiatric Symptoms in Prion Disease. *American Journal of Psychiatry* [Internet]. 1º de março de 2014 [citado 29 de setembro de 2025]; Disponível em: <https://psychiatryonline.org/doi/10.1176/appi.ajp.2013.12111460>

141. 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. *Movement disorders : official journal of the Movement Disorder Society* [Internet]. junho de 2021 [citado 29 de setembro de 2025];36(6). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33405293/>

142. Ben-Shachar D. The bimodal mechanism of interaction between dopamine and mitochondria as reflected in Parkinson's disease and in schizophrenia. *Journal of neural transmission (Vienna, Austria : 1996)* [Internet]. fevereiro de 2020 [citado 29 de setembro de 2025];127(2). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31848775/>

143. Ali S, Santomauro D, Ferrari AJ, Charlson F. Schizophrenia as a risk factor for cardiovascular and metabolic health outcomes: a comparative risk assessment. *Epidemiology and psychiatric sciences* [Internet]. 9 de fevereiro de 2023 [citado 29 de setembro de 2025];32. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36756905/>

144. 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* [Internet]. 2019;95(5):179–97. Disponível em: <http://dx.doi.org/10.2183/pjab.95.014>

145. Park H, Poo MM. Neurotrophin regulation of neural circuit development and function. *Nat Rev Neurosci* [Internet]. janeiro de 2013;14(1):7–23. Disponível em: <http://dx.doi.org/10.1038/nrn3379>

146. Biogenesis and Function of the NGF/TrkA Signaling Endosome. Em: International Review of Cell and Molecular Biology [Internet]. Academic Press; 2015 [citado 27 de maio de 2025]. p. 239–57. Disponível em: <http://dx.doi.org/10.1016/bs.ircmb.2014.10.002>
147. Luza S, Opazo CM, Bousman CA, Pantelis C, Bush AI, Everall IP. The ubiquitin proteasome system and schizophrenia. The lancet Psychiatry [Internet]. junho de 2020 [citado 29 de setembro de 2025];7(6). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/32061320/>
148. Surya A, Bolton BM, Rothe R, Mejia-Trujillo R, Leonita A, Zhao Q, et al. Differential impacts of ribosomal protein haploinsufficiency on mitochondrial function. The Journal of cell biology [Internet]. 3 de março de 2025 [citado 29 de setembro de 2025];224(3). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/39786340/>
149. Roberts RC. Mitochondrial dysfunction in schizophrenia: With a focus on postmortem studies. Mitochondrion [Internet]. janeiro de 2021 [citado 29 de setembro de 2025];56. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33221354/>
150. 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. Psychiatric genetics [Internet]. 1º de fevereiro de 2021 [citado 29 de setembro de 2025];31(1). Disponível em: <https://pubmed.ncbi.nlm.nih.gov/33395218/>
151. 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. Journal of psychiatric research [Internet]. outubro de 2023 [citado 29 de setembro de 2025];166. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/37659266/>

ANEXO

Anexo I – Artigo publicado

Rebouças et al. *BMC Psychiatry* (2025) 25:1177
<https://doi.org/10.1186/s12888-025-07524-6>

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
arturqueiroz@iocruz.br

¹Instituto de Pesquisa Clínica e Translacional (IPCT), Faculdade Zarns, Clariens Educação, Salvador, Bahia, Brazil

²Multinational Organization Network Sponsoring Translational and Epidemiological Research (MONSTER) Initiative, Salvador, Brazil

³Escola Bahiana de Medicina e Saúde Pública (EBMSP), Salvador 40290-150, Brazil

⁴Transplant / Oncology & Infectious Disease Program, School of Medicine, Johns Hopkins University, 2000 East Monument St, Baltimore, MD 21205, USA

⁵Laboratório 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 post-mortem 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 have 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 "*absMax-Mean*". 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 also 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 $|\log_{2}FC| > 1$ 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 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' 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

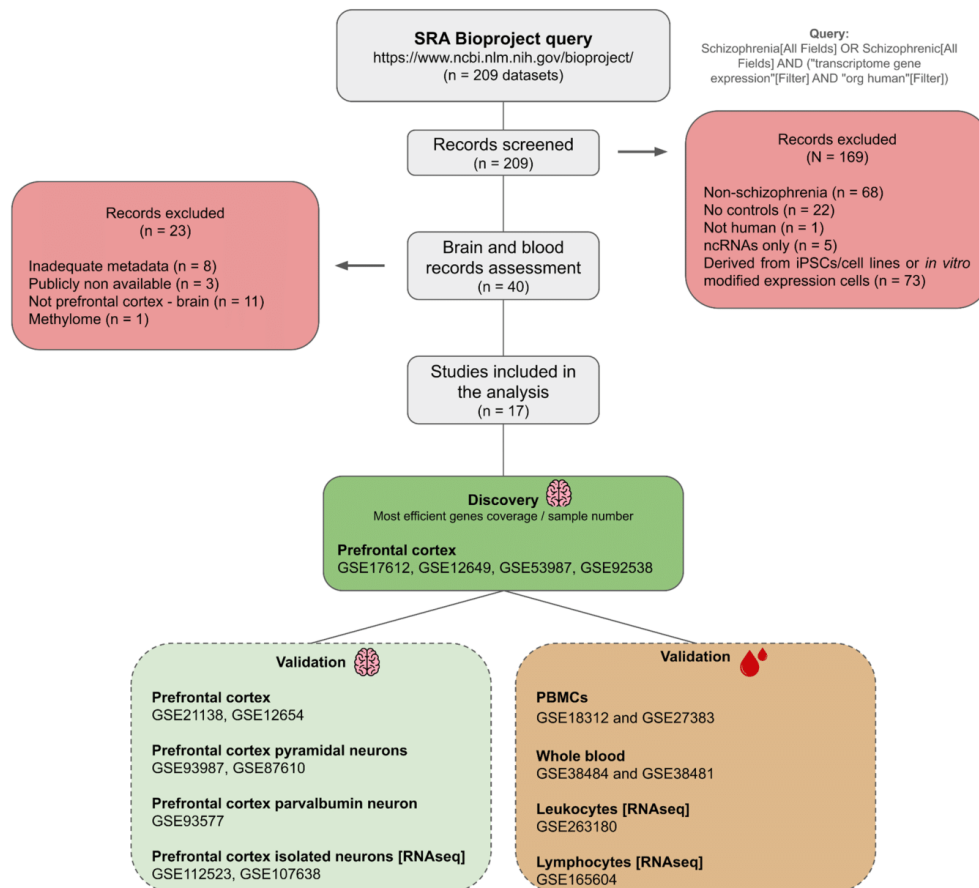


Fig. 1 Sampling flow chart. Flow chart depicting the sampling process for this Individual participant data meta-analysis, with details about exclusion criteria and numbers of excluded/included studies

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

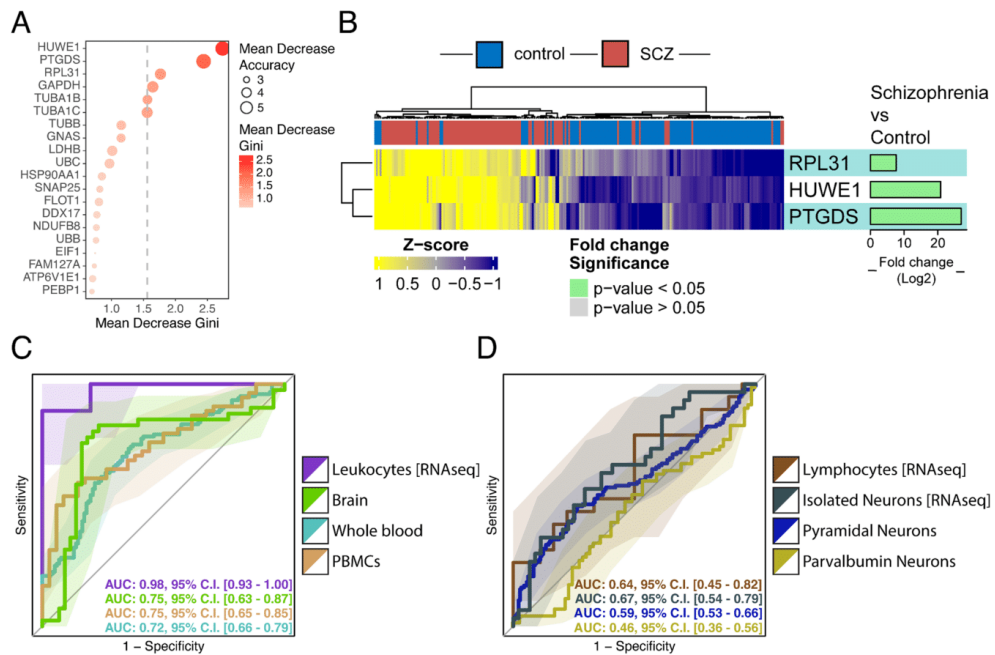


Fig. 2 Most informative genes identification and discriminative performance assessment. **A** Dot plot displaying the random forest's importance values used to select the most informative genes to separate SCZ from control samples. *Mean decrease accuracy* is related to the models' accuracy drop when 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

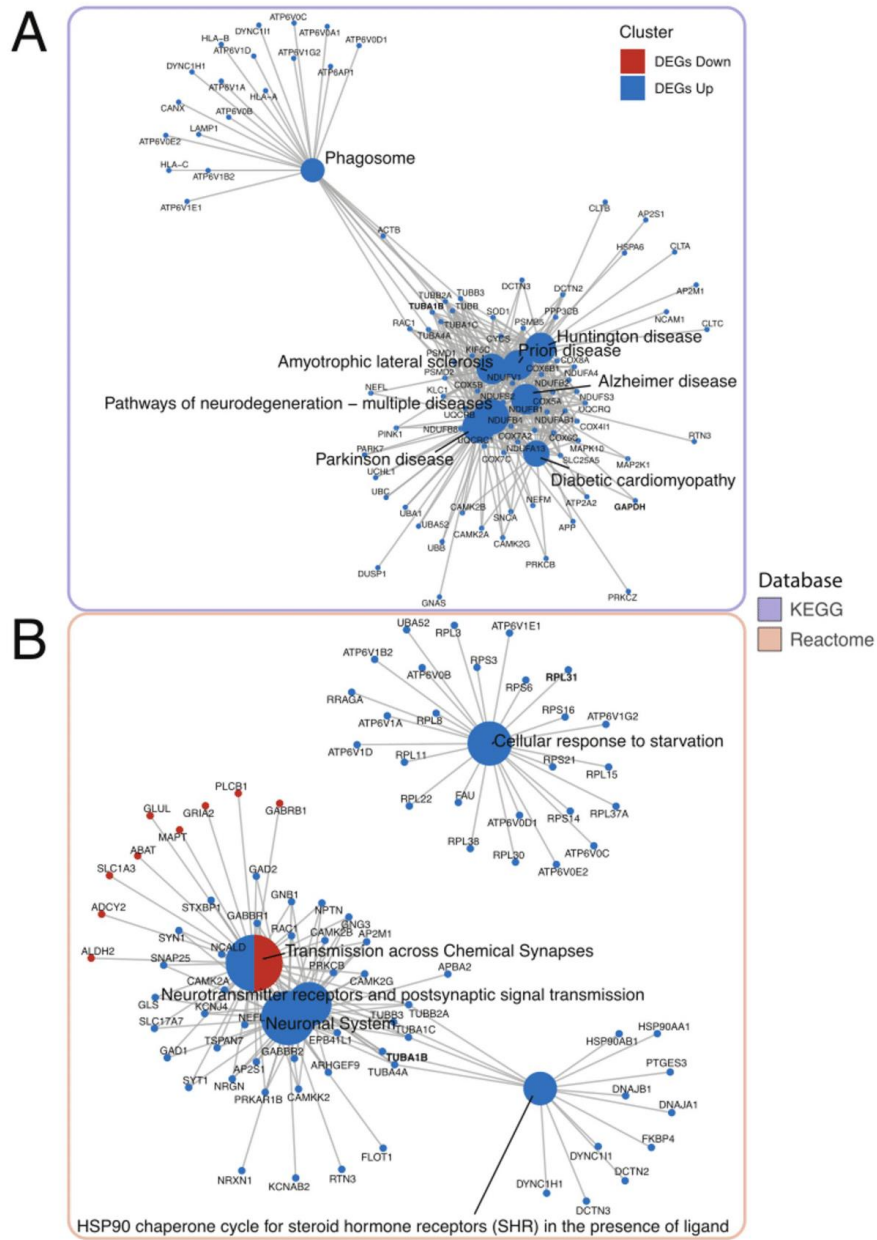


Fig. 3 Network of the most informative genes in the enrichment analysis. Enriched pathways in both KEGG and Reactome databases. The network displays the pathways that were enriched using the up and downregulated DEGs as input, highlighting the pathways containing genes selected by the random forest. 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

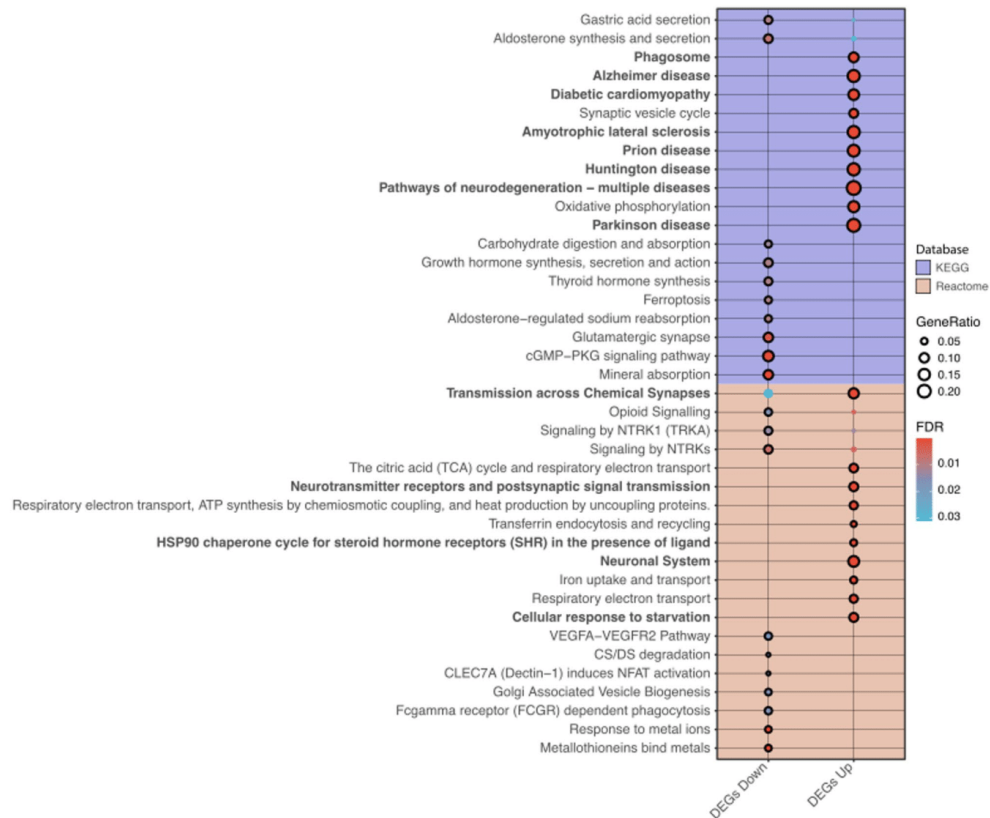


Fig. 4 Dot plot for the enrichment analysis top pathways. Enriched pathways in both KEGG and Reactome databases. The dot plot displays the pathways that were enriched using the up and downregulated DEGs as input. The dot size represents the pathways' gene ratio, while the color represents its P-value 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 "Fcγ 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
CI.	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

- 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.
- Schizophrenia [Internet]. [cited 2024 Mar 5]. Available from: <https://www.who.int/news-room/fact-sheets/detail/schizophrenia>
- Keshavan MS, Collin G, Guimond S, Kelly S, Prasad KM, Lizano P. Neuroimaging in schizophrenia. *Neuroimaging Clin N Am*. 2020;30(1):73–83.
- Fischer BA, Carpenter WT Jr. Will the Kraepelinian dichotomy survive DSM-V? *Neuropsychopharmacology*. 2009;34(9):2081–7.
- 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.
- Levin HS, Eisenberg HM, Benton AL. Frontal lobe function dysfunct. 1991. 460 p.
- 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.
- 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.
- 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.
- Lai CY, Scarr E, Udawela M, Everall I, Chen WJ, Dean B. Biomarkers in schizophrenia: A focus on blood based diagnostics and therapeutics. *World J Psychiatry*. 2016;6(1):102–17.
- 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 JJ, 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 JJ, 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. Post-mortem 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. Hofman 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, Glautier 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 JT, Guardavaccaro D, Jiang R, Pagano M, Guillemot F, et al. The HECT-domain ubiquitin ligase Hwul1 controls neural differentiation and proliferation by destabilizing the N-Myc oncogene. *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. Davaliev 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 ME, 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.