Coupling of autism genes to tissue-wide expression and dysfunction of synapse, calcium signalling and transcriptional regulation

Jamie Reilly1*, Louise Gallagher2,3, Geraldine Leader4, Sanbing Shen1,5*

1 Regenerative Medicine Institute, School of Medicine, Biomedical Science Building, National University of Ireland (NUI) Galway, Galway, Ireland, 2 Discipline of Psychiatry, School of Medicine, Trinity College Dublin, Dublin, Ireland, 3 Trinity Translational Medicine Institute, Trinity Centre for Health Sciences—Trinity College Dublin, St. James’s Hospital, Dublin, Ireland, 4 Irish Centre for Autism and Neurodevelopmental Research (ICAN), Department of Psychology, National University of Ireland (NUI) Galway, Galway, Ireland, 5 FutureNeuro Research Centre, Royal College of Surgeons in Ireland (RCSI), Dublin, Ireland

* J.Reilly12@nuigalway.ie (JR); Sanbing.shen@nuigalway.ie (SS)

Abstract

Autism Spectrum Disorder (ASD) is a heterogeneous disorder that is often accompanied with many co-morbidities. Recent genetic studies have identified various pathways from hundreds of candidate risk genes with varying levels of association to ASD. However, it is unknown which pathways are specific to the core symptoms or which are shared by the co-morbidities. We hypothesised that critical ASD candidates should appear widely across different scoring systems, and that comorbidity pathways should be constituted by genes expressed in the relevant tissues. We analysed the Simons Foundation for Autism Research Initiative (SFARI) database and four independently published scoring systems and identified 292 overlapping genes. We examined their mRNA expression using the Genotype-Tissue Expression (GTEx) database and validated protein expression levels using the human protein atlas (HPA) dataset. This led to clustering of the overlapping ASD genes into 2 groups; one with 91 genes primarily expressed in the central nervous system (CNS geneset) and another with 201 genes expressed in both CNS and peripheral tissues (CNS+PT geneset). Bioinformatic analyses showed a high enrichment of CNS development and synaptic transmission in the CNS geneset, and an enrichment of synapse, chromatin remodelling, gene regulation and endocrine signalling in the CNS+PT geneset. Calcium signalling and the glutamatergic synapse were found to be highly interconnected among pathways in the combined geneset. Our analyses demonstrate that 2/3 of ASD genes are expressed beyond the brain, which may impact peripheral function and involve in ASD co-morbidities, and relevant pathways may be explored for the treatment of ASD co-morbidities.
can be accessed via Biocductor in R, using the code provided in the supplementary information. The networks from Huri can be accessed from NDEX (http://www.ndexbio.org/#/user/69e7b1d-8981-11ea-aaef-0ac135e8bacf). Information from the scoring systems used can be found in the supplementary information sections of their respective publications. Differentially Expressed genes can be found in supporting information from respective publications and in Geo2r (GSE42133 for Pramparo, GSE28521 for Voineagu). Code is provided for data access to HPA and GeneOverlap and generating figures. Genelists were obtained from pysgenet(http://www.pysgenet.org/web/PsyGenET/menu), Harmonizome(https://maayanlab.cloud/Handromize/dataset/GWASdb +SNP-Disease+Associations) and the supplementary information from respective publications.

**Funding:** Funding was provided by Science Foundation Ireland (Grant 13/1A/1787 to S.S. and L.G., 16/RC/3948 to FutureNeuro), and National University of Ireland Galway (Grant RSU002 to S. S.) for funding the research.

**Competing interests:** The authors have declared that no competing interests exist.

Our analyses suggest that a third of ASD risk genes (CNS geneset) is specifically expressed in the CNS, which are involved in brain development, synaptic function and ion transport, whereas the majority of ASD factors are highly expressed in both CNS and peripheral tissues (CNS+PT geneset), with pathways of brain development, chromatin organisation and gene regulation, which may account for ASD peripheral comorbidity.
Methods

Datasets and shortlisting of the ASD risk factors

Five datasets were chosen as the starting point of this study [28, 31–34], and high-ranking genes were shortlisted from each dataset with defined criteria (Table 1, Fig 1). For the Exome Aggregation Consortium (EXAC) with exome sequencing data from ~60,000 cases (S1 Table), a high intolerance to mutation of pLi > 0.9 was applied, where pLi indicated the level of intolerance to mutations in a given gene, with many containing loss of function variants. In the Krishnan’s geneset (S2 Table) created using human disease databases of GAD, OMIM, HUGE and SFARI in December 2013, along with a brain-specific functional interaction network, a q-value of < 0.05 was used, where q-value was the probability of the gene being an ASD risk factor.

Table 1. Overview of five datasets and the independent scoring systems used to shortlist ASD genes for overlapping analysis.

<table>
<thead>
<tr>
<th>Score name</th>
<th>Data Source</th>
<th>Starting genes</th>
<th>Score Threshold</th>
<th>Shortlisted genes</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAC</td>
<td>Exome sequences from 60,706 individuals</td>
<td>15735</td>
<td>pLi &gt; 0.9 with &gt;90% negative effect upon mutation</td>
<td>3126</td>
<td>Lek et al. [31]</td>
</tr>
<tr>
<td>SFARI</td>
<td>SFARI GENE (Jan 2019)</td>
<td>1053</td>
<td>All genes recorded in Jan 2019</td>
<td>1053</td>
<td>Gene.SFARI.org (Jan 2019)</td>
</tr>
<tr>
<td>Krishnan</td>
<td>SFARI, OMIM, GAD, HUGE (up to 2013)</td>
<td>25825</td>
<td>q value &lt; 0.05</td>
<td>3225</td>
<td>Krishnan et al. [32]</td>
</tr>
<tr>
<td>Duda</td>
<td>Microarray data from human, mouse and rat. Protein interaction databases (MIPS, BIOGRID, MINT, IntAct)</td>
<td>21115</td>
<td>Top 10th percentile</td>
<td>2111</td>
<td>Duda et al. [34]</td>
</tr>
<tr>
<td>Zhang</td>
<td>Mouse CNS Microarray Data, Genes homologous to human</td>
<td>15950</td>
<td>Positive DAMAGE score (D&gt;0)</td>
<td>7189</td>
<td>Zhang et al. 2017</td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0242773.t001

Fig 1. Flowchart of the analysis for the paper: Arrow indicates the steps of each analysis.

https://doi.org/10.1371/journal.pone.0242773.g001
candidate after multiple testing for false positivity. For the Zhang's geneset (S3 Table) made using CNS microarray expression data from six brain regions derived from mice and validated using data from exome sequencing studies [11, 14, 17, 29, 36, 37], mutations of the human homologues with a (DAMAGE) score of $D \geq 0$ were shortlisted, where positive D-score was a measure of a mutation’s likeliness to be associated with ASD. For the Duda’s list (S4 Table) which was created from de novo mutation analysis, protein-protein interaction and phenotype information, the top 10% of genes in the list were selected, as this was used as a cut off in the original publication. Finally, for the SFARI collection, all genes up to January 2019 were included (S5 Table).

**Overlap of shortlisted genes with SFARI**

After shortlisting of high-ranking genes from each dataset, the Jvenn program was applied to identify the common ASD risk genes among the 5 sources [38]. A list of 519 genes, which were overlapped among 4 of 5 sources, was extracted (Table 1, Fig 1). Among them, 319 genes appeared in the SFARI database were taken forward for expression and enrichment analyses.

**Filtering using gene expression analysis data**

Since ASD has a wide array of peripheral co-morbidities, we believe that some ASD genes are expressed in peripheral tissues. To explore this possibility, mRNA expression data were downloaded from the GTEx consortium (v7) in the form of median TPM (Transcripts Per Million) values from all tissues [39]. We excluded low expression genes based on the average of the median (TPM < 3) in both CNS and PT groups. Additionally, we removed genes with a SFARI score of 6, which were not likely to be associated with ASD. The expression data was applied to the 319 selected genes and uploaded to Morpheus (https://software.broadinstitute.org/morpheus) to generate heatmap (with settings for clustering: hierarchical, euclidean distance, linkage method complete, clustering based on columns). Genes with extremely low levels of mRNA expression, at an average TPM $\leq 3$ in both the brain and peripheral tissues, were excluded from the subsequent analyses (S8 Table).

**Overview of protein expression levels**

As an additional level of verification, we used HPA data (v19.3) obtained via HPA analyze [40], a Bioconductor program that runs in R, to assess protein expression levels across multiple tissues among the two genesets, and used ggplot2 [41] to visualise the data.

**Tissue specific interaction networks**

The expression of ASD genes in other tissues indicates that they may interact with other factors in tissue-specific networks. To explore this, we used tissue-specific networks generated from Huri to see if ASD candidate genes were present in other networks, and if they had any interacting partners within these networks.

**Functional enrichment analysis of the final geneset**

The final list of ASD common risk factors was analyzed through STRING program for pathway analyses, except for SHANK3 that was misidentified as HOMER2 in the current human database of STRING v11. The resulting GO Terms (Biological Processes, Cellular Components, and Molecular Function) and KEGG pathways were downloaded. The same list was also loaded into Cytoscape [42] to identify sub-clusters of genes in interaction network.
Analysis of genesets for co-morbid phenotypes

To examine potential association of the ASD genesets with occurring co-morbidities, an over-representation analysis (ORA) was carried out using the tool WebGestalt [43] to assess other co-morbid conditions linking to the ASD genesets. The Human Phenotype Ontology database was used for the analysis [44]. The top 50 terms were used as a cut off to balance between the co-morbidities reported in ASD and to ensure that the final lists are not too broad and overly diluted.

Comparison of shortlisted ASD geneset with ASD expression studies

To examine the utility of the ASD geneset in the literature of ASD gene expression studies, we compared the ASD geneset with DEGs reported from post-mortem brain [45], blood [46–48] and GI tissue [46, 49], as well as iPSC-derived cell models [50–55], to see if any of the ASD genes were significantly up or downregulated. The DEGs were obtained using autism versus control group with FDR (adj p-value) at 0.05, except for Voineagu [56] and Pramparo [47], which we used Geo2R [57] using autism versus controls as groups to obtain DEGS at 0.05 FDR (adj p-value).

Expression of genes across brain development and sex-bias

We used CSEA [58] tool to check for enrichment of our genes for human gene expression across developmental periods and brain tissues in the Brainspan dataset. We also used the genesets from the publications [59, 60] to see if any of our shortlisted genes had sex-bias expression in prenatal stages [59], or if there was sex-specific splicing in our geneset [60] using the bioconductor package GeneOverlap [61].

Comparison with psychiatric and peripheral diseases

Our functional analyses of the ASD genesets showed enrichment for processes in areas relating to cardiac function and insulin. In addition, it is known that many ASD risk genetic factors share molecular pathways with other psychiatric conditions. To future explore this, we downloaded genesets for schizophrenia (SCZ), bipolar disorder (BP), and major depressive disorder (MDD) from Psygenet [62]. We also downloaded genelists from Harmonizome database [63, 64] for 4 peripheral conditions of arrhythmia (ARY), type 1 (T1D) and type 2 (T2D) diabetes based on our results, and inflammatory bowel disease (IBD) based on co-morbidity of gastrointestinal issues with ASD for comparison of ASD risk factors in the current study.

Results

To explore the hypothesis that ASD candidate genes were recurrent across different scoring systems, we selected the SFARI dataset and four other published scoring systems for the current study (Table 1, Fig 1). Consequently, we selected 3126 genes out of 15735 from the EXAC dataset with pLi ≥ 90% chance of intolerance to loss of function [31], 3225 genes from Krishnan’s data [32] based on q < 0.05, 7189 genes out of 15950 from Zhang’s data [33] with a positive DAMAGE score (D > 0), and 2111 genes from the top 10% of the 21115 genes in the Duda’s data [34], and all 1053 SFARI genes as of Jan 2019 [28].

These shortlists were subsequently analysed by Jvenn web tool for overlapping ASD genes [38], and 114 genes were found to be shared by all five shortlists (Fig 2, S6 Table). Considering that some well-known ASD genes such as SHANK3 and CHD8 were excluded from the 114 geneset, we adjusted to include the genes that were overlapped in 4 of the 5 scoring systems, which resulted in a total to 519 ASD risk genes. Further examination of the 519 genes for the
presence in the SFARI dataset has narrowed down the list to 319 SFARI genes, which were highly ranked in >3 of 4 other independent studies besides the SFARI dataset (Table 1, Fig 1).

Expression pattern of ASD genes

To assess the gene expression across different tissues, we next examined body-wide expression of the 319 genes using the GTEx dataset containing standardized mRNA expression in units of TPM. This further reduced the ASD genes to 292 genes to filter out low abundance genes with TMP $\geq 3$ in both CNS and PT (S7 Table).

The expression levels of the 292 genes were presented in the Heatmap, with high expression coloured in red, low expression in green and no expression in grey. Analysis of the 292 genes with GTEx (S8 Table) showed that the ASD genes were clustered into 2 groups (Fig 3); 91 genes were mainly expressed in the CNS with TMP $< 3$ in PT, whereas the remaining 201 genes were ubiquitously expressed in both the CNS and PT.

A similar expression profile for proteins was observed in HPA data (Figs 4 and 5), whereas CNS geneset showed protein expression mostly in the CNS (Fig 4), the CNS+PT geneset showed protein expression in both the CNS and other tissue types (Fig 5). We also identified that some of the ASD factors in the CNS+PT geneset (S1–S19 Figs) were indeed interacting with other proteins in tissue-specific networks (Table 2). Interestingly HDAC4 (S6) in the colon and STX1A (S15) in the pituitary gland were found to be tissue-specific genes in these organs.

Enriched neurodevelopment, synaptic function, and ion transport in the CNS geneset

To assess the biological context of ASD factor, the STRING (v 11) program was used to analyse two groups of the ASD factors, respectively [65]. The 91 CNS-specific geneset gave rise to 305 “Biological Processes” (S10 Table), 73 “Cellular Components” (S11 Table) and 62 “Molecular Function” (S12 Table). Go term results revealed an enrichment for neuronal development and synaptic function in the CNS geneset (Fig 6A).

We also visualized the interaction network using Cytoscape (V3.7). Among the 91 CNS genes (Table 3), 47 (colored in blue) were involved in “Nervous system development” (FDR = 1.39E-18), 31 (colored in red) were linked to “Trans-synaptic signalling” (FDR = 2.41E-25), and 30 were associated with “Ion transmembrane transport” (colored in yellow, FDR = 1.55E-14, Fig 6A), which included 6 calcium channels (ATP2B2, CACNA1A,
CACNA1B, CACNA1D, CACNA1G, CACNA2D3), 3 sodium channels (SCN1A, SCN2A, SCN8A), 4 potassium (HCN1, KCND2, KCNQ2, KCNQ3) channels, 6 glutamatergic receptors (GRIA1, GRID1, GRIK2, GRIN1, GRIN2A, GRIN2B), 5 GABAergic receptors (GABRA1, GABRA3, GABRA4, GABRA5, GABRB3) and 5 transporters (SLC12A5, SLC1A2, SLC24A2, SLC30A3, SLC4A10).

Consistently, the synapses (42/91 genes, FDR = 4.06E-30), neuronal projection (44/91, FDR = 9.54E-28) and ion channel complex (24/91, FDR = 7.59E-22, Table 3) were enriched in “Cell Components”. The ion-gated channel activity (24/91, FDR = 5.78E-19) and neurotransmitter receptor activity (13/91, FDR = 3.60E-13) including glutamate (7/91, 2.95E-09, GRIA1, GRID1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1) and GABA receptor activity (6/91, 4.12E-08, GABBR2, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3), were the most significant “Molecular Functions”. The “KEGG pathway” analyses (S13 Table) demonstrated that Glutamatergic synapse (12/91, 2.46E-11, CACNA1A, CACNA1D, DLGAP1, GRIA1, GRIK2, GRIK1, GRIN2A, GRIN2B, GRM1, HOMER1, SHANK2, SLC1A2), GABAergic synapse (11/91, 3.44E-11, CACNA1A, CACNA1B, CACNA1D, GABBR2, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3, GAD1, SLC12A5), Calcium signalling pathway (11/91, 1.80E-08, ATP2B2, CACNA1A, CACNA1B, CACNA1D, CACNA1G, CAMK2A, ERBB4, GRIK1, GRIN2A, GRM1, NOS1) and Circadian entrainment (9/91, 1.74E-08, CACNA1D, CACNA1G, CAMK2A,
GRIA1, GRIN1, GRIN2A, GRIN2B, NOS1, NOS1AP) were the top pathways in the CNS geneset (Table 3). Together, these data suggest that the 91 CNS-specific ASD risk factors are involved in regulation of brain development, E/I balance and calcium signalling, which are closely related to the ASD core features, and to the CNS comorbidity such as epilepsy, intellectual disability and sleeping disorders.

**Enriched chromatin organisation and gene regulation in CNS+PT geneset**

Analyses of the 200 CNS+PT genes resulted in 546 "Biological Processes" (S14 Table), 127 "Cellular Components" (S15 Table), and 123 "Molecular Function" (S16 Table). Like the CNS geneset, CNS development and synapse are the top pathways in the CNS+PT ASD geneset. This included Nervous system development (72/200 genes in blue, FDR = 7.50E-17, Fig 6B), Modulation of chemical synaptic transmission (20/200, FDR = 1.74E-08) and Trans-synaptic signalling (22/200, FDR = 3.16E-08) as top "Biological Processes", Synapse (46/200 in red, FDR = 6.44E-18, Fig 6B) as the most significant "Cellular Components", Ion binding (106/200, FDR = 2.09E-08) in the "Molecular Function" (Table 3), and Long-term potentiation (6/200, FDR = 0.003), Glutamatergic (7/200, FDR = 0.006), and Dopaminergic synapse (7/200, FDR = 0.011) identified in the "KEGG" pathways (Table 3, S17 Table).

However, the most prominent feature of the CNS+PT ASD geneset was transcription regulation, and this included Nucleus (128/200, FDR = 1.73E-14) in the "Cellular Components",
Fig 5. Protein expression of CNS+PT geneset from HPA across tissues. Genes are displayed in order of decreasing expression.

https://doi.org/10.1371/journal.pone.0242773.g005
### Table 2. Interacting ASD genes in tissue-specific networks

<table>
<thead>
<tr>
<th>Tissue</th>
<th>ASD genes</th>
<th>Interacting partners</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart left ventricle</td>
<td>TCF4, RFX3, HDAC4, <strong>POGZ</strong>, <strong>ANKRD11</strong>, RERE, YY1, CMIP, QRICHI</td>
<td>MYH7B, TCF24, ASB15, SNRPC, TRIM54, FSD2, FHL2</td>
</tr>
<tr>
<td>Heart-left atrial appendage</td>
<td>HDAC4, <strong>ANKRD11</strong>, TCF4, CMIP, QRICHI</td>
<td>TRIM54, ASB15, MYH7B, TNN1, TCF24, FSD2</td>
</tr>
<tr>
<td>Artery-tibial</td>
<td><strong>ANKRD11</strong></td>
<td>NOV</td>
</tr>
<tr>
<td>Small Intestine-terminal ileum</td>
<td><strong>POGZ</strong>, CLSTN3, NSD1</td>
<td>A1CF, AO1C, BCL2L15, AGR2, OLFM4</td>
</tr>
<tr>
<td>Pancreas</td>
<td>MEIS2, DAGLA, STX1A, TSC1, ARNT2, TCF4, <strong>POGZ</strong>, CLSTN3, <strong>ANKRD11</strong></td>
<td>FAM136A, LHPP1L5, PNLP1R1, SERP1, ENKD1, SIM1, NEUROD, BCL2L15, BANF1, A1CF, OLFM4, TMEM97</td>
</tr>
<tr>
<td>Stomach</td>
<td><strong>PRICKLE1</strong>3, <strong>ANKRD11</strong>, TCF4, <strong>POGZ</strong>, NSD1</td>
<td>BCL2L15, ODAM, AGR2, JRK</td>
</tr>
<tr>
<td>Esophagus-mucosa</td>
<td>Dyrk1A, ARNT2, WAC, CATNAL1, TSC1, RNF38, EXOC5, YY1, HDAC4, MEIS2, <strong>POGZ</strong>, <strong>ANKRD11</strong></td>
<td>DTX2, BICD2, USH1G, TXN, CYST1T, LGAL57B, LGALS7</td>
</tr>
<tr>
<td>Esophagus-muscularis</td>
<td>USP7</td>
<td>ANKS1A</td>
</tr>
<tr>
<td>Muscle-skeletal</td>
<td>NPSA2, TSC1, MEIS2, TCF4, CMIP, <strong>POGZ</strong>, <strong>ANKRD11</strong>, SRSF11, HDAC4, STX1A, QRICHI1, CACNB2</td>
<td>KRT31, ATG4A, CDR2L, TRIB3, MYF5, FAM22B, FAM166B, BICD2, ME2FA, ME2FC, GOLGA2, SXX1P, AES, NGLY1, YWHAE, DUPD1, FHL3, TADA2B, TRIM27, CALCOCO2, INCA1, OIPS, LBX1, NFKBID, ASB15, CEPI, USP6, ZMYND12, VPS2, HEXIM2, TRIM54, FSD2, FCHSD2, KLHL38, HOOK2</td>
</tr>
<tr>
<td>Colon-sigmoid</td>
<td>TCF4, <strong>POGZ</strong>, QRICHI1, TSC1, YY1, SAE1</td>
<td>ME2FC, MOEOX2, CDDNBDP1, TRIM27, ODAM, INCA1, CDPF1, PAX8, FCHSD2, CDC3C16, NFKBID, CDR2L, HOOK2, CDR2, NFYC, GOLGA2, VPS22, BICD2, TFP11, BZLF1, CALCOCO2, PICK1, CCD3C15, CADPS, MTUS2, YWHAE, AES, SXX1P, BEGAIN, CEPI, USP6, FSD2, ZMYND12, TSGA10, HAPI</td>
</tr>
<tr>
<td>Colon-transverse</td>
<td>HDAC4, SAE1, <strong>POGZ</strong>, <strong>ANKRD11</strong>, CLSTN3, EXOC5, KATNAL1, MEIS2, NSD1, SATB2, TCF4</td>
<td>SATB2, BCL2L15, KLC4, PTK6, A1CF, ABHD11, AGR2, NXP2E2, AO1C</td>
</tr>
<tr>
<td>Kidney</td>
<td>CACNB2, TCF4, SAE1, <strong>POGZ</strong>, MEIS2, PHF12</td>
<td>MCCD1, AO1C, ATP6V0D2, CLCNKA, TMEM174, PAX8, A1CF</td>
</tr>
<tr>
<td>Pituitary</td>
<td>QRICHI1, CIC, YY1, BAZ2B, USP7, USP15, HDAC4, <strong>ANKRD11</strong>, EPI400, STXB1P, STX1A, MBDS, DYSPL2, WASFI1, <strong>POGZ</strong>, DYSPL3, DAGLA, TCF4, MEIS2, DYNCHH1, NPSA2, EXOC5</td>
<td>RAB3LI1, BLOC156, ZNF696, ZNF440, PLN, SEC22A, TMEM254, KRT40, RMDN2, CIGALT1, NAPB, ZNF6, ZNF250, APOL2, STX12, VSTM4, BRD8, NEUROD4, AOC3, CENPP, CASC4, NINJ2, VAMP1, CLCNKA, TMEM41A1, ZNF136, AB13, STX7, STX2, VTT1B, STX4, CD81, EXOC5, APOL3, RHBD2D2, ARL13B, VAMP4, STX3, BET1, ZNF12, TRAP3P1, UBE2I, MAPK1, PGAP2, EBA9, ZNF65, SMIM1, DEUP1, DDX49, STX10, ANKD46, TRIM38, EFHC1, SERP1, C2orf82, CLEC1A, AIG1, CLIN6, TXLNA, SERP2, VAMP5, JAGN1, TMEM120A, TSNARE1, CAC1, CAC2, HOXH1, ATRAP, SNAP47, STX16, USE1, CXCL16, B1T2N2A2, MAL, ZFP11, FAMM3, BPIFA1, SPICE1, GOLGA2, CYB5B, GMAPS, TMEM128, STXI1, NGK7, STX6, LHPP5L, CMTM7, STX5, NRM, AQPS, VAMP3, BNIP1, LHX3, ETNK2, RN4, TXB1, ZNF835, CDC37, ZNF441, KIFC3, STX3, TMEM100, ZNF707, TMEM60</td>
</tr>
<tr>
<td>Lung</td>
<td>PHF12</td>
<td>ATP6V0D2</td>
</tr>
<tr>
<td>Artery coronary</td>
<td><strong>ANKRD11</strong></td>
<td>NOV</td>
</tr>
<tr>
<td>Artery aorta</td>
<td><strong>ANKRD11</strong></td>
<td>NOV</td>
</tr>
<tr>
<td>Adipose subcutaneous</td>
<td>ARNT2</td>
<td>SIM1</td>
</tr>
<tr>
<td>Spleen</td>
<td>SAE1, <strong>ANKRD11</strong>, TCF4, WASFI1, CLSTN3, <strong>POGZ</strong>, STX1A, MEIS2, CAMK2B, TLK2, KHD83S, MARK1</td>
<td>POU2AF1, INPP5D, PAX5, DOCK2, NFAM1, GRB2, CCM2L, TCF23, TRAF3P3, FOLR3, NKG7, RASAL3, AB13, TCI1B</td>
</tr>
<tr>
<td>Adrenal Gland</td>
<td>TCF4, STX1A, EPI400, QRICHI1, TSC1, <strong>ANKRD11</strong></td>
<td>MTFRII, TMEM41A, NOV, CHCHD2, TRAPPCC2L, RMDN2, FAM166B, ALAS1</td>
</tr>
</tbody>
</table>

Regulation of gene expression (96/200 genes in red, FDR = 4.21E-12, Fig 6B), Chromatin organization (44/200 in green, FDR = 1.44E-18, Fig 6B) and Histone modification (26/200, FDR = 2.08E-12) identified in the "Biological Processes", Chromatin binding (29/200, FDR = 3.00E-11), Transcription factor binding (24/200, FDR = 2.26E-06), DNA binding (53/
In addition, Circadian entrainment (5/200 genes, FDR = 0.034, CACNA1C, CAMK2B, ITPR1, PLCB1, PRKCB), WNT signalling (10/200, FDR = 3.1E-04, APC, CAMK2B, CHD8, CREBBP, CSNK1E, GSK3B, PLCB1, PRICKLE1, PRICKLE2, PRKCB), Thyroid hormone signalling (8/200, FDR = 1.5E-03, ATP1A3, CREBBP, GSK3B, MED13, MED13L, NCOR1, PLCB1, PRKCB), Aldosterone synthesis and secretion (8/200, FDR = 5.0E-04, ATP1A3, CACNA1C, CAMK2B, DAGLA, ITPR1, NR4A2, PLCB1, PRKCB), Gastric acid secretion (5/200, FDR = 0.0179, ATP1A3, CAMK2B, ITPR1, PLCB1, PRKCB), Insulin secretion (9/200, FDR = 7.16E-05, ATP1A3, CACNA1C, CAMK2B, KCNMA1, PLCB1, PRKCB, RAPGEF4, SNAP25, STX1A), Salivary secretion (5/200, FDR = 0.0298, ATP1A3, ITPR1, KCNMA1,
<table>
<thead>
<tr>
<th>Term ID</th>
<th>Term description (Background Gene Count)</th>
<th>ASD genes (91 genes) centered on brain development, synapse and ion transport</th>
<th>FDR</th>
<th>Matching proteins in the network (IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0007399 (Biol. Proc.)</td>
<td>Nervous system development (2206)</td>
<td>47 (CNS)</td>
<td>1.39E-19</td>
<td>ATP2B2, BCL11A, CAMK2A, CNTN6, CNTNP2, CNTNA2, CNTND2, CUX2, DAB1, DLX1, DSCAM, ELAVL3, ERBB4, FEZ2, FOXG1, GABRA4, GABRA5, GABRB3, GAP43, GDA, GRIN1, GRIN2A, GRIN2B, HCN1, KCNQ2, KIRREL3, LRRC7, MYT1L, NOS1, NRXN1, PLXNA4, RBPFOX1, RELN, RIMS1, ROBO2, SCN2A, SCN8A, SHANK2, SLC12A5, SLC1A2, SLC4A10, SLITRK5, SRRM4, SYT1, SYT17, TBR1, UNC13A</td>
</tr>
<tr>
<td>GO:0043005 (Cell. Comp.)</td>
<td>Neuron projection (1142)</td>
<td>44 (CNS)</td>
<td>9.54E-28</td>
<td>ANKS1B, CACNA1B, CADM2, CAMK2A, CDH8, CNKSR2, CNTNP2, CNTNA2, CNTND2, DAB1, DSCAM, FRMPD4, GABRB2, GABRA5, GAD1, GAP43, GRIAT1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1, HCN1, HOMER1, KCND2, KCNQ2, KCNQ3, KIRREL3, LRRC4, LRRC7, NOS1, NRXN1, RELN, ROBO2, SCN1A, SCN2A, SCN8A, SHANK2, SLC12A5, SLC1A2, SLC4A10, SYT1, UNC13A</td>
</tr>
<tr>
<td>GO:0099537 (Biol. Proc.)</td>
<td>Trans-synaptic signalling (408)</td>
<td>31 (CNS)</td>
<td>2.41E-25</td>
<td>CACNA1B, CACNA1D, CDH8, DLGAP1, DLGAP2, GABBR2, GABRA1, GABRA5, GABRB3, GAD1, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1, HOMER1, KCND2, KCNQ2, KCNQ3, NOS1, NRXN1, RIMS1, SLC12A5, SLC1A2, SLITRK5, SYT1, SYT17, UNC13A</td>
</tr>
<tr>
<td>GO:0045202 (Cell. Comp.)</td>
<td>Synapse (107)</td>
<td>42 (CNS)</td>
<td>4.06E-30</td>
<td>ANKS1B, ATP2B2, CACNA1B, CADM2, CAMK2A, CDH8, CNKSR2, CNTNA2, DAB1, DLGAP1, DLGAP2, DSCAM, FRMPD4, GABRB2, GABRA1, GABRA3, GABRA4, GABRB3, GAD1, GAP43, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1, HOMER1, KCND2, KCNQ2, KCNQ3, NOS1, NRXN1, RIMS1, SHANK2, SLC30A3, SYT1, SYT17, UNC13A</td>
</tr>
<tr>
<td>GO:0030594 (Mol. Funct.)</td>
<td>Neurotransmitter receptor activity (849)</td>
<td>13 (CNS)</td>
<td>3.11E-13</td>
<td>GABRB2, GABRA1, GABRA3, GABRA4, GABRB3, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1</td>
</tr>
<tr>
<td>GO:0008066 (Mol. Funct.)</td>
<td>Glutamate receptor activity (27)</td>
<td>7 (CNS)</td>
<td>2.73E-09</td>
<td>GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1</td>
</tr>
<tr>
<td>GO:0016917 (Mol. Funct.)</td>
<td>GABA receptor activity (22)</td>
<td>6 (CNS)</td>
<td>3.86E-08</td>
<td>GABRR2, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3</td>
</tr>
<tr>
<td>GO:0034220 (Biol. Proc.)</td>
<td>Ion transmembrane transport (995)</td>
<td>30 (CNS)</td>
<td>1.55E-14</td>
<td>ATP2B2, CACNA1A, CACNA1B, CACNA1D, CACNA1G, CACNA2D3, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1, KCND2, KCNQ2, KIRREL3, LRRC4, LRRC7, NOS1, NRXN1, RIMS1, SHANK2, SLC12A5, SLC1A2, SLC24A2, SLC30A3, SLC4A10</td>
</tr>
<tr>
<td>GO:0034702 (Cell. Comp.)</td>
<td>Ion channel complex (278)</td>
<td>24 (CNS)</td>
<td>7.59E-22</td>
<td>CACNA1A, CACNA1B, CACNA1D, CACNA1G, CNTNP2, DPP10, GABRA1, GABRA3, GABRA4, GABRA5, GBRR3, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIK1, GRIN2A, GRIN2B, HCN1, KCND2, KCNQ2, KCNQ3, KCNQ4, SCA1A, SCA2A, SCA8A</td>
</tr>
<tr>
<td>GO:0022839 (Mol. Funct.)</td>
<td>Ion gated channel activity (329)</td>
<td>24 (CNS)</td>
<td>4.34E-19</td>
<td>CACNA1A, CACNA1B, CACNA1D, CACNA1G, CACNA2D3, GABRA1, GABRA3, GABRA4, GABRA5, GBRR3, GLRA2, GRIA1, GRIK1, GRIK2, GRIN1, GRIN2A, GRIN2B, HCN1, KCND2, KCNQ2, KCNQ3, KCNQ4, SCA1A, SCA2A, SCA8A</td>
</tr>
<tr>
<td>GO:0005262 (Mol. Funct.)</td>
<td>Calcium channel activity (114)</td>
<td>9 (CNS)</td>
<td>7.18E-08</td>
<td>CACNA1A, CACNA1B, CACNA1D, CACNA1G, CACNA2D3, GABRA1, GABRA2, GABRA3, SLC2A4A2</td>
</tr>
<tr>
<td>GO:0007399 (Biol. Proc.)</td>
<td>Nervous system development (2206)</td>
<td>72 (CNS+PT)</td>
<td>7.50E-17</td>
<td>ANK2, ANK3, APBB1, ARHPAG3P, ARID1B, ARNT2, ATRA, AUTS2, CAMK2B, CAMSAP2, CHD5, CHD7, CIC, CLSTN3, CNR1, CNTN4, CSNK1E, DAGLA, DLG4, DMD, DYSPL2, DYSPL3, DYRK1A, EPHB2, EXT1, GIGYF2, GSK3B, HDAC4, JARID2, KDM6A, KDM6B, KIFSC, KIT, MAP2, APK1P2, MARK1, MBDS, MECPR2, MYH10, MYO5A, NAV2, NEGR1, NEO1, NFI, NFB, NIPBL, NLGN2, NPAS2, NR4A2, NRKAM, NRXN2, NRXN3, NTRK3, OPHN1, PLCB1, PRICKLE1, PTBP2, RERE, SATB2, SETD2, SLIT3, SMARCA2, SMARCC2, SNAP25, SPAST, STXBP1, SYN1, TAOK2, TCF4, TSC1, WASF1</td>
</tr>
<tr>
<td>GO:0050804 (Biol. Proc.)</td>
<td>Modulation of chemical synaptic transmission (316)</td>
<td>20 (CNS+PT)</td>
<td>1.74E-08</td>
<td>CAMK2B, CLSTN2, CLSTN3, CNR1, CNTN4, DLG4, EPHB2, GRK5, KCNBI1, KIT, MAPK4P2, MECPR2, MEF2I, NLGN2, OPHN1, RIMS3, SNAP25, STX1A, STXBP1, SYN1</td>
</tr>
<tr>
<td>GO:0099537 (Biol. Proc.)</td>
<td>Trans-synaptic signalling (408)</td>
<td>22 (CNS+PT)</td>
<td>3.16E-08</td>
<td>ARID1B, CACNB2, CASK, CLSTN3, CNR1, DAGLA, DLG4, EPHB2, GRK5, GSK3B, MAPK4P2, MECPR2, MYO5A, NFI, NRXN2, RIMS3, SLC6A1, SNAP25, STX1A, STXBP1, SYN1, SYNJ1</td>
</tr>
</tbody>
</table>

(Continued)
Table 3. (Continued)

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Term description (Background Gene Count)</th>
<th>ASD genes</th>
<th>FDR</th>
<th>Matching proteins in the network (IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0045202 (Cell. Comp.)</td>
<td>Synapse (849)</td>
<td>45 (CNS +PT)</td>
<td>3.25E-17</td>
<td>ANK2, ANK3, APBB1, ARHGAP32, ARHGAP33, ATP1A3, CASCA1C, CADM1, CAMK2B, CASK, CLSTN2, CLSTN3, CNR1, CPEB4, DAGLA, DLG4, DMD, DMXL2, EPHB2, GPHN, GRIK5, GSK3B, HDAC4, ITPR1, KCNB1, MAP2, MAPK8IP2, MECP2, MYH10, NF1, NGLN2, NRCAM, NRXN2, OPIN1, PDE4B, PSD3, RIMS3, SNAP25, STX1A, STXBP1, STXBP5, SYN1, SYNE1, SYNJ1, WASF1</td>
</tr>
<tr>
<td>GO:0005634 (Cell. Comp.)</td>
<td>Nucleus (6892)</td>
<td>128 (CNS +PT)</td>
<td>1.73E-14</td>
<td>ANKRD11, APBP1, APC, ARHGAP32, ARID1B, ARNT2, ASH1L, ATRX, AUTS2, BA22, BRD4, BTAF1, CAMK2B, CAMTA1, CASK, CDK13, CHD3, CHD5, CHD7, CHD8, CIC, CMIP, CPEB4, CREBBP, CSNK1E, CTFC, DCUN1D1, DDX3X, DMD, DOCK4, DRTK1A, EHM1, EP400, EPC2, EPB2, FBXO11, FOXP1, GGNBP2, GRIK5, GSK3B, HCF1C, HDAC4, HERC2, HRNRNP1, HUWE1, INTS6, IRF2BP1, ITPR1, JARID2, JMJD1C, KAT6A, KATNAL1, KDM4B, KDM6A, KDM6B, KHDRBS3, MAP2, MBDS, MECP2, MED13, MED13L, MEIS2, MKL2, NAV2, NCO1, NEO1, NF1, NF1B, NF1X, NIPBL, NPAS2, NR2F1, NR3C2, NRG1, NRG2, NRG3, OGD1, OGD2, OSM, PAF2, PHF2, PHF8, PRC1, PRC2, PRKCB, PRPL3, PRPF9, PTBP2, QRICH1, RA11, RB1CC1, RBM27, RERE, RFX3, RNF38, SAE1, SATB2, SETBP1, SETD2, SMARC2, SMARC2, SPAST, SBF511, STX1A, STXBP1, SYN1, SYNE1, TAO2, TCF4, TGL2, TOP1, TRIM33, TRIP12, TSC1, UBN2, UBR5, UPF2, USP15, USP7, WAC, WD-F3, WDR26, YEATS2, YY1, ZMYND11, ZNF462, ZNF827</td>
</tr>
<tr>
<td>GO:0006325 (Biol. Proc.)</td>
<td>Chromatin organization (683)</td>
<td>44 (CNS +PT)</td>
<td>1.44E-18</td>
<td>APBB1, ARID1B, ASH1L, ATRX, AUTS2, BRD4, CHD3, CHD5, CHD7, CHD8, CREBBP, CTFC, EHM1, EP400, EPC2, HCF1C, HDAC4, HRNRNP1, HUWE1, JARID2, JMJD1C, KAT6A, KDM4B, KDM6A, KDM6B, MECP2, NCO1, NRG1, OGD1, PHF2, PRKCB, RERE, SATB2, SETD2, SMARC2, SMARC2, TGL2, TOP1, USP15, USP7, WAC, YEATS2, ZMYND11, ZNF462</td>
</tr>
<tr>
<td>GO:003682 (Mol. Func.)</td>
<td>Chromatin binding (501)</td>
<td>29 (CNS +PT)</td>
<td>2.62E-11</td>
<td>APBB1, ASH1L, ATRX, AUTS2, BRD4, CHD3, CHD5, CREBBP, CTFC, EHM1, EP400, EPC2, HCF1C, HDAC4, HRNRNP1, JARID2, JMJD1C, KAT6A, KDM4B, KDM6A, KDM6B, MECP2, NCO1, NRG1, OGD1, PHF2, PRKCB, RERE, SATB2, SMARC2, TOP1, WAC, ZMYND11</td>
</tr>
<tr>
<td>GO:0016570 (Biol. Proc.)</td>
<td>Histone modification (347)</td>
<td>26 (CNS +PT)</td>
<td>2.08E-12</td>
<td>APBB1, ASH1L, CHD3, CHD5, CREBBP, EHM1, EP400, EPC2, HCF1C, HDAC4, HRNRNP1, JARID2, JMJD1C, KAT6A, KDM4B, KDM6A, KDM6B, MECP2, NCO1, NRG1, OGD1, PHF2, PRKCB, RERE, SATB2, SMARC2, TOP1, USP15, USP7, WAC, YEATS2</td>
</tr>
<tr>
<td>GO:0042393 (Mol. Func.)</td>
<td>Histone binding (188)</td>
<td>12</td>
<td>2.99E-05</td>
<td>APBB1, ATRX, BRD4, CHD5, CHD8, PHF2, PHIP, PRKCB, SMARC2, USP15, YEATS2, ZMYND11</td>
</tr>
<tr>
<td>GO:0008134 (Mol. Func.)</td>
<td>Transcription factor binding (610)</td>
<td>24</td>
<td>2.05E-06</td>
<td>APBB1, ARNT2, CDK13, CREBBP, DDX3X, FOXP1, GSK3B, HCF1C, HDAC4, HRNRNP1, JARID2, JMJD1C, KAT6A, MECP2, MED13, MEIS2, NCO1, NRG1, NRG2, PRKCB, TGF4, TRIP12, USP7, YEATS2</td>
</tr>
<tr>
<td>GO:0010468 (Biol. Proc.)</td>
<td>Regulation of gene expression (4533)</td>
<td>96 (CNS +PT)</td>
<td>4.21E-12</td>
<td>ANK2, ANK3, APBB1, ARID1B, ARNT2, ASH1L, ATRX, AUTS2, BA22, BRD4, BTAF1, CAMTA1, CDK13, CHD3, CHD5, CHD7, CHD8, CIC, CREBBP, CTFC, DAPK1, DDX3X, DRTK1A, EHM1, EP400, EPC2, FOXP1, GGNBP2, GIFY2, GSK3B, HCF1C, HDAC4, HRNRNP1, IRF2BP1, JARID2, JMJD1C, KAT6A, KDM4B, KDM6A, KDM6B, KHDRBS3, KIT, MECP2, MED13, MED13L, MEIS2, MKL2, NCO1, NEO1, NF1, NF1B, NF1X, NIPBL, NPAS2, NR2F1, NR3C2, NRG1, NRG2, NRG3, OGD1, OGD2, PAF2, PHF2, PHIP, PLCB1, POZG, PRICKE1, PRKCB, PTBP2, QRICH1, RA11, RB1CC1, RERE, RFX3, SATB2, SETBP1, SETD2, SLIT3, SMARC2, SMARC2, TCF4, TNRCE8B, TOP1, TRIM33, TSC1, UBR5, USP15, USP7, VIP, WAC, YEATS2, YY1, ZMYND11, ZNF462, ZNF827</td>
</tr>
<tr>
<td>GO:0003677 (Mol. Func.)</td>
<td>DNA binding (2457)</td>
<td>53</td>
<td>5.62E-06</td>
<td>ARID1B, ARNT2, ASH1L, ATRX, BA22, BTAF1, CAMTA1, CDK13, CHD3, CHD5, CHD7, CHD8, CIC, CREBBP, CTFC, DDX3X, EP400, FOXP1, HDAC4, HRNRNP1, HUWE1, JARID2, JMJD1C, KAT6A, KDM6A, KDM6B, MECP2, MEIS2, NCO1, NF1B, NF1X, NPAS2, NR2F1, NR3C2, NRG1, NRG2, NRG3, OGD1, OGD2, PAF2, PRICKE1, RA11, RERE, RFX3, SATB2, SETBP1, SMARC2, SMARC2, TCF4, TOP1, TRIM33, USP15, YY1, ZMYND11, ZNF462, ZNF827</td>
</tr>
</tbody>
</table>

(Continued)
Table 3. (Continued)

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Term description (Background Gene Count)</th>
<th>ASD genes</th>
<th>FDR</th>
<th>Matching proteins in the network (IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0043167 (Mol. Func.)</td>
<td>Ion binding (6066)</td>
<td>106 (CNS +PT)</td>
<td>1.42E-08</td>
<td>AGAP1, ARIHGAP, ASAP2, ASH1L, ATP1A3, ATP8A1, ATRX, BAZ2B, BTAF1, CACNA1C, CACNA2D1, CAMK2B, CASK, CDC42BPB, CDK13, CHD3, CHD5, CHD7, CHD8, CLSTN2, CLSTN3, CPEB4, CREBBP, CSNK1E, CTCF, DAGLA, DAPK1, DDX3X, DMD, DYSF, DST, DYNCH1L, DYRK1A, EHMT1, EP400, EPHB2, EXT1, FXO1, FGD1, FOXP1, GPN1, GSK3B, HADA, HERC2, HNRNP1, IRF2BPL, ITPR1, JMJ1C, KAT6A, KATNAL1, KCND3, KCNMA1, KDM4B, KDM6A, KDM6B, KIF21B, KIF5C, KIT, MARK1, MYH10, MYO5A, NAV2, NFI, NPS2, NR2F1, NR3C2, NR4A2, NRXN2, NRXN3, NSD1, NTRK3, OGT, OPHN1, PDE4A, PDE4B, PHF2, PHF3, PLCB1, POGZ, PRKCI, PRKCI, PRKCB, RAII, RAPGEF4, RBM27, RERE, RNF38, SETD2, SLC5A1, SLIT3, SMARCA2, SPAST, SYN1, TAOK2, TTK, TOP1, TRIM33, TRIO, UBR5, WDFY3, YTHDC2, YY1, ZMYND11, ZNF462, ZNF827</td>
</tr>
</tbody>
</table>

PLCB1, PRKCB) and Pancreatic secretion (5/200, FDR = 0.0343, ATP1A3, ITPR1, KCNMA1, PLCB1, PRKCB) were also detected as significant KEGG pathways in the CNS+PT ASD geneset.

In consistency with this, STRING analysis of the combined 291 ASD genes gave rise to 47 "KEGG pathways" (S18 Table), 677 "Biological Processes" (S19 Table), 149 "Cellular Components" (S20 Table) and 177 "Molecular Function" (S21 Table). The top enriched pathways of the combined ASD risk factors included Nervous system development (119/291 genes, FDR = 8.85E-34), Synapse (87/291, 1.9E-43), Trans-synaptic signalling (53/291, 1.04E-28), Ion transport (44/291, 4.79E-15) and Neurotransmitter receptor activity (14/291, 8.86E-08). The CNS+PT ASD geneset was also enriched for Secretion, Thyroid function, and Cellular signalling. Interestingly, 10 pathways from both ASD genesets appeared to overlap, which included Glutamatergic, Dopaminergic, Cholinergic synapses, Circadian entrainment, cAMP and Retrograde endocannabinoid signalling (Table 3). Calcium channels, Glutamatergic and GABAergic receptors appeared to link to most of the pathways.

E/I balance, and calcium signalling are central to ASD

KEGG analyses generated 26 pathways for the CNS (S13 Table) geneset and 31 pathways for the CNS+PT geneset (S17 Table). The top pathways from the CNS geneset corresponded to Glutamatergic synapse (12/91 genes, 2.46E-11), GABAergic synapse (11/91, 3.44E-11), Neuroactive ligand-receptor interaction (14/91, 1.42E-09), Retrograde endocannabinoid signalling (11/91, 3.77E-09), Calcium signalling pathway (11/91, 1.80E-08) and MAPK signalling (6/91, 0.0105). The CNS+PT ASD geneset was also enriched for Secretion, Thyroid function, and Cellular signalling. Interestingly, 10 pathways from both ASD genesets appeared to overlap, which included Glutamatergic, Dopaminergic, Cholinergic synapses, Circadian entrainment, cAMP and Retrograde endocannabinoid signalling (Table 3). Calcium channels, Glutamatergic and GABAergic receptors appeared to link to most of the pathways.

To further investigate the interconnectivity among the KEGG pathways (S18 Table), we constructed an interaction matrix with the combined ASD geneset KEGG pathways (S22 Table), similar to a previous analysis [30]. The calcium signalling (27/47), MAPK Signalling (12/47) and cAMP signalling (6/47) were identified as highly interconnected signalling pathways. Pre-synaptically, the genes in calcium signalling also appeared frequently in the other pathways, particularly the genes encoding the pore-forming subunit of voltage-gated calcium channel (CACNA1A, CACNA1B, CACNA1C, CACNA1D), the 13P receptor IPTR1, PLCB and the calmodulin proteins (CAMK2A, CAMK2B) are connected to the neurotransmitter releases.
Post-synaptically, Glutamatergic synapse (8/47) had the largest number of interactions with other pathways, which was followed by Dopaminergic synapse (7/47). The NDMAR (GRIN2A, GRIN2B, GRIN1) and AMPA (GRIA1) receptors were also the highly interconnected notes in neural and synaptic, calcium and cAMP signalling pathways. In summary, this overlap analyses demonstrated that the E/I balance, and calcium signalling are the most significant pathways linking to the ASD core symptoms, and transcriptional regulation to the ASD comorbidities.

Enriched epilepsy/seizures in the CNS geneset and congenital abnormalities/developmental delay in CNS+PT geneset

The results from WebGestalt (Fig 7, Supplementary links 1 and 2) showed that epilepsy and seizures were enriched in 19 of the top 50 phenotype terms and movement disorders in 9 of the 50 terms in the CNS geneset. In the CNS+PT set, there was enrichment for behaviour issues (9/50) such as self-injurious (FDR = 1.04E-07), aggressive behaviour (1.04E-11), and congenital abnormalities of the face (25/50 items).

High proportion of the ASD genes were dysregulated in ASD

We compared the ASD geneset with DEGs between ASD and controls in the literature and found that 201 of the 292 genes in our ASD geneset were dysregulated, at least once across multiple ASD gene expression studies (Table 4, S23 Table). Recurrent DEGs such as RELN, FOXP1, GAD1, NRXN1, FOXG1 and CAMK2A were present in multiple studies (S23 Table). These data suggest that not only mutations but also dysregulated expression of the ASD risk genes can be linked to the development of ASD.

Table 4. Genes dysregulated in ASD expression studies that overlapped with our geneset.

<table>
<thead>
<tr>
<th>Study</th>
<th>Tissue type</th>
<th>genes overlap</th>
<th>up</th>
<th>down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gupta [66]</td>
<td>Cortical post-mortem</td>
<td>22</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>Parikshak [67]</td>
<td>Cortical post-mortem</td>
<td>24</td>
<td>4</td>
<td>20</td>
</tr>
<tr>
<td>Voineagu [56]</td>
<td>Cortical post-mortem</td>
<td>48</td>
<td>39</td>
<td>9</td>
</tr>
<tr>
<td>Walker (2013) [49]</td>
<td>colon</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Walker (2016) [46]</td>
<td>blood and colon</td>
<td>70</td>
<td>5</td>
<td>37&quot;&quot;</td>
</tr>
<tr>
<td>Mariani [50]</td>
<td>Organoids</td>
<td>89</td>
<td>86'</td>
<td>3'</td>
</tr>
<tr>
<td>Gresi Olivera [54]</td>
<td>IPSC neurons</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>DeRosa [53]</td>
<td>IPSC neurons</td>
<td>15</td>
<td>4'</td>
<td>11'</td>
</tr>
<tr>
<td>Velmeshev [68]</td>
<td>Cortical tissue</td>
<td>38</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td>Herrero [69]</td>
<td>Amygdala post-mortem</td>
<td>9</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Chien [48]</td>
<td>Lymphoblastoid cell lines</td>
<td>9</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Ginsberg [70]</td>
<td>Cortical tissue</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Garbett [45]</td>
<td>Superior Temporal Gyrus</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Breen [55]</td>
<td>IPSC neurons and NPCs</td>
<td>10</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>Wang (2015) [64]</td>
<td>IPSC neurons and NPCs</td>
<td>84</td>
<td>63</td>
<td>21</td>
</tr>
<tr>
<td>Wang (2017) [52]</td>
<td>Organoids</td>
<td>24</td>
<td>17</td>
<td>7</td>
</tr>
<tr>
<td>Pramparo [47]</td>
<td>Leukocytes</td>
<td>15</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

* denotes genes which have simultaneous up/down expression
** denote genes for blood and colon samples respectively.

https://doi.org/10.1371/journal.pone.0242773.t004
ASD genes enriched in cortex development and sex-bias brain expression

We compared ASD shortlist with CSEA dataset and found that 280 of the 292 ASD genes were mapped to CSEA, with an enrichment in the cortex across most timepoints (S24 Table), with most significant enrichment in the early-mid fetal stage of brain development at pSI 0.05 (p-value = 5.427e-16, FDR = 3.256e-14) pSI 0.01 (p-value = 5.560e-07 FDR = 3.336e-05) and pSI 0.001 p-value = 8.912e-04, FDR = 0.053) (Fig 8D). In addition, there was also enrichment of genes in the striatum at the early-mid fetal stage (p-value = 1.872e-08, FDR = 2.808e-07), in the cerebellum at childhood (p-value = 7.576e-07, FDR = 7.576e-06) and adolescence (p-value = 9.970e-04, FDR = 0.005) (Fig 8D). In addition, there was also enrichment of genes in the striatum at the early-mid fetal stage (p-value = 1.872e-08, FDR = 2.808e-07), in the cerebellum at childhood (p-value = 7.576e-07, FDR = 7.576e-06) and adolescence (p-value = 9.970e-04, FDR = 0.005), in the thalamus in early fetal (p-value = 0.010, 0.043) and neonatal period (p-value = 4.333e-04, FDR = 0.003), and in the amygdala from mid-early (p-value = 0.009, FDR = 0.042) to mid-late (p-value = 0.017, FDR = 0.064) fetal stages. These data suggest that the ASD geneset plays essential role in brain development, especially corticogenesis.

To investigate if the ASD genes are sex-related, we compared the ASD geneset with genes which were known to have sex-biased expression in prenatal male and female (S25 Table). Significant correlations were found with genes bias for female cerebellar cortex, dorsolateral...
prefrontal cortex, medial prefrontal cortex, orbital frontal cortex, inferolateral temporal cortex and caudal superior temporal cortex (Fig 8A), and with genes bias for male primary somatosensory cortex, mediodorsal nucleus of thalamus and striatum (Fig 8B). In addition, 34 of the 292 genes were found to have sex-biased gene-splicing in at least one brain region (Fig 8C, S25 Table). These genes are likely to contribute to sex-bias occurrence of the ASD.

**ASD genes present in other conditions**

We next compared the ASD genes with other neuropsychiatric conditions, including schizophrenia, bipolar and major depression, and peripheral conditions such as arrhythmia, inflammatory bowel disease and type 1 and type 2 diabetes (Fig 9, S25 Table), and identified 94 genes which were identified in Psychgenet database (Fig 9A), and involved in synaptic transmission. The remaining ASD-unique genes were enriched for chromatin organization and gene expression. We also found significant overlaps of the ASD genes with factors associated with psychiatric and peripheral conditions (Fig 9B), including a large overlap with type 2 diabetes (134/
Coupling of autism genes to tissue-wide expression and dysfunction of synapse, signalling and gene regulation
These data suggest a common disturbance in neuronal communication with CNS other neuropsychiatric disorders and gene dysregulation with peripheral conditions.

**Discussion**

It is becoming apparent that ASD genes could influence other organ systems. This is reflected by the many co-morbidities occurring outside the CNS such gastrointestinal issues, metabolic disorders, auto-immune disorders, tuberous sclerosis, attention-deficit hyperactivity disorder, and sensory problems associated motor problems. However, little attention has been given to related organs of major comorbidities. Here we have identified 319 overlapping ASD candidates among the four independent scoring systems and the SFARI database [28, 31–34]. We also introduced gene expression using the GTEx database [39, 71], which consists of mRNA data of 53 human tissues from approximately 1000 individuals at the age of 21–70. This resulted in a shortlist of 292 common ASD candidate genes with mRNA expression at TPM \( \geq 3 \) transcripts. This also categorized the ASD factors into 2 genesets, the CNS-specific geneset of 91 genes (with a TMP \(< 3 \) in PT) and the CNS+PT geneset of 201 genes. This was validated at the protein level across these tissues in the human protein atlas (HPA) dataset and Huri, showing that that ASD genes are not only expressed in other organs outside the brain, but also appear to interact with other proteins in tissue-specific networks.

STRING analyses show that the CNS geneset is enriched for nervous development, glutamatergic/GABAergic synapses, and calcium signalling. Phenotype analysis also showed high enrichment for epilepsy and seizures. Both results support the hypothesis that disruption of E/I balance during CNS development as a major feature of ASD [72], which are related to CNS co-morbidities such as epilepsy occurring in 30% of ASD at severe end of the spectrum.

The expression of 201 ASD candidate genes in CNS+PT suggest that ASD genes may influence not only the CNS but also peripheral systems in the body. This genaset is enriched for nervous development and synapse, as well as for chromatin organisation and gene regulation, which are consistent with previous reports from exome sequencing studies [12, 14, 16, 17, 29]. Therefore, the genes involved in chromatin organisation and gene regulation could have an influence in peripheral co-morbidities. Many of the genes in our ASD genaset also show dysregulated expression in multiple studies (Table 4) pertaining to cortical tissue and iPSC derived models, and in the few studies carried on gene expression in the gastrointestinal tract of ASD.

Strong candidate genes such as \textit{CHD8}, \textit{POGZ} and \textit{DYRK1A} were previously reported to be associated with not only autism but also gastrointestinal issues, facial dysmorphisms, visual and feeding problems [73–76]. Indeed, facial dysmorphism is also a recurrent phenotype that emerges among various subgroups of autistics with known genetic mutations [77–79]. Some enrichment of ASD genes reported to overlap with heart development (S10-21) and congenital heart deformation [80, 81], and a high rate of ASD diagnosis was reported among children with congenital heart defects [82]. \textit{POGZ} and \textit{ANKRD11} are also present in many tissue interaction networks (Table 2) which are involved in neural proliferation. The \textit{POGZ} is a zinc finger protein interacting with the transcription factor SP1 [83], and \textit{ANKRD11} is a chromatin regulator, modulating histone acetylation and inhibiting ligand-dependent activation of transcription [84]. \textit{ANKRD11} mutations have been associated with diseases with distinctive craniofacial

Fig 9. Overlap of ASD genes with other conditions. A) Comparison of the ASD geneset with Psychnet (SCZ, BP, MDD) defines a network of 94 overlapping genes (green) and 198 ASD-unique genes (blue). B) matrix of overlaps between ASD/Psychiatric (SCZ, BP, MDD) and peripheral (IBD, ARY, T1D and T2D) conditions. C) Network of ASD genes and overlaps with peripheral (IBD, ARY, T1D and T2D) conditions, with orange circle for non-overlapping genes. IBD—Inflammatory bowel disease, ARY—Arrhythmia, T1D - type 1 diabetes, T2D - type 2 diabetes. SCZ—schizophrenia, BP—bipolar disorder, MDD—major depressive disorder.

292, FDR = 3e-42). These data suggest a common disturbance in neuronal communication with CNS other neuropsychiatric disorders and gene dysregulation with peripheral conditions.
features, short stature, skeletal anomalies, global developmental delay, seizures and intellectual disability [85].

The strong enrichment for neuronal processes and functions in tissues beyond the brain is of curious interest, but researchers are starting to explore other aspects of ASD that could have an impact on other organs such as the heart via the sympathetic and parasympathetic nervous systems [86], and the gastrointestinal tract via the enteric nervous system [87]. In fact there is growing evidence that heart rate is affected among Autistics [88–91], along with evidence that ASD genes could be involved in aspects of gastrointestinal development and function [74, 92–97]. There is a potential for more work on how the parts of the brain control heart rate and gastrointestinal, how they are altered in autism, or even if reported autonomic issues in co-morbidities such gastroesophageal reflux [98, 99] hold true in the autistic population as well. The expression of proteins such as STX1A, SNAP25 and FOXP1 expressed in endocrine tissues could be of interest in ASD, given how knockouts in these genes can impact the development and function of certain parts of the endocrine system [100–102] and how genes involved in neurotransmission could also be involved in secretion of hormones [103].

Another unaddressed question is if the peripheral nervous system and peripheral organs are affected by mutations in addition to the CNS. Amongst the results we found enrichment for neuromuscular and cardiac function in STRING analysis (S9-20). Some animal models such as FOXP1, SHANK3, NOS1 and CHD8 [74, 92, 94, 96, 97] have been developed for functional analysis of ASD candidate genes in the gastrointestinal system, which indicates that ASD genes may play an important role in this organ [95], and yet most research have been mainly focused on the brain in both human and animal models. A greater utilisation of the animal models to explore other systems such as the cardiac and gastrointestinal systems would be welcome.

In fact, 88/292 genes in our ASD geneset already have existing genetic mouse models, as well as rescue models for 28/292 genes according to the SFARI database (July 2020) at the time of writing. They are helpful in understanding function of these genes, and may also assist drug development to remediate related pathways, not just in the brain, but also throughout the peripheral nervous system that connects to co-morbidity organs, and even in the peripheral organs themselves.

The interconnectivity analyses from the current study reveal calcium, MAPK and glutamatergic signalling as three highest interconnected pathways, all are also involved with each other based on the interaction matrix. This is in line with a previous publication that ASD factors are converged upon MAPK and calcium signalling [30]. It is worth to note that MAPK signalling is also interlinked with calcium signalling in this study. Ten of the 14 MAPK pathway members, CACNA1A, CACNA1B, CACNA1C, CACNA1D, CACNA1G, CACNA2D1, CACNA2D3, CACNB2, ERBB4 and PRKCB, are overlapped with the calcium signalling (Table 5), and 8 of them are calcium channels. Furthermore, calcium channels appear in 16 top KEGG pathways including glutamatergic, GABAergic, dopaminergic, cholinergic and serotonergic synapses of the ASD genes (Table 5). Our results add to the evidence that calcium and glutamatergic signalling are the significant components in ASD pathways.

Calcium signalling is a highly integral system in the human body and is increasingly shown to be implicated in ASD [104, 105]. In neurons, the arrival of the electric current induces Ca\(^{2+}\) influx via voltage-gated calcium channels, and this triggers exocytosis and neurotransmitter release [106]. The voltage-gated calcium channels are tetramers containing three auxiliary subunits (β, α2δ, γ) and one pore-forming α1 subunit. Eight calcium channels (CACNA1A, CACNA1B, CACNA1C, CACNA1D, CACNA1G, CACNA2D1, CACNA2D3, CACNB2) are present in our 291 ASD geneset. Experiments using fibroblasts from monogenic [107] and non-syndromic autistic subjects [108] demonstrate aberrant calcium signalling mediated by I3PR. In
Table 5. Converging ASD candidate genes on E/I balance and calcium signalling pathway.

<table>
<thead>
<tr>
<th>Term ID</th>
<th>Term description (Background Gene Count)</th>
<th>ASD genes</th>
<th>FDR</th>
<th>Matching proteins in the network (IDs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa04724</td>
<td>Glutamatergic synapse (112)</td>
<td>12 (CNS)</td>
<td>2.46E-11</td>
<td>CACNA1A, CACNA1D, DLGAP1, GRIA1, GRIK2, GRIN1, GRIN2A, GRIN2B, GRM1, HOMER1, SHANK2, SLC1A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 (CNS +PT)</td>
<td>0.0179</td>
<td>CACNA1C, DLG4, GRIK5, ITPR1, PLCB1, PRKCB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19 (Com)</td>
<td>1.21E-10</td>
<td>CACNA1A, CACNA1C, CACNA1D, DLG4, DLGAP1, GRIA1, GRIK2, GRIK5, GRIN1, GRIN2A, GRIN2B, GRM1, HOMER1, SHANK3, ITPR1, PLCB1, PRKCB, SHANK2, SLC1A2</td>
</tr>
<tr>
<td>hsa04727</td>
<td>GABAergic synapse (88)</td>
<td>11 (CNS)</td>
<td>3.44E-11</td>
<td>CACNA1A, CACNA1B, CACNA1D, GABBR2, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3, GAD1, SLC12A5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 (Com)</td>
<td>2.48E-09</td>
<td>CACNA1A, CACNA1B, CACNA1C, CACNA1D, GABBR2, GABRA1, GABRA3, GABRA4, GABRA5, GABRB3, GAD1, GPHN, PRKCB, SLC12A5, SLC6A1</td>
</tr>
<tr>
<td>hsa04728</td>
<td>Dopaminergic synapse (128)</td>
<td>8 (CNS)</td>
<td>2.16E-06</td>
<td>CACNA1A, CACNA1B, CACNA1D, CAMK2A, GRIA1, GRIN2A, GRIN2B, SCN1A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 (CNS +PT)</td>
<td>0.0126</td>
<td>CAMK2C, CAMK2B, GSK3B, ITPR1, KIF5C, PLCB1, PRKCB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15 (Com)</td>
<td>7.84E-08</td>
<td>CACNA1A, CACNA1B, CACNA1C, CACNA1D, CAMK2A, CAMK2B, GRIA1, GRIN2A, GRIN2B, GSK3B, ITPR1, KIF5C, PLCB1, PRKCB, SCN1A</td>
</tr>
<tr>
<td>hsa04725</td>
<td>Cholinergic synapse (111)</td>
<td>6 (CNS)</td>
<td>0.0001</td>
<td>CACNA1A, CACNA1B, CACNA1D, CAMK2A, KCNQ2, KCNQ3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11 (Com)</td>
<td>2.36E-05</td>
<td>CACNA1A, CACNA1B, CACNA1C, CACNA1D, CAMK2A, CAMK2B, ITPR1, KCNQ2, KCNQ3, PLCB1, PRKCB</td>
</tr>
<tr>
<td>hsa04726</td>
<td>Serotonergic synapse (112)</td>
<td>5 (CNS)</td>
<td>0.0011</td>
<td>CACNA1A, CACNA1B, CACNA1D, GABBR3, KCN2D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 (Com)</td>
<td>0.00062</td>
<td>CACNA1A, CACNA1B, CACNA1C, CACNA1D, GABBR3, ITPR1, KCN2D, PLCB1, PRKCB</td>
</tr>
<tr>
<td>hsa04720</td>
<td>Long-term potentiation (64)</td>
<td>6 (CNS)</td>
<td>6.89E-06</td>
<td>CAMK2A, GRIA1, GRIN1, GRIN2A, GRIN2B, GRM1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6 (CNS +PT)</td>
<td>0.0025</td>
<td>CACNA1C, CAMK2B, CREBBP, ITPR1, PLCB1, PRKCB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12 (Com)</td>
<td>3.14E-08</td>
<td>CACNA1C, CAMK2A, CAMK2B, CREBBP, GRIA1, GRIN1, GRIN2A, GRIN2B, GRM1, ITPR1, PLCB1, PRKCB</td>
</tr>
<tr>
<td>hsa04730</td>
<td>Long-term depression (60)</td>
<td>4 (CNS)</td>
<td>0.0011</td>
<td>CACNA1A, GRIA1, GRM1, NOS1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 (Com)</td>
<td>0.00051</td>
<td>CACNA1A, GRIA1, GRM1, ITPR1, NOS1, PLCB1, PRKCB</td>
</tr>
<tr>
<td>hsa04024</td>
<td>cAMP signaling pathway (195)</td>
<td>8 (CNS)</td>
<td>3.21E-05</td>
<td>ATPT2B, CACNA1D, CAMK2A, GABBR2, GRIA1, GRIN1, GRIN2A, GRIN2B</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7 (CNS +PT)</td>
<td>0.0382</td>
<td>ATPT1A3, CACNA1C, CAMK2B, CREBBP, PDE4A, PDE4B, RAPGEF4</td>
</tr>
<tr>
<td>hsa04010</td>
<td>MAPK signaling pathway (293)</td>
<td>6 (CNS)</td>
<td>0.0105</td>
<td>CACNA1A, CACNA1B, CACNA1D, CACNA1G, CACNA2D3, ERBB4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14 (Com)</td>
<td>0.0013</td>
<td>CACNA1A, CACNA1B, CACNA1C, CACNA1D, CACNA1G, CACNA2D1, CACNA2D3, CACNB2, ERBB4, KIT, MAP2K8P2, NFX1, PRKCB, TAOK2</td>
</tr>
<tr>
<td>hsa04020</td>
<td>Calcium signaling pathway (179)</td>
<td>11 (CNS)</td>
<td>1.80E-08</td>
<td>ATPT2B, CACNA1A, CACNA1B, CACNA1D, CACNA1G, CAMK2A, ERBB4, GRIN1, GRIN2A, GRM1, NOS1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16 (Com)</td>
<td>7.79E-07</td>
<td>ATPT2B, CACNA1A, CACNA1B, CACNA1C, CACNA1D, CACNA1G, CAMK2A, CAMK2B, ERBB4, GRIN1, GRIN2A, GRM1, ITPR1, NOS1, PLCB1, PRKCB</td>
</tr>
<tr>
<td>hsa04713</td>
<td>Circadian entrainment (93)</td>
<td>9 (CNS)</td>
<td>1.74E-08</td>
<td>CACNA1D, CACNA1G, CAMK2A, GRIA1, GRIN1, GRIN2A, GRIN2B, NOS1, NOS1AP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 (CNS +PT)</td>
<td>0.0339</td>
<td>CACNA1C, CAMK2B, ITPR1, PLCB1, PRKCB</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14 (Com)</td>
<td>1.98E-08</td>
<td>CACNA1C, CACNA1D, CACNA1G, CAMK2A, CAMK2B, GRIA1, GRIN1, GRIN2A, GRIN2B, ITPR1, NOS1, NOS1AP, PLCB1, PRKCB</td>
</tr>
</tbody>
</table>

(Continued)
addition, mutations of calcium channels have been found in ASD, for example, \textit{CACNA1A} rs7249246/rs12609735 were associated with Chinese Han ASD [109], and \textit{CACNA1A} mutations in Epileptic Encephalopathy [110, 111], gain of function of \textit{CACNA1C} in Timothy syndrome with ASD [112] and recurring CNVs of \textit{CACNA2D3} [29, 36, 113, 114]. Exome sequencing has identified various mutations of \textit{CACNA1D} in ASD [14, 16, 29, 36, 115], epilepsy [116] developmental delay [117] endocrine issues [117, 118], \textit{CACNA2D1} in epilepsy and intellectual disability [119] and \textit{CACNB2} mutations in ASD [120, 121]. \textit{CACNA1C} (rs1024582) and \textit{CACNB2} (rs2799573) polymorphisms were suggested as the common risks across seven brain diseases [122]. Therefore, dysregulated calcium and synaptic signalling could be a commonly perturbed pathway in ASD and frequently occurring comorbidities.

Calcium channels are coupled to neuronal transmission, and E/I imbalance was proposed as a common ASD pathway previously [123]. For example, increased calcium signalling is found in \textit{NRXN1α+/-} neurons derived from ASD induced pluripotent stem cells with increased expression of voltage-gated calcium channels [124]. In the current study, KEGG pathways show that \textit{CACNA1A}, \textit{CACNA1C}, \textit{CACNA1D} are involved in glutamatergic synapse; \textit{CACNA1A}, \textit{CACNA1B} and \textit{CACNA1D} in cholinergic synapse; and \textit{CACNA1A}, \textit{CACNA1B}, \textit{CACNA1C} and \textit{CACNA1D} in GABAergic, dopaminergic, and serotonergic synapses. The glutamatergic and GABAergic transmission are the major excitatory and inhibitory pathways in the CNS, which are recurrently featured in “Biological Processes”, “Molecular Functions”, “Cell Components” and KEGG pathways of the 291 ASD candidate genes. Various glutamatergic receptor genes (\textit{GRIA1}, \textit{GRIK2}, \textit{GRIK5}, \textit{GRIN1}, \textit{GRIN2A}, \textit{GRIN2B}, \textit{GRM1}, \textit{GRM5}, \textit{GRM7}), scaffolding components of synapses (\textit{SHANK2, SHANK3, DLG4, DLGAP1}) and transport of glutamate (\textit{SLC1A2}) were all involved in ASD and other comorbidities [110, 125–130].
Remarkably, various calcium signalling members are commonly appeared in other pathways that are perturbed in ASD (Fig 5C–5D, Table 5). For example, 11/14 molecules in Circadian entrainment, 10/14 in MAPK signalling, 7/15 in Wnt signalling, 10/10 in Oxytocin signalling, 9/12 in Aldosterone synthesis and secretion, 8/17 in Retrograde endocannabinoid signalling, 6/11 in insulin secretion, were found in calcium signalling pathway, which could be of great interest given their role in the pancreas [131]. In the top KEGG pathways identified from the CNS geneset (CNS), the CNS+PT geneset and combined (Com) 292 ASD candidates, calcium signalling members (bold) appeared in all other KEGG pathways. Therefore, calcium signalling is likely to play a major role not only in ASD but also in ASD comorbidities.

**Limitations**

Like other meta-analyses, biases and limitations should be considered in pathway analyses. We first assumed that any significant ASD candidate genes shall appear in 4/5 independent datasets. However, some strong candidate genes, such as *FMR1* and *PTEN*, did not appear in the final 291 gene list. While they have much evidence to support their role in ASD, the genes did not overlap due to differences in ranking criteria of different systems. For example, *PTEN* was in the top ranking in SFARI, Zhang and EXAC, however it was not shortlisted from Duda’s and Krishnan’s score systems. Likewise, *FMR1* was ranked top in Duda’s, Zhang’s and SFARI, but did not pass the threshold in EXAC or Krishnan’s dataset. The biological influence of *FMR1* and *PTEN* in ASD is very significant. For example, *FMR1* is known to target 126/291 ASD candidates in the current study, and *PTEN* is a target of another strong ASD candidate, *CHD8*, on our list.

We also filtered out 200 genes (S26 Table), which were overlapped by four independent scoring systems but not existed in the SFARI database. 66 of them are targets of the *FMR1*, and 36 are targets of *CHD8*. Some of them like *BCL11B, DPF1, ETV1, NFASC, PAK7, PLXNC1, SCN3A, SLC17A7* and SLITRK1 were also dysregulated in cerebral organoids derived from autistics with large head circumference [50], while others (*NEED4L, RICTOR, SLC8A1, CELF2, MLLT3, PPFIA2, EPHA7, LRRC4C* and *RORB*) were identified from very recent single cell sequencing of autism cortical tissue [68]. This suggests they would still be modulated as downstream targets in some cases of ASD.

In the current study we hypothesized that a strong ASD candidate gene should be highly expressed in the brain and/or relative PT with strong comorbidity. However, this assumption could be potentially challenged by the following scenario. (1) If an ASD gene were development-specific but had low abundance in adult post-mortem tissues, it would be filtered out by GTEx dataset derived from donors of 20–79 age; (2) If a gene were highly expressed in a small subset of specific nuclei of the brain, which might appear not abundant in the total brain RNA; (3) If an ASD gene were modulated by ethnic genetic background, as 85.2% of GTEx donors were Caucasian European subjects, 12.7% were African-American, 1.1% were Asian and 0.3% were American-Indian. As for protein expression, the HPA is an evolving database, and genes with no protein expression data now can be updated in future releases. The same goes for cell types in tissue, which will also be incomplete, given the myriad cell types that are present in all organs. The use of single cell sequencing technology and flow cytometry will be useful in addressing the issue [132, 133].

It is interesting that many of the ASD genes are found to have sex-bias expression and/or splicing, which may be associated with sex-bias diagnosis of ASD [134–136]. We are limited in understanding its biological base by the lack of transcriptomic analysis of ASD genes. Many of the transcriptomic studies have focused on male subjects, or the ratio of female samples are too few to illustrate any sex-specific effects in ASD. It is suggested that future transcriptomic
studies should incorporate an even number of male and female subjects in both groups of case and control. Organoid cell lines may also be created from both sexes to make up for this short coming.

It would be desirable if gene expression datasets are available to compare from control and ASD patients at different time points not just in the brain, but also across the entire body. A recent publication [137] has proposed a paediatric cell atlas, which would collate and characterise gene expression at the single cell level across multiple tissues and time points of human development from birth to adulthood. This would be a fantastic initiative if it fully goes ahead, as such a resource would bring great insight into the co-morbidities associated with ASD. Single-cell expression data from cortical tissue of ASD subjects has become available to researchers recently [68], which could be a good starting point to analyse cell-types of interest and explore ASD heterogeneity. The availability of iPSCs from different ASD cases allows to culture and analyse different cell types in both CNS and peripheral organs, which are not easily accessed by conventional methodologies. This can be useful to explore how ASD genes may influence the biological processes during brain development, neuronal function, as well as cells of comorbidity peripheral organs.

Conclusion

By utilizing multiple scoring systems, we have identified recurrent ASD candidate genes, with convergence on multiple pathways and processes involved in ASD (Fig 10). The use of GTEx and HPA data also gives a glimpse into their body-wide expression patterns, which has not been explored previously using ASD gene lists, which we have done so in this study. The bioinformatic analyses of CNS-specific and/or CNS+PT candidate genesets enable us to pinpoint CNS development, E/I balance and calcium signalling as important pathways involved in not just ASD but also brain comorbidity such as epilepsy. The analysis of CNS+PT geneset suggests chromosomal organisation/transcription regulation, calcium-interconnected MARK/ WNT and secretion as major pathways with disruptive behaviour, developmental delay as well
as congenital abnormalities. Calcium signalling is highly interconnected amongst pathways, which could be informative in exploring complications and co-morbidities associated with ASD where calcium signalling could be involved, especially those subsets of autistic individuals who harbour mutations in these genes that can result in channelopathies.

Supporting information

S1 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S2 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S3 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (PNG)

S4 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S5 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S6 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney,
Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S7 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S8 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S9 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S10 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S11 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

**S12 Fig. ASD genes are present in tissue-specific interaction networks.** Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.
S13 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S14 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S15 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S16 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S17 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S18 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach. (TIF)

S19 Fig. ASD genes are present in tissue-specific interaction networks. Yellow nodes highlight ASD genes, green are tissue-specific genes, red edges are interactions between ASD genes and other partners. The graphs are in order of appearance; Adrenal Gland, Adipose subcutaneous, Artery aorta, Artery coronary, Artery tibial, Colon sigmoid, Colon transverse, Esophagus-
mucosa, Esophagus-muscularis, Heart-left atrial appendage, Heart-left ventricle, Kidney, Lung, Muscle-skeletal, Pancreas, Pituitary, Small Intestine-terminal ileum, Spleen, Stomach.

S1 Table. The shortlisted genes from the scoring systems, which can be found in the supporting excel file.
(XLSX)

S2 Table. The shortlisted genes from the scoring systems, which can be found in the supporting excel file.
(XLSX)

S3 Table. The shortlisted genes from the scoring systems, which can be found in the supporting excel file.
(XLSX)

S4 Table. The shortlisted genes from the scoring systems, which can be found in the supporting excel file.
(XLSX)

S5 Table. The shortlisted genes from the scoring systems, which can be found in the supporting excel file.
(XLSX)

S6 Table. Jvenn overlap data. This can be found in the supporting excel file.
(XLSX)

S7 Table. HPA data relating to ASD genes. This can be found in supporting excel file.
(XLSX)

S8 Table. The 292 ASD gene list with attached scoring information: This can be found in the supporting excel file.
(XLSX)

S9 Table. The expression data of the 292 genes, this is found in the supporting excel file.
(XLSX)

S10 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S11 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S12 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S13 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S14 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)
S15 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S16 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S17 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S18 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S19 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S20 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S21 Table. STRING results for the CNS, CNS+PT and all 292 genes, these are found in the supporting excel files.
(XLSX)

S22 Table. Pathway interaction matrix, this is found in the supporting excel file.
(XLSX)

S23 Table. Dysregulated genes overlapped with our dataset with up/down regulation and fold changes. This can be found in supporting excel file.
(XLSX)

S24 Table. Enrichment Results from CSEA analysis of genes in different brain regions across multiple timepoints, this is found in the supporting excel file.
(XLSX)

S25 Table. Gene lists for sex-biased expression in male and female prenatal brain, sex-biased splicing, and psychiatric conditions from Psygenet, and peripheral conditions from Harmonizome database. These are found in the supporting excel file.
(XLSX)

S26 Table. Table of non-SFARI Overlapping genes. This is found in the supporting excel file.
(XLSX)

S1 File. The web links to full WebGestalt results are attached as html documents.
(ZIP)

(R)

S3 File. HPA-code final.
(R)
Acknowledgments
The Authors wish to thank all those who contributed to this manuscript.

Author Contributions
Conceptualization: Jamie Reilly.

Formal analysis: Jamie Reilly.

Supervision: Sanbing Shen.

Visualization: Jamie Reilly, Sanbing Shen.

Writing – original draft: Jamie Reilly.

Writing – review & editing: Louise Gallagher, Geraldine Leader, Sanbing Shen.

References


41. Wickham H (2016) ggplot2 - Elegant Graphics for Data Analysis. 2nd ed. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4


Coupling of autism genes to tissue-wide expression and dysfunction of synapse, signalling and gene regulation


Coupling of autism genes to tissue-wide expression and dysfunction of synapse, signalling and gene regulation


