scientific reports

Check for updates

OPEN

Gene network analysis identifies dysregulated pathways in an autism spectrum disorder caused by mutations in Transcription Factor 4

Lucas M. de Carvalho[1,2]✉, Vinicius M. A. Carvalho[1,6], Antonio P. Camargo[1,3]✉ & Fabio Papes[1,4,5]✉

Pitt–Hopkins syndrome (PTHS) is a rare neurodevelopmental monogenic disorder in the autistic spectrum caused by mutations in the Transcription Factor 4 gene. Even though the genetic etiology is known, the molecular mechanisms underlying PTHS remain poorly understood. To gain insight into the disease's pathophysiology, we set out to identify genes and pathways putatively involved in the pathology through co-expression and gene hub analyses using transcriptomic data from neural progenitor cells, neurons, and brain organoids derived from PTHS patients. Our results revealed several groups of co-expressed genes that are differentially regulated in PTHS neural cells compared to controls. These groups were enriched for genes involved in neural development and function, including synaptic transmission, membrane excitability, and cell adhesion. We identified several hub genes (highly connected nodes within gene networks that are central in these modules), including some that encode proteins involved in histone modification, synaptic vesicle trafficking, and cell signaling. Furthermore, we found that the differential expression of hub genes in PTHS neural cells was associated with altered cellular processes linked to neurodevelopment, such as cell–cell communication and irregular synaptic networks. Notably, we identified a set of hub genes related to the histone gene family, which is associated with neuronal differentiation and may contribute to PTHS pathogenesis and potentially serve as a biomarker for disease prognosis. Our results support the notion that PTHS involves alterations in neural development and function, particularly in excitatory neurons. The groups of co-expressed genes and hub genes we identified provide new insights into the molecular mechanisms underlying PTHS pathogenesis and could potentially be targeted for therapeutic intervention.

Keywords Pitt–Hopkins syndrome, Autism, Co-expression analysis, Hub genes, Interactome

Pitt–Hopkins syndrome (PTHS) is a rare neurodevelopmental disorder caused by mutations in the transcription factor gene *TCF4*, resulting in a range of phenotypic anomalies. These include distinctive facial features and neuropsychiatric abnormalities such as profound cognitive impairment, motor delay, and delayed language development, which collectively place this disease in the autistic spectrum[1]. Despite its significant impact on affected individuals and their families, the underlying molecular mechanisms of PTHS remain poorly understood.

In recent years, high-throughput omics technologies, such as RNA-Seq and single-cell RNA-Seq, have provided a wealth of data that can be used to gain insight into the pathophysiology of complex diseases. In

[1]Department of Genetics, Evolution, Microbiology, and Immunology, Institute of Biology, University of Campinas, Campinas, Sao Paulo 13083-862, Brazil. [2]Sao Francisco University, Bragança Paulista, Sao Paulo 12916-900, Brazil. [3]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Department of Pediatrics, School of Medicine, University of California San Diego, La Jolla, CA 92037, USA. [5]Center for Medicinal Chemistry, University of Campinas, Campinas, Sao Paulo 13083-886, Brazil. [6]Present address: Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA 94305, USA. ✉email: lucas.miguel@usf.edu.br; antoniop.camargo@lbl.gov; papesf@unicamp.br

particular, co-expression network analysis has emerged as a powerful tool for identifying groups of co-expressed genes (modules), which highlight functionally related gene sets, and for investigating their relationships with disease phenotypes[2]. By grouping co-expressed genes, researchers can infer functional relationships and find potential biomarkers or therapeutic targets in disease pathways. Furthermore, the use of network analysis to identify potential biomarkers has significantly improved our knowledge about several diseases that affect the nervous system—including Alzheimer's disease, autism, and viral infection[3–12].

In this study, we aimed to shed light onto the molecular mechanisms underlying PTHS pathophysiology by performing protein–protein interaction (PPI) network and gene co-expression analyses on RNA-Seq data from neurons and neural progenitor cells (NPC) obtained from PTHS patients and their respective parents (controls). We chose these types of analyses because they allow us to map gene interactions on multiple levels, providing a more comprehensive understanding of how different genes, including hub genes, may contribute to disease phenotypes. First, we predicted a putative set of molecular interactions (interactome) for each neural cell type (NPCs or neurons) to identify interactions not observed in regular differential gene expression analyses. Subsequently, we searched for submodules within the network with significantly distinguishable gene expression patterns between PTHS patients and parental controls. We also identified modules of co-expressed genes and investigated their associations with different aspects of neuronal differentiation and function.

By leveraging the power of high-throughput omics technologies, PPI network analysis, and gene co-expression analysis, we found previously unknown genetic and molecular interactions that could shed light onto the molecular mechanisms underlying PTHS and support further investigations on the pathophysiology of this disease. Ultimately, our findings may pave the way for the development of novel therapies and treatment to improve the lives of those affected by Pitt–Hopkins Syndrome.

Materials and methods

RNA-Seq data acquisition

RNA-seq data from samples obtained from PTHS patient and their respective parental controls were previously generated[13] and deposited in the Gene Expression Omnibus (GEO) database (accessions: GSE159392, GSE159860, and GSE189121). Results relative to differentially expressed genes (DEGs) in these samples were based on the data presented in our previous study[13].

Interactome generation and submodule identification

We obtained the differentially expressed genes (DEGs) for each neural cell type, representing two stages of neural differentiation (NPCs and neurons) from the list of DEGs common to all parent–child comparisons through differential expression analysis. To build the interactome, both lists were submitted to the STRING v11 database[14], yielding score values that represent the reliability of the inferred interactions. We set the minimum score threshold to 0.9 (in a scale from 0 to 1), to build a network of highly reliable interactions.

The Molecular Complex Detection (MCODE)[15] plug-in for Cytoscape[16] software was used to identify highly interconnected regions in the protein–protein interaction networks. Such regions represent potential molecular complexes and can be leveraged to extract key subnetworks. The following parameters were used: *degree cutoff* = 2, *node score cutoff* = 0.2, *node density cutoff* = 0.1, *Max depth* = 100, *K-core* = 2, and *cutoff score* > 5.

Statistical test

We employed hypergeometric tests to assess the statistical significance of the enrichment of a given feature in the protein–protein networks. This test allows the calculation of the probability of observing a certain number of nodes with high expression per patient cell in each network, given the total number of nodes in each network and the total number of nodes in both networks combined.

Construction of the gene co-expression network

Clusters of co-expressed genes (modules) were identified using the Weighted Gene Co-expression Network Analysis (WGCNA, version 1.72-1) R library[17]. First, the gene expression matrix was filtered to remove lowly expressed genes and samples with high missing values using the *goodSampleGene* function, which identifies samples and genes with too many missing entries and genes with zero variance, and a soft-thresholding power was selected using the scale-free topology criterion. Co-expression modules were identified using the *blockwiseModules* function, with a minimum module size of 30[18]. The module eigengene (ME) was calculated for each module, and modules with high correlation coefficients (|correlation| > 0.9) were merged.

To identify key co-expression modules, we applied a gene significance (GS) cut-off and a module membership (MM) cut-off. GS is a measure of the association between gene expression and the neural cell type (NPC or neuron), and MM is a measure of the correlation between gene expression and module eigengene expression (*signedKME* function). We used a GS cut-off of 0.2, such that only genes with a strong correlation with the phenotype were included in the analysis. We also used a MM cut-off of 0.8, meaning that genes with a high correlation with the module eigengene were considered to be highly connected within the module. Moreover, the hub gene identification per module was based on high MM membership (MM > 0.9).

Functional enrichment analysis

Functional enrichment analysis was performed using the clusterProfiler v4.6.2 package[19] to identify the biological processes and pathways associated with each module. The significantly enriched Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG)[20] pathways (Benjamini–Hochberg FDR $\leq$ 0.05) were selected for downstream analysis.

Single-cell analysis

Single-cell RNA sequencing data was processed using the Seurat package v4.0[21]. The violin plots and statistical tests were produced using the Seurat object available in our publication[13]. The Wilcoxon rank-sum test was employed to determine the significance of gene expression differences between clusters.

Jaccard index calculation

To evaluate the overlap between gene sets, we calculated the Jaccard index, a statistical measure of similarity that quantifies the proportion of shared elements between two sets. It is defined as the size of the intersection divided by the size of the union of the two sets.

In this study, the sets were defined as follows: (1) a reference gene set of interest (e.g., disease-associated genes, pathway-specific genes, or other curated gene lists) and (2) the genes identified in the NPC or neuron interactomes. For each comparison, the Jaccard index ($0 \leq J(A,B) \leq 1$) was computed using Eq. (1):

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

where A represents the reference gene set, and B represents the genes in the respective interactome. The numerator $|A \cap B|$ indicates the number of genes common to both sets, while the denominator $|A \cup B|$ represents the total number of unique genes in either set. This analysis was performed for various gene sets against the NPC and neuron interactomes.

Results and discussion

Analysis of the PTHS interactome reveals dysregulation of key genes involved in brain development and neuronal function

Previously published lists of DEGs between PTHS individuals and their parents provide limited insight into the molecular mechanisms underlying the disease, as they lack evidence for the physical interactions between dysregulated proteins or the genetic interaction relationships between upstream and downstream DEGs. Causal relationships cannot be established without such information, highlighting the need for further investigation using a systems biology approach to uncover the underlying molecular pathways involved in PTHS pathogenesis. To generate such interactomes, we utilized the STRING database, which offers a comprehensive and up-to-date resource of known and predicted protein interactions. We constructed two interactomes (for NPCs and neurons) by adding edges between DEGs with high interaction confidence scores (Supplementary Material 1).

The resulting NPC interactome contained 325 nodes and 504 edges, while the neuronal interactome exhibited 673 nodes and 1897 edges. Inspection of the NPC network revealed that it is enriched for genes exhibiting higher expression in PTHS individuals when compared to their corresponding parents (red nodes in Fig. 1; hypergeometric test *p*-value: 5.05e−34; over-enrichment: 1.53-fold), while the neuron interactome displayed more genes that are downregulated in the patients (blue nodes in Fig. 2; *p*-value: 7.58e−49; under-enrichment: 1.41-fold). This pattern was reversed when we evaluated DEGs that are more highly expressed in the parents, as these genes are significantly overrepresented in the neuron interactome (hypergeometric test *p*-value: 9.88e−43; over-enrichment: 1.27-fold) and underrepresented in the NPC interactome (*p*-value: 1.18e−35; under-enrichment: 2.23-fold). These findings suggest that, in the context of PTHS, the genes present in these two networks exhibit contrasting gene expression patterns.

Within the NPC interactome, we found a set of interacting genes that are involved in cell–cell communication and signaling (Fig. 1A). This group includes several genes coding for thrombospondin genes (*THBS2*, *THSD4,* and *THSD1*), some of which are upregulated (blue nodes in Fig. 1C) while others are downregulated in PTHS (red nodes). Atypical expression of these genes has been linked to cortical dysfunction in neurological disorders characterized by irregular synaptic networks[22]. The same thrombospondin genes were also found to be hub genes in the NPC interactome (Supplementary Material 8—Fig. S1 and Table S1), and have been shown to be biologically important for neuronal function[23]. The fact that these genes were identified as hub genes (and are, therefore, central to maintaining the network stability and structure) suggests that their dysregulation could result in compounded genetic effects and thus in a wide range of disease phenotypes. The high-score interactions between thrombospondin genes we observed are based on the curated Reactome database[24], as shown by the score associated with the STRING platform (Supplementary Material 1). The interactions among these components are based on the *O*-glycosylation pathway of TSR domain-containing proteins, which regulates many biological processes such as Notch signaling, inflammation, wound healing, angiogenesis, and neoplasia[25]. In particular, dysregulation of thrombospondin expression may imply aberrant neural progenitor cell cycle regulation or proliferation, which is in keeping with our previously published observation that PTHS neural progenitor cells have diminished proliferative activity[13].

Within the NPC interactome, we also found an interaction group that encompasses two sets of dysregulated genes investigated in our previous study[13]: CDKNs (cyclin-dependent kinase inhibitors, e.g., *CDKN2A* and *CDKN1A*) and *TFAP2C/TFAP2A* (Fig. 1B). CDKNs were investigated in the context of replicative senescence and found to be upregulated in the slowly-dividing, senescent neural progenitor cells of PTHS individuals. This is in accordance with the known general role of *CDKN* components as cyclin inhibitors that interrupt the cell cycle and prevent the progenitors from dividing. Interestingly, other studies observed that TFAP transcription factors regulate the expression of *CDKN* genes[26,27]. Given that both CDKNs and TFAP2C/TFAP2A are down-regulated in the interactome, this co-dysregulation suggests a coordinated disruption in their regulatory pathway, which

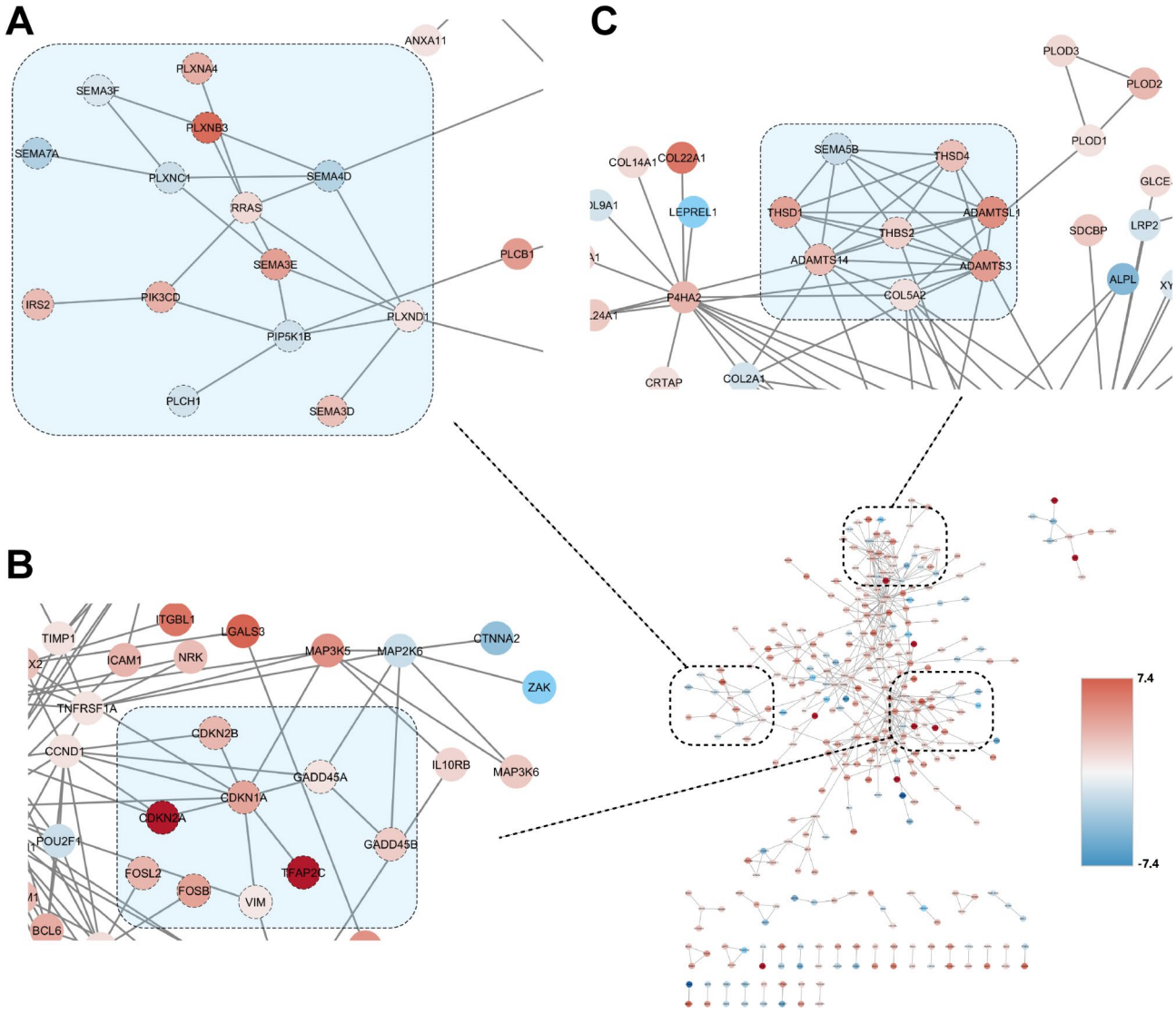


Fig. 1. Interactome constructed based on DEGs obtained from NPCs of PTHS patients. The nodes represent genes, while the edges represent protein–protein interactions. The colors represent the log2 fold-change (log2FC) of the genes between the group of PTHS individuals and the parents group (control). Genes upregulated in the PTHS samples are shown in red, while downregulated genes are shown in blue. Three sectors of the interactome map are shown at higher magnification to reveal the identities of relevant genes: (**A**) Semaphorin genes and their Plexin receptors; (**B**) CDKN genes and components involved in cell cycle regulation; (**C**) Thrombospondin genes involved in chemotactic signaling.

could explain why patient cells stop proliferating and differentiate early[13], a phenomenon that could be further investigated in future studies.

In the neuron interactome, some nodes that are crucial in the network, due to the high number of connections they make with other genes (Fig. 2), are related to the function of excitatory neurons, which our previous study showed to be present in lower numbers in the developing neural tissue of brain organoids derived from PTHS patients and in the cortical tissue of a PTHS post-mortem sample[13]. These genes encode ion channels and proteins relevant for neurotransmitter and synaptic function; while the function of these genes in excitatory neuron physiology has been widely investigated, the significance of its dysregulation in the context of Pitt–Hopkins syndrome is relevant as it may explain the loss of excitatory neurons in the developing cortex of these individuals and provides a new framework for understanding the pathological mechanisms and for identifying targets for future pharmacological intervention.

A potentially relevant gene hub is composed of synaptosome-associated protein 25 (*SNAP25*), synaptotagmin 1 (*SYT1*), and synapsin III (*SYN3*), as well as genes associated with synaptic vesicles, such as vesicle-associated membrane protein (*VAMP*) family members (Fig. 2B). These genes maintain significant connections with many nodes clustered in the central part of the network, all of which are involved in neurotransmission. These results stress the value of performing interactome network analysis, as the identification of such interaction relationships

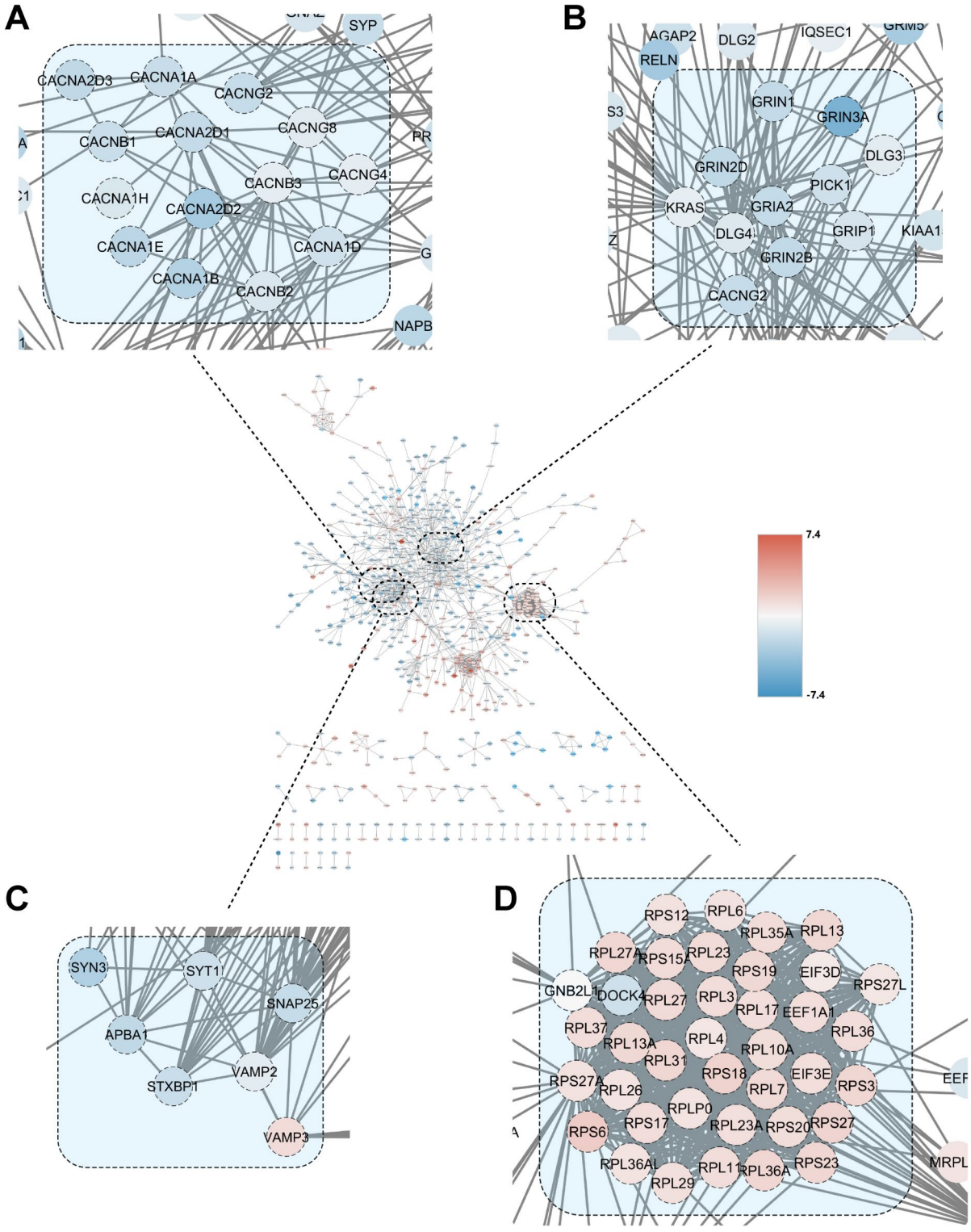


Fig. 2. Interactome network map constructed based on differentially expressed genes (DEGs) obtained from neuronal cells of Pitt–Hopkins Syndrome (PTHS) patients. Nodes represent genes, whereas edges represent protein–protein interactions. Four groups of genes are highlighted in each panel: (**A**) glutamate receptor subunits; (**B**) synapse-related genes; (**C**) ion channels; (**D**) ribosomal proteins. The colors represent the log2 fold-change (log2FC) of the genes between the group of PTHS individuals and the parents group (control). Genes upregulated in PTHS patients are shown in red, while downregulated genes are shown in blue.

between DEGs cannot be easily compiled from lists of differentially expressed genes. The *SNAP25* gene, which codes for a presynaptic plasma membrane protein involved in regulating neurotransmitter release, displayed higher expression in parent cells than in PTHS cells in the glutamatergic (Glut) lineage of iPSC-derived brain cortical organoids, while the *SYT1* and *SYN3* genes exhibited no difference, according to pseudotime single-cell RNA-Seq analysis (Supplementary Material 6—Fig. 1S). The lower *SNAP25* expression may indicate that PTHS patients have a deficiency in neuronal impulse propagation resulting from synaptic vesicle dysregulation. Importantly, the interaction between dysregulated synapse-related genes provides a putative mechanistic explanation for the altered electrical activity phenotypes observed in the neurons of PTHS patients[13] and may provide a new avenue for identifying target genes of interest for pharmacological study. Notably, another cluster in the neuron interactome network contains genes encoding ion channels (Fig. 2A), providing additional insight into the underlying pathology of the altered neuronal electrical activity in PTHS.

Lastly, we performed enrichment analysis of KEGG pathways in the neuronal interactome. Noteworthy pathways related to synaptic function or transmission include axon guidance (FDR = 6.99e−08; hsa04360), calcium signaling pathway (FDR = 1.80e−05; hsa04020), cAMP signaling pathway (FDR = 3.62e−05; hsa04024), dopaminergic synapse (FDR = 1.95e−06; hsa04728), GABAergic synapse (FDR = 1.10e−05; hsa04727), glutamatergic synapse (FDR = 3.45e−12; hsa04724), and synaptic vesicle cycle (FDR = 5.43e−06; hsa04721). Among all genes involved in these processes, we highlight the G protein subunit γ^3 (*GNG3)* gene, which exhibits higher expression in parent cells than in PTHS, in both Glut and GABA lineages of organoids, according to pseudotime analysis in single-cell transcriptomic data (Supplementary Material 6—Fig. 2S). This gene is a member of dopaminergic and glutamatergic synapse pathways and has been associated as a hub gene in the interactome of neuronal cells derived from Parkinson's disease patients[28]. Given its role as a modulator or transducer in various transmembrane signaling systems, we hypothesize *GNG3* may be linked to reduced signaling in PTHS neurons.

The overlap between the genes found in the NPC interactome and the neuron interactome was 55 genes, resulting in a Jaccard index of 0.0588, indicating that, although there is some overlap between the two interactomes, the intersection is relatively small. These results suggest that the profiles of dysregulated genes in the two stages of neural differentiation—NPCs and neurons—is dissimilar and that the phenotypes observed in PTHS individuals may be the summation of alterations in distinct cell types due to largely non-overlapping sets of aberrant molecular pathways.

Interactome submodule association

Submodule identification provides clusters of highly connected genes that could be used as markers associated with PTHS progression or severity and/or provide insight into potential therapeutic targets for the disease. Submodules often represent functional clusters of proteins or genes that work together in specific biological processes, allowing researchers to detect key pathways involved in a particular condition. Also, it reduces the complexity of vast biological networks by segmenting them into smaller, more interpretable parts, making data mining more manageable.

We identified submodules within the interactomes and tried to find those that are expressed exclusively in either the parent or PTHS patient samples, in both NPC and neuronal cells. Using MCODE, we found 3 and 6 submodules within the NPC and neuron interactomes, respectively (Supplementary Material 2). Among the NPC interactome submodules, none were found to be exclusively expressed either in PTHS patients or in the parents (Supplementary Material 8—Fig. S2), but we identified groups of genes (submodules 1, 2, 4 and 5) with exclusive expression in the parents within the neuron interactome (Fig. 3).

Among the submodules with higher expression in PTHS patients, submodule 1 contains genes coding for ribosomal genes (Fig. 2D), raising the question of whether the upregulation of ribosomal genes in PTHS neurons leads to aberrant protein translation during neurodevelopment in the patients. To validate the difference in ribosomal gene expression between parent and patient (PTHS), we searched for genes in the Ribosomal Protein Small (RPS) and Large (RPL) families in single-cell data from our previous publication[13]. Several members were found to be differentially expressed in our bulk RNA-Seq data from neurons, and the expression profiles of some genes in these families were significantly different along the neuronal differentiation pseudotime in single-cell RNA-Seq data from organoids (Supplementary Material 6—Fig. 3S).

Similar to submodule 1, submodule 2 of the neuron interactome is characterized by the prevalence of genes upregulated in PTHS. GO and KEGG enrichment analysis revealed that this submodule is enriched for processes and pathways related to type I interferon signaling pathway (FDR = 4.15e−21), cellular response to type I interferon (FDR = 4.15e−21), response to type I interferon (FDR = 4.17e−21) and defense response to virus (FDR = 2.63e−12). Moreover, the genes present within this submodule were also selected as hub genes in the neuron interactome (Supplementary Material 8—Fig. S3 and Table S2), further strengthening our hypothesis that they participate in PTHS pathogenesis. The existence of genes related to interferon could be indicative of neuroinflammatory responses [29] in the PTHS neural tissue, a possibility that could be further investigated in cellular models of neuroimmune interactions, including the understudied role of microglia in PTHS.

Submodule 4 displays a unique set of genes that are mostly downregulated in PTHS neurons, enriched for genes related to the glutamatergic synapse pathway, including *GRIN2D*, *GRIA2*, *DLG4*, *GRIN1*, *GRIN3A,* and *GRIN2B* (enrichment FDR = 2.170e−10). *GRIN* genes encode N-methyl-D-aspartate (NMDA) glutamate receptor subunits[30], characteristic of excitatory neurotransmission in the brain[31] and shown to be dysregulated in several neurodevelopmental disorders[32–34]. For example, *GRIN2B*, associated with various neurodevelopmental disorders (MIM #616139 and #613970), exhibits higher expression in the parent sample as compared to PTHS in the Glut lineage, according to single-cell transcriptomic data (Supplementary Material 6—Fig. 4S). The clear dysregulation of *GRIN* gene expression in the neuron interactome strongly supports the idea that glutamate receptor molecular pathology accompanies PTHS pathophysiology, and we hypothesize that the low activity of

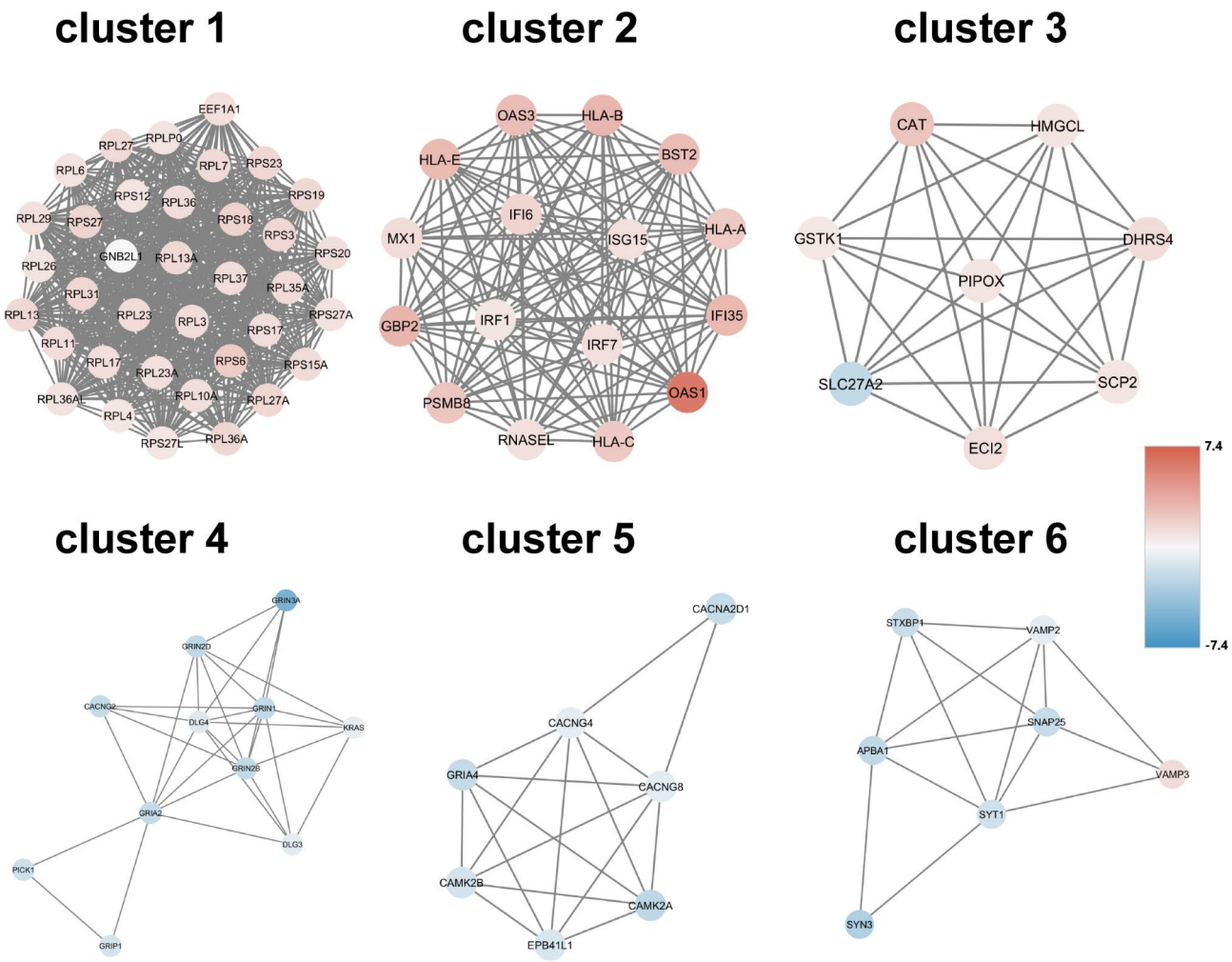


Fig. 3. Submodules within the neuron interactome. The colors represent the log2 fold-change (log2FC) of the genes between the group of PTHS individuals and the parents group (control). Genes upregulated in PTHS patients are shown in red, while downregulated genes are shown in blue.

NMDA receptors may partly explain the deficits in neuronal firing activity previously observed in PTHS brain organoids and patient-derived neurons in vitro[13].

Submodule 5 also contains genes downregulated in PTHS neuronal cells, enriched for pathways related to adrenergic signaling (FDR = 5.88–09), arrhythmogenic right ventricular cardiomyopathy (FDR = 1.38e−05), cardiac muscle contraction (1.38e−05), hypertrophic cardiomyopathy (FDR = 1.38e−05), and dilated cardiomyopathy (FDR = 1.38e−05). Genes present in submodule 5 show a strong association and are prevalent in a curated database with experimental biochemical data, based on the STRING scores obtained. Among the connected genes, glutamate receptor 4 (*GRIA4*) plays a critical role in excitatory synaptic transmission and interacts with genes coding for voltage-dependent calcium channels (*CACNG4* and *CACNG8*, which regulate the properties of synapses and trafficking of synaptic components) as well as with calcium/calmodulin-dependent kinase genes (*CAMK2A* and *CAMK2B*, responsible for dendritic spine and synapse formation, neuronal plasticity, and neurotransmitter release). Additionally, the interaction between *GRIA4* and the Erythrocyte Membrane Protein Band 4.1 gene (*EPB41*) is described as conferring stability and plasticity to the neuronal membrane (https://www.proteinatlas.org/ENSG00000088367-EPB41L1). Analysis of the *EPB41* gene in singl e-cell data revealed high expression in the GABA lineage of parent cells (Supplementary Material 6—Fig. 5S), indicating a deficiency in neuronal membrane plasticity in PTHS. Considering the interactions identified in this submodule, its high expression in control neurons and the gene functions, we hypothesize that the neurons from PTHS patients have impaired vesicle trafficking, synapse formation, and neuronal plasticity. This group of genes could become potential markers associated with the disease and help explain PTHS pathophysiology.

The interactome analysis performed here revealed novel connections between dysregulated genes, including biological processes and pathways that may be explored in the context of PTHS. Moreover, our findings suggest that PTHS is characterized by profound alterations in a range of cellular functions, including synaptic transmission, membrane ion permeability, and neuronal differentiation, with possible implications for how this and other similarly complex neurodevelopmental disorders are studied in the future.

Gene co-expression analysis

To identify clusters of co-expressed genes that have a common expression pattern in different cell types and in different individuals, revealing putative genetic associations in the context of PTHS, we applied the WGCNA method to RNA-Seq expression data from neurons and NPCs obtained from individuals affected by Pitt–Hopkins Syndrome and their parents. After filtering out lowly expressed genes and batch effects, we constructed a co-expression network using the weighted adjacency matrix approach with a soft-thresholding power of 20, followed by calculation of the topological overlap measure (TOM) between genes to identify modules of co-expressed genes. We identified a total of 73 modules of co-expressed genes, ranging in size from 30 to 2259 genes. To investigate the relationships between the identified modules and the type of cell (NPC or neuron), we calculated the module-sample relationships using the module eigengene (ME), a single representative value of the gene expression profile for each module.

We found that several co-expression modules were significantly associated with cell type (FDR-adjusted *p*-value < = 0.05 and |correlation|> 0.25) (Fig. 4). GO and KEGG enrichment analysis was performed on all key modules (Supplementary Material 3). In particular, the *red* module was positively associated with NPCs from PTHS patients and negatively associated with neurons from PTHS patients, while the *tan* module was positively associated with neurons from PTHS patients. The *floralwhite* module was positively associated with neurons from parents, and the *pink* module was positively associated with NPCs from parents and negatively associated with neurons from PTHS patients.

To explore the putative function of each identified key gene co-expression module, we performed GO and KEGG enrichment analysis using the clusterProfiler package. Among the biological processes enriched in the *red* module, we highlight the following categories: axonemal dynein complex assembly (GO:0,070,286; q-value = 4.56e−12), axoneme assembly (GO:0,035,082; q-value = 1.44e-33), cerebrospinal fluid circulation (GO:0,090,660; q-value = 2.94e−08), and GABAergic neuron differentiation (GO:0,097,154; q-value = 1.90e−06). These terms are connected to a variety of cellular functions, including cell motility, signaling, neuronal function, and differentiation. Of particular significance is the GABAergic neuron differentiation category, related to the process of neurogenesis of gamma-aminobutyric acid (GABA)-sensitive inhibitory neurons during brain development. As this *red* module is positively associated with NPCs from PTHS patients and negatively associated with neurons from PTHS patients, these data corroborate previous findings by our group showing that neurodevelopment in PTHS organoids involves defective formation of GABAergic neurons[13], a piece of result that points to the need of further pharmacological research to tackle this alteration in patients in the future. Regarding the interactome hub genes, the *red* module has the *PSMB8* and *OAS1* hub genes from the neuron interactome.

GO analysis of the *tan* module revealed enrichment for nucleosome assembly (GO:0,006,334, q-value = 3.49e−62), chromatin remodeling (GO:0,006,338, q-value = 2.78e−44), telomere organization (GO:0,032,200, q-value = 2.27e−23), and innate immune response in mucosa (GO:0,002,227, q-value = 0.0033). The positive association of the *tan* module with processes for maintaining genome stability during proper development and cellular differentiation is particularly noteworthy and suggests that abnormal neural development in PTHS patients may be linked to dysregulation of chromatin remodeling, in accordance with the histone deacetylase impairment previously reported in PTHS mouse models[35].

The *pink* module is simultaneously correlated with genotype and neural cell type, as it is positively associated with NPCs from parents and negatively associated with neurons from PTHS patients, suggesting a dysfunction in processes characteristic of the normal neuronal cell. Indeed, enrichment analysis revealed processes related to regulation of mTOR signaling (GO:0,032,006, q-value = 9.13e−06), axonal transport (GO:0,098,930, q-value = 2.64e−05), synaptic vesicle cytoskeletal transport (GO:0,099,514, q-value = 0.00029), synapse organization (GO:0,050,808, q-value = 0.049), cellular response to nutrient levels (GO:0,031,669, q-value = 0.044), cellular monovalent inorganic cation homeostasis (GO:0,030,004, q-value = 9.46e−04), regulation of neurotransmitter secretion (GO:0,046,928, q-value = 0.037), signal release from synapse (GO:0,099,643, q-value = 0.02), regulation of NMDA receptor activity (GO:2,000,310, q-value = 0.0062), and regulation of neurotransmitter secretion (GO:0,046,928, q-value = 0.0348). Most of these processes are evidently critical for neuronal function, particularly for synaptic communication and proper trafficking of cargo along neuronal processes, including the transport of neurotransmitter-filled vesicles to the synapse for release. Dysregulated genes associated with neuronal function have already been reported in the context of PTHS models[36], but the co-expression of these genes now reveals the extent of the concomitant dysregulation of several genes in this category, providing the framework for understanding PTHS at the systems level and for generating a more complete picture of the molecular pathology. Additionally, some enriched processes in the *pink* module are related to mitochondrial function and ATP production, such as ATP synthesis coupled electron transport (GO:0,042,773, q-value = 0.0050), respiratory electron transport chain (GO:0,022,904, q-value = 0.0127), and oxidative phosphorylation (hsa00190, q-value = 4.81e−05), which play key roles in neuronal physiology and energy metabolism. These findings stress the aberrant neuronal function characteristic of PTHS patients shown in previous publications[13,35].

Known genes related to PTHS and their association in co-expression analyses

One of the potential uses of our co-expression analysis is to verify which sets of genes are co-expressed (with a high correlation coefficient) with genes previously studied in the context of PTHS. For example, it has been shown[13] that PTHS neurons, in which *TCF4* expression is impaired, have diminished expression of the Contactin Associated Protein 2 gene (*CNTNAP2*), which is mutated in patients exhibiting another type of autism-spectrum neurodevelopmental disorder exhibiting symptoms overlapping with PTHS[37]. Moreover, our prior publication described that PTHS NPCs have impaired proliferation due to decreased activity in the Wnt signaling pathway[13].

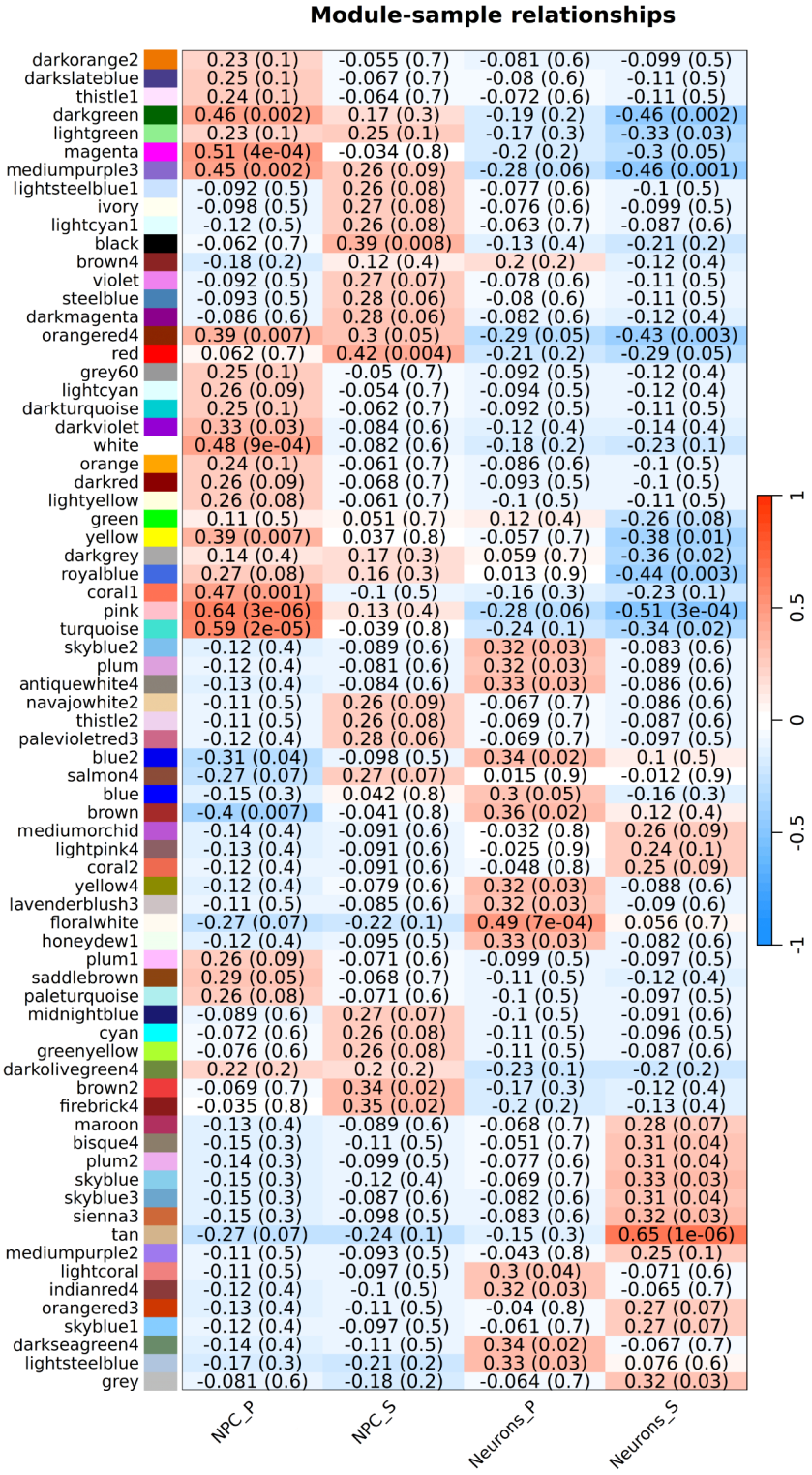


Fig. 4. Modulus-sample relationship after analysis of the WGCNA results, describing both cell type (NPC or neuron) and genotype (PTHS and parent) effects. The heatmap shows a correlation value and a *p*-value (in parenthesis) between the gene expression pattern and the eigengene of each module. High positive and high negative correlations are shown in red and blue, respectively. Legend: NPC_P: NPCs from parents; NPC_S: NPCs from PTHS patients; Neurons_P: neuronal cells from parents; Neurons_S: neuronal cells from PTHS patients.

However, it is not known how *TCF4* and *CNTNAP2* dysregulation are linked to the aberrant expression of Wnt signaling components, a piece of knowledge that may partly explain PTHS pathophysiology.

Indeed, among the genes co-expressed with *CNTNAP2* with the highest correlation coefficients, we found Adenomatosis Polyposis Coli Tumor Suppressor (*APC*), one of the key regulators of the Wnt signaling pathway, as well as *WNT10B*, *WNT4*, *WNT7A*, and *WNT8A* (Supplementary Material 5). We also found the cyclin kinase inhibitor genes *CDKN1B* and *CDKN2D* among the genes co-expressed with *CNTNAP2*, which may help link the aberrant replicative senescence observed in PTHS NPCs[13] with the dysregulation in *CNTNAP2*. Evidently, correlation does not imply causation, and further experiments are needed to define if *CNTNAP2* is upstream or downstream of these important regulators.

Interestingly, *CNTNAP2* is co-expressed with several calcium voltage-gated channel genes in the *CACN* family, potassium voltage-gated channels in the *KCN* family, and sodium channels in the *SCN* family (Annex 4). Given the severe alterations in intrinsic electrical properties and membrane excitability of PTHS neurons[13,35], it will be interesting to focus on the mechanistic study of those genes to verify if they are causally implicated with PTHS neuronal pathology or if they can be targeted pharmacologically to ameliorate the phenotypic abnormalities present in the patients' neural tissue.

Another gene that is co-expressed with *CNTNAP2* in PTHS is *SOX4* (Supplementary Material 5), which codes for an SRY-related HMG-box (SOX) protein family member that orchestrates neuronal differentiation[38] and was shown by our group to be dysregulated in PTHS NPCs and neurons, contributing to the lower differentiation rate of NPCs into cortical neurons in vitro[13]. The possibility that *SOX4* acts downstream of *TCF4* through *CNTNAP2* is suggested by the co-expression analyses performed here and merits further mechanistic investigation to determine the sequence of dysregulated events that lead from mutations in *TCF4* to aberrant neurogenesis in the developing cortex of PTHS individuals.

Finally, we verified that a set of genes co-expressed with *CNTNAP2* and *SOX4* is clearly involved with neuronal maturation. We highlight the co-expression of *BASP1* (Brain Abundant Membrane Attached Signal Protein 1) and *CRMP1* (Collapsin Response Mediator Protein 1), which are markers of the leading process of maturing neurons. Both *BASP1* and *CRMP1* exhibit higher expression in parent cells as compared to PTHS cells in the GABA lineage (Supplementary Material 6—Fig. 6S). These results strongly suggest that the PTHS brain may be characterized by deficits in cellular maturation during cortical development, an unexplored topic that needs further investigation.

In combination, the interactome and co-expression analyses performed here revealed novel molecular players and interactions that may not only be explored in future studies to decipher PTHS pathophysiology in different cell types and stages of brain development but also help select new targets for pharmacological research to treat this severe form of autism.

Integration with known disease-association gene databases

The study conducted by Sakai et al.[39] aimed to uncover shared molecular mechanisms underlying autism spectrum disorders (ASDs) by developing a protein interaction network based on ASD-associated genes. Through a yeast-two hybrid (Y2H) screen, the authors identified 539 candidate proteins that interact with syndromic ASD and associated proteins, forming 848 novel interactions, most of which were previously unreported (Supplementary Material 7). We analyzed the network described by Sakai et al. and identified Jaccard overlap indices of 0.01609195 and 0.03451178 with our NPC and neuron interactomes, respectively, indicating a limited overlap between the lists of dysregulated genes from both studies. This interesting result points to the existence of unique molecular aspects in PTHS, which require further investigation and highlight the importance of comparing across studies (although it should be noted that the results of each gene network study are highly influenced by the specific details of the network construction process).

The study by Konopka et al.[40] utilized weighted gene co-expression network analysis (WGCNA) to investigate transcriptional programs involved in human neural differentiation and their relationship with neuropsychiatric disorders, such as autism. Neurons derived from normal human neural progenitors (NHNPs) were differentiated into post-mitotic states and analyzed over time using genomic microarrays. WGCNA revealed that several autism spectrum disorder-associated genes are highly co-expressed during differentiation, and the authors highlighted four co-expression modules: *black* (415 genes), *pink* (1,048 genes), royal blue (947 genes), and *turquoise* (253 genes) (Supplementary Material 7). The highest and lowest Jaccard indices with our NPC network were found in the *black* (0.03676) and *turquoise* (0.01071) modules, respectively. Meanwhile, the highest and lowest Jaccard indices with our neurons network were observed in the *royal blue* (0.03293) and *turquoise* (0.02015) modules, respectively. The authors noted the presence of genes related to autism, epilepsy, or both in the network, as identified through the SFARI gene database (http://gene.sfari.org). Some of these highlighted genes across the four modules were found in our interactomes (e.g. *SERPINE1*, *SYN1*, *NRXN1*, *SHANK2* and *CACNA2D2*), except for genes associated with both autism and epilepsy.

Willsey et al.[41] investigated the convergence of nine high-confidence autism spectrum disorder-associated genes (hcASD) across different developmental periods, brain regions, and cell types. They used co-expression networks based on these genes as "seeds" and expression data spanning various human brain regions during development[42]. The study observed an enrichment of likely ASD-associated genes (pASD) and discussed the top 20 genes related to them. The authors described findings in three networks generated from comparisons of three developmental brain periods (Supplementary Material 7): (1) early fetal to early mid-fetal, (2) early mid-fetal to late mid-fetal, and (3) neonatal & early infancy to early childhood. The Jaccard indices for all three networks with our NPC interactome were low (range 0.002–0.0079; e.g., *KCNC1* and *ZBTB16*), but higher with our neuron interactome (range 0.02205–0.02469; e.g., *CACNA1B*, *CACNG4*, *GRIN2B*, *SYN2* and *GRIA4*). The co-expression networks demonstrated enrichment during critical periods of fetal development, aligned with the transition from progenitors to mature neurons. This probably reflects the greater functional complexity of neurons and

their interactive genetic profiles in processes such as synaptogenesis. Consequently, networks focused on these periods may capture genes more relevant to neuronal processes compared to progenitor stages, ensuring greater intersection with our mature neuron interactome.

The genes from the SFARI database (https://gene.sfari.org/database/human-gene/), a comprehensive resource cataloging 1203 autism-associated genes (Supplementary Material 7), showed low intersection with our interactomes, reflected by Jaccard indices of 0.02624 for NPCs (39 common genes) and 0.0735 for mature neurons (128 common genes; e.g. *SNAP25*, *VAMP2*, *GRIN2B*, *SYT1*, *CNTNAP2*, *DLG4* and *GRIA2*). Functional enrichment of the gene lists from the intersections shows that, for the NPC interactome, metabolic pathways related to cardiopathies were identified, such as Arrhythmogenic right ventricular cardiomyopathy (hsa05412; p-adj = 1.2e−3), Hypertrophic cardiomyopathy (hsa05410; p-adj = 1.3e−2), and Dilated cardiomyopathy (hsa05414; p-adj = 1.3e−2). Functional enrichment for the intersection of genes present in the neuron interactome revealed pathways related to synapses, such as Glutamatergic synapse (hsa04724; p-adj = 2.6e−7), Synaptic vesicle cycle (hsa04721; p-adj = 6.1e−6), Dopaminergic synapse (hsa04728; p-adj = 2.8e−5), and Serotonergic synapse (hsa04726; p-adj = 2.1e−3).

The small intersection between the SFARI database and our dataset of dysregulated genes is interesting and unexpected, revealing that PTHS is characterized by unique molecular alterations not shared by other autism disorders, a piece of result that may direct future efforts to identify therapeutic targets for this and other related neurodevelopmental diseases. On the other hand, the higher Jaccard index between SFARI and our neuron interactome in comparison with the lower Jaccard index between SFARI and our NPC interactome may be attributed to the less specialized nature of neural progenitors, underscoring the importance of advancing neural cells to later stages of differentiation when modelling neuropsychiatric disorders in vitro.

Molecular association of PTHS with brainstem-related neurodevelopmental disorders

Some of the most debilitating clinical symptoms found in PTHS children are respiratory anomalies, characterized by episodes of hyperventilation, usually followed by apnea[1]. The investigation of neurodevelopmental disorders associated with the brainstem is therefore essential to understanding the shared molecular pathways and biological mechanisms underlying these conditions, particularly those that may drive the dysregulation of central nervous system control of respiratory rhythm. In this section, we explore and compare the results obtained in our analysis with findings from other disorders, such as Congenital Central Hypoventilation Syndrome (CCHS), Pitt–Hopkins-like Syndrome 2 (PTHSL2), Developmental and Epileptic Encephalopathy 2 (DEE2), and Rett Syndrome (RTT). This approach allows for identifying connections between relevant genes and modules, enhancing the understanding of the molecular mechanisms underlying PTHS and its relationship with other neurodevelopmental disorders.

Congenital Central Hypoventilation Syndrome (CCHS) (MIM 209,880) is characterized by respiratory difficulties due to the inability to detect low oxygen or high carbon dioxide levels, caused by mutations in the *PHOX2B* gene. *PHOX2B* was not present in the interactomes generated in our study but was identified in the *white* module (439 nodes and 63,242 edges) of co-expression and is considered a hub gene of that module (ME = 0.9525) (Supplementary Material 4), with 326 high confidence interactions (Supplementary Material 5). Furthermore, the *white* module is considered a key module and is positively associated with NPCs from parents (r = 0.48; *p*-value = 9e−04).

Pitt–Hopkins-like Syndrome 2 (PTHSL2) (MIM 614,325) is a rare developmental condition associated with autism spectrum disorder, caused by alterations in the *NRXN1* gene. This gene was found in our neuronal interactome, with neighbors including *NXPH1*, *DLG4*, *SHANK2*, *CBLN1*, *SYT1*, and *LRRTM1*. Co-expression analysis revealed that *NRXN1* is in the *turquoise* module (4073 nodes and 5,025,402 edges), being a hub gene (ME = 0.9563) (Supplementary Material 4) with 3796 high confidence interactions in that module (Supplementary Material 5). The *turquoise* module is a key module, positively associated with NPCs from parents (r = 0.59; p = 2e−5) and negatively associated with neurons from PTHS patients (r = −0.34; *p* = 0.02). Interestingly, two articles link failures in neural development to *NRXN1* deletions. Lam et al.[43] explore an autism spectrum disorder disease model using induced pluripotent stem cells (iPSCs) derived from an individual with biallelic deletion of NRXN1-alpha. Using the BackSPIN algorithm, the study identified gene-enriched lists for different cell types during neurogenesis and neuronal differentiation. Comparing the list of 486 genes (Supplementary Material 7) with our neuron interactome, which is associated with mature neurons, resulted in a Jaccard index of 0.0745, with 80 intersecting genes (e.g., *CAMK2B*, *NRXN1*, *GNG3*, *THSD7A*, *CNTN1*, *SHANK2* and *GRIA2*). Functional enrichment of the intersection showed that the most enriched biological processes included chemical synaptic transmission (GO:0,007,268; p-adj = 1.7e−10), nervous system development (GO:0,007,399; p-adj = 2.9e−05), and calcium ion transport (GO:0,006,816; p-adj = 2.3e−05). Additionally, Di et al.[44] investigated the role of *NRXN1* in neurodevelopmental disorders (NDDs), focusing on behavioral and molecular changes in the prefrontal cortex (PFC). Using tandem mass tag (TMT)-based proteomic analysis, the authors identified 130 differentially expressed proteins in PFC neurons of *NRXN1* knockout mice. Intersections of this list with our NPC interactome yielded two genes (*PFKP* and *TFAP2C*) and 10 genes with our neuron interactome (*NRXN1*, *ANXA1*, *SSR3*, *SPOCK3*, *CHGB*, *SPTB*, *HADHB*, *GPC2*, *GFAP*, and *JUP*). Analyzing the 3796 genes connected to *NRXN1* in the *turquoise* module revealed intersections of 110 and 19 genes with the Lam and Di studies, respectively.

Developmental and Epileptic Encephalopathy 2 (DEE2) (MIM 300,672) is a severe form of epilepsy with developmental delay and early-life seizures, associated with mutations in the *CDKL5* gene. This gene was not present in our interactomes but was found in the *turquoise* module of co-expression, connected to 2657 genes (Supplementary Material 5). A study by Liao and Lee[45] investigated the molecular mechanisms underlying neuronal hyperexcitability in *Cdkl5* mutant mice using transcriptomic analysis across different ages (P7, P12, and P17), covering the transient manifestation period of epileptic activities. From all statistically significant

differentially expressed genes provided by the authors (Supplementary Material 7), we identified nine intersecting genes with *CDKL5*'s first neighbors in the *turquoise* module (*GRM2*, *KCNQ3*, *GLRA2*, *SLC8A3*, *ADCY5*, *DRD2*, *GRM4*, *GAD1*, and *ST18*), all of which are important regulators of synaptic transmission.

Lastly, Rett Syndrome (RTT) (MIM 312,750) is a neuropsychiatric disorder primarily affecting girls, leading to loss of motor and language skills, caused by mutations in the *MECP2* gene. This gene was found in the *yellow* module of co-expression (1697 nodes and 972,732 edges) and is a hub gene of the module (ME = 0.9348). Additionally, the *yellow* module is a key module, positively associated with NPCs from parents (r = 0.39; $p = 0.007$) and negatively associated with neurons from PTHS patients (r = − 0.38; $p = 0.01$). The *MECP2* gene has 1,629 interactions in the module. A study by Haase et al.[46] used transcriptomic analyses of iPSCs derived from RTT patients and healthy controls, employing weighted gene correlation network analysis (WGCNA). The study compared gene expression profiles across dysregulated modules preserved in iPSCs and neurons derived from RTT patient iPSCs. Of the 1726 genes analyzed, 27 were downregulated and 13 were upregulated in the derived neurons. Among the 40 genes in the study (Supplementary Material 7), only *BAAT* was a first neighbor of *MECP2* in our network analyses.

Characterization of OMIM disease-associated genes in the network

To explore the relationship between disease-associated genes and our interactomes, we searched the OMIM database (December 23, 2024 version), using a selection of terms to cover a range of relevant diseases, including "neurodevelopmental disorder", "brain development OR brain disorder", "immune dysregulation", "endocrine", "hypoventilation" and "hyperventilation". The intersection between the lists of genes found in each category with those present in our interactomes was evaluated (Supplementary Material 8—Table S3) to determine the Jaccard index for each network and the number of shared genes. The results from the OMIM search are provided in Supplementary Material 7.

For the term "neurodevelopmental disorder", we found 581 genes in OMIM, 10 of which are present in our NPC interactome and 53 of which are found in our neuron interactome. We observed a greater overlap of genes in this category with the neuronal interactome compared to the NPC interactome (Jaccard indices: for neurons = 0.044; for NPCs = 0.011). Terms related to brain development ("brain development OR brain disorder") yielded a significant intersection, with 10 genes in the NPC interactome and 29 in the neuronal interactome, although the Jaccard indices were still relatively low (0.015 for NPCs and 0.029 for neurons). For terms under "immune dysregulation", we observed an even smaller intersection, with only 3 genes in both interactomes, resulting in very low Jaccard indices (0.007 for NPCs and 0.004 for neurons), indicating that genes involved in immune dysregulation are underrepresented in our interactomes. For the term "endocrine" (509 genes), an intersection of 14 genes was found with our NPC interactome and 26 with our neuronal interactome, presenting slightly higher Jaccard indices (0.017 for NPCs and 0.022 for neurons). Interestingly, for the terms "hypoventilation" and "hyperventilation," the intersection was much smaller, with a reduced number of genes in our interactomes, especially for "hypoventilation," where no genes were shared with the NPC interactome.

In principle, one would expect high overlap between the genes and networks involved in Pitt–Hopkins Syndrome and other neurodevelopmental disorders, but our results appear to show otherwise. These findings may pave the way towards deciphering the uniqueness of the molecular dysregulation present in PTHS, allowing the scientific community to raise further mechanistic hypothesis and propel directed therapeutic research to meet the medical needs faced by patients with this devastating disorder.

Conclusions

Our study revealed extensive dysregulation in the expression of several genes and pathways in human cellular samples from PTHS individuals, in accordance with the notion that this is a complex neurodevelopmental disorder involving alterations in a multitude of genes related to brain development and neuronal function, including some involved in such distinct processes as neuronal excitability, synapse formation, neurotransmission, and neuronal maturation.

The interactome reconstruction steps allowed us to identify gene interactions, key pathways and biological processes associated with PTHS pathogenesis, such as those related to neuronal function, cell–cell communication, signaling, and neuronal maturation. Our exploratory analysis of the interactome discovered genetic or biochemical interactions that could raise mechanistic hypotheses regarding PTHS pathophysiology, an outcome that simple differential gene expression analysis could not reveal. By focusing on particular interactome submodules, we also identified novel genes that were downregulated in PTHS patients, providing further insights into possible molecular mechanisms underlying the disease, including genes that encode cyclin inhibitors and transcription factors expressed during progenitor differentiation that may explain why PTHS progenitor cells exhibit precocious cellular replicative senescence[13].

The co-expression analysis performed on NPC and neurons from both parents and PTHS patients allowed us to identify several modules of genes that are significantly dysregulated in the disease, suggesting disrupted pathways and biological processes. Moreover, we identified several key modules and hub genes in these modules, such as those related to histone family genes, which could serve as potential positive prognostic factors for PTHS patients. The differences observed in the hub genes between the NPC cells and neurons of parents and PTHS patients suggest potential differences in disease progression and severity, highlighting the importance of studying different cell types to gain a better understanding of the disease.

We also investigated the intersection between the interactomes analyzed in our study and the lists of dysregulated genes in other studies and in disease databases related to autism and neuropsychiatric disorders. Although the intersection between genes associated with these diseases and our NPC and neuron interactomes was shown to be relatively low, we identified some relevant genes that may contribute to our understanding of conditions caused by mutations in the Transcription Factor 4 gene and related disorders. Our findings revealed

important differences in PTHS, highlighting the novelty of our work in identifying aspects related to this disease that are not present in other forms of autism or disorders of neurodevelopment. This analysis provides new insights into how gene interactions may contribute to the underlying processes of these diseases, enabling the prioritization of targets for functional studies and potential clinical interventions. It is also important to mention that *TCF4* is regarded as a risk gene for schizophrenia[47] and the results described here may thus be relevant for understanding these other types of neuropsychiatric conditions.

In conclusion, our integrative approach combining co-expression analysis, hub gene identification, single-cell analysis and interactome reconstruction allowed us to obtain a comprehensive understanding of PTHS pathogenesis and identify potential therapeutic targets for future study.

Data availability

Data supporting the findings in this study are included within the Supplementary Material. Large tables containing the lists of differentially expressed genes in each parent-patient pair and the differentially expressed genes common to all pairs in bulk RNA sequencing and single-cell RNA sequencing experiments are publicly available without restriction from the Zenodo repository (https://doi.org/10.5281/zenodo.6325406). Raw RNA sequenci ng and single cell RNA sequencing data analyzed during the current study are available in the Gene Expression Omnibus (GEO) repository of the National Center for Biotechnology Information (NCBI), under persistent accession numbers GSE159392, GSE159860, and GSE189121.

Received: 15 October 2024; Accepted: 4 February 2025
Published online: 10 February 2025

References

1. Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R. & Papes, F. Transcription factor 4 and its association with psychiatric disorders. *Transl. Psychiatry* https://doi.org/10.1038/s41398-020-01138-0 (2021).
2. van Dam, S., Võsa, U., van der Graaf, A., Franke, L. & de Magalhães, J. P. Gene co-expression analysis for functional classification and gene–disease predictions. *Brief Bioinform.* **19**, 575–592 (2018).
3. Hu, R. T. et al. Co-expression Network analysis reveals novel genes underlying Alzheimer's disease pathogenesis. *Front. Aging Neurosci.* https://doi.org/10.3389/fnagi.2020.605961 (2020).
4. Liang, J. W. et al. Application of weighted gene co-expression network analysis to explore the key genes in Alzheimer's disease. *J. Alzheimers Dis.* **65**, 1353–1364 (2018).
5. Santri, I. N. et al. Identification of hub genes and potential biomarkers for childhood asthma by utilizing an established bioinformatic analysis approach. *Biomedicines* **10**, 2311 (2022).
6. Hasankhani, A. et al. Differential co-expression network analysis reveals key hub-high traffic genes as potential therapeutic targets for COVID-19 pandemic. *Front. Immunol.* **12**, 5371 (2021).
7. Chen, H., Yang, J. & Wu, W. Seven key hub genes identified by gene co-expression network in cutaneous squamous cell carcinoma. *BMC Cancer* **21**, 1–12 (2021).
8. Li, X. et al. Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. *Sci. Rep.* **7**, 1–13 (2017).
9. Zheng, P. F., Chen, L. Z., Guan, Y. Z. & Liu, P. Weighted gene co-expression network analysis identifies specific modules and hub genes related to coronary artery disease. *Sci. Rep.* **2021**(11), 1–13 (2021).
10. Jin, F. & Duan, C. Identification of immune-infiltrated hub genes as potential biomarkers of Moyamoya disease by bioinformatics analysis. *Orphanet. J. Rare Dis.* **17**, 1–12 (2022).
11. Zuo, D. et al. Identification of hub genes and their novel diagnostic and prognostic significance in pancreatic adenocarcinoma. *Cancer Biol. Med.* **19**, 1029–1046 (2022).
12. Li, X. et al. Integrated analysis of brain transcriptome reveals convergent molecular pathways in autism spectrum disorder. *Front. Psychiatry* **10**, 706 (2019).
13. Papes, F. et al. Transcription Factor 4 loss-of-function is associated with deficits in progenitor proliferation and cortical neuron content. *Nat. Commun.* **2022**(13), 1–26 (2022).
14. Szklarczyk, D. et al. STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
15. Bader, G. D. & Hogue, C. W. V. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinform.* **4**, 2 (2003).
16. Shannon, P. et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498 (2003).
17. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 1–13 (2008).
18. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* https://doi.org/10.2202/1544-6115.1128 (2005).
19. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innov.* **2**, 100141 (2021).
20. Kanehisa, M., Furumichi, M., Sato, Y., Matsuura, Y. & Ishiguro-Watanabe, M. KEGG: Biological systems database as a model of the real world. *Nucleic Acids Res.* **53**, D672–D677 (2025).
21. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
22. Risher, W. C. & Eroglu, C. Thrombospondins as key regulators of synaptogenesis in the central nervous system. *Matrix Biol.* **31**, 170 (2012).
23. Wang, B., Guo, W. T. & Huang, Y. Thrombospondins and synaptogenesis. *Neural Regen Res.* **7**, 1737 (2012).
24. Gillespie, M. et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
25. Moremen, K. W., Tiemeyer, M. & Nairn, A. V. Vertebrate protein glycosylation: Diversity, synthesis and function. *Nat. Rev. Mol. Cell Biol.* **13**, 448–462 (2012).
26. Scibetta, A. G., Wong, P. P., Chan, K. Y. V., Canosa, M. & Hurst, H. C. Dual association by TFAP2A during activation of the p21cip/CDKN1A promoter. *Cell Cycle* **9**, 4525–4532 (2010).
27. Wong, P.-P. et al. Histone demethylase KDM5B collaborates with TFAP2C and Myc to repress the cell cycle inhibitor p21(cip) (CDKN1A). *Mol. Cell Biol.* **32**, 1633–1644 (2012).
28. Quan, W. et al. Identification of Potential core genes in Parkinson's disease using bioinformatics analysis. *Parkinsons Dis.* **2021**, 1–10 (2021).
29. Roy, E. R. et al. Type I interferon response drives neuroinflammation and synapse loss in Alzheimer disease. *J. Clin. Invest.* **130**, 1912–1930 (2020).

30. Vyklicky, V. et al. Structure, function, and pharmacology of NMDA receptor channels. *Physiol. Res.* **63**, S191–S203 (2014).
31. Pizarro, F. J., Jewett, M. C., Nielsen, J. & Agosin, E. Growth temperature exerts differential physiological and transcriptional responses in laboratory and wine strains of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **74**, 6358–6368 (2008).
32. Platzer, K. & Lemke, J. R. GRIN2B-Related Neurodevelopmental Disorder. *GeneReviews®* (2021).
33. Platzer, K. & Lemke, J. R. GRIN1-Related Neurodevelopmental Disorder. *GeneReviews®* (2021).
34. Platzer, K., Krey, I. & Lemke, J. R. GRIN2D-Related Developmental and Epileptic Encephalopathy. *GeneReviews®* (2022).
35. Kennedy, A. J. et al. Tcf4 regulates synaptic plasticity, DNA methylation, and memory function. *Cell Rep.* **16**, 2666–2685 (2016).
36. Phan, B. D. N. et al. A myelin-related transcriptomic profile is shared by Pitt–Hopkins syndrome models and human autism spectrum disorder. *Nat. Neurosci.* **23**, 375–385 (2020).
37. Zweier, C. et al. CNTNAP2 and NRXN1 are mutated in autosomal-recessive Pitt–Hopkins-like mental retardation and determine the level of a common synaptic protein in drosophila. *Am. J. Hum. Genet.* **85**, 655 (2009).
38. Chen, C., Lee, G. A., Pourmorady, A., Sock, E. & Donoghue, M. J. Orchestration of neuronal differentiation and progenitor pool expansion in the developing cortex by SoxC genes. *J. Neurosci.* **35**, 10629–10642 (2015).
39. Sakai, Y. et al. Protein interactome reveals converging molecular pathways among autism disorders. *Sci. Transl. Med.* **3**, 86ra49 (2011).
40. Konopka, G. et al. Modeling the functional genomics of autism using human neurons. *Mol. Psychiatry* **17**, 202 (2011).
41. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997 (2013).
42. Kang, H. J. et al. Spatiotemporal transcriptome of the human brain. *Nature* **478**, 483 (2011).
43. Lam, M. et al. Single cell analysis of autism patient with bi-allelic NRXN1-alpha deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality. *Exp. Cell Res.* **383**, 111469 (2019).
44. Wu, D. et al. NRXN1 depletion in the medial prefrontal cortex induces anxiety-like behaviors and abnormal social phenotypes along with impaired neurite outgrowth in rat. *J. Neurodev. Disord.* **15**, 1–14 (2023).
45. Liao, W. & Lee, K. Z. CDKL5-mediated developmental tuning of neuronal excitability and concomitant regulation of transcriptome. *Hum. Mol. Genet.* **32**, 3276–3298 (2023).
46. Haase, F., Gloss, B. S., Tam, P. P. L. & Gold, W. A. Wgcna identifies translational and proteasome-ubiquitin dysfunction in rett syndrome. *Int. J. Mol. Sci.* **22**, 9954 (2021).
47. Hill, M. J. et al. Knockdown of the schizophrenia susceptibility gene TCF4 alters gene expression and proliferation of progenitor cells from the developing human neocortex. *J. Psychiatry Neurosci.* **42**, 181–188 (2017).

Acknowledgements

The authors thank Drs. Gonçalo Pereira and Marcelo Carazzolle for resources, Ana Paula Ferreira, Welbe Bragança, and the CQMED staff for administrative and technical help. This work was supported by grants #2020/11451-7 (F.P.), #2022/12762-1 (F.P.) and #2019/12914-3 (L.M.C.) and by fellowship #2018/04240-0 (A.P.C.) from the Sao Paulo Research Foundation (FAPESP).

Author contributions

L.M.C: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing- Original draft preparation, Writing-Review and editing, Visualization; V.M.A.C: Investigation, Methodology, Data curation, Writing-Review and editing; A.P.C: Conceptualization, Writing-Original draft preparation, Data curation, Writing-Review and editing; F.P: Conceptualization, Investigation, Methodology, Data curation, Writing-Original draft preparation, Resources, Writing-Review and editing.

Declarations

Competing interests

The University of California San Diego has filed a patent application (WO2022072709A1, published on 04-07-2022), in which FP is an inventor, regarding methods to provide correction of TCF4 expression in PTHS that were based on results described in another publication[13]; this other publication contains publicly available gene expression datasets that have been used to perform the analysis in the present paper. Other authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-89334-0.

Correspondence and requests for materials should be addressed to L.M.C., A.P.C. or F.P.

Reprints and permissions information is available at www.nature.com/reprints.

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

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/.

© The Author(s) 2025