

SCIENTIFIC REPORTS



OPEN

A gene expression signature in developing Purkinje cells predicts autism and intellectual disability co-morbidity status

Harry Clifford¹, Anna Dulneva¹, Chris P. Ponting¹, Wilfried Haerty^{1,2} & Esther B. E. Becker¹ 

Autism spectrum disorder (ASD) is a complex neurodevelopmental disease whose underpinning molecular mechanisms and neural substrates are subject to intense scrutiny. Interestingly, the cerebellum has emerged as one of the key brain regions affected in ASD. However, the genetic and molecular mechanisms that link the cerebellum to ASD, particularly during development, remain poorly understood. To gain insight into the genetic and molecular mechanisms that might link the cerebellum to ASD, we analysed the transcriptome dynamics of a developing cell population highly enriched for Purkinje cells of the mouse cerebellum across multiple timepoints. We identified a single cluster of genes whose expression is positively correlated with development and which is enriched for genes associated with ASD. This ASD-associated gene cluster was specific to developing Purkinje cells and not detected in the mouse neocortex during the same developmental period, in which we identified a distinct temporally regulated ASD gene module. Furthermore, the composition of ASD risk genes within the two distinct clusters was significantly different in their association with intellectual disability (ID), consistent with the existence of genetically and spatiotemporally distinct endophenotypes of ASD. Together, our findings define a specific cluster of ASD genes that is enriched in developing PCs and predicts co-morbidity status.

Autism spectrum disorder (ASD) is a highly prevalent, complex group of neurodevelopmental diseases defined by deficits in social cognition and communication as well as restricted interests and repetitive behaviours. Beyond these core features, ASD is often associated with variable co-morbid conditions including a low nonverbal intelligence quotient, motor deficits and epilepsy¹. ASD is highly heritable, and recent advances in genomic technology have led to the identification of several hundred genetic variants associated with ASD¹. This considerable genetic heterogeneity of ASD combined with its broad clinical phenotype present a major challenge to our understanding of the underlying disease pathophysiology. The primary molecular mechanisms and also the neural substrates that cause ASD remain largely to be elucidated.

Interestingly, the cerebellum has emerged as one of the key brain regions affected in autism^{2,3}. Imaging meta-analysis has revealed a significant reduction of distinct grey matter areas in the cerebellum in ASD⁴, whose degree predicts the severity of core autism symptoms⁵. In particular, Purkinje cells (PCs), which constitute the sole output neurons of the cerebellar cortex, are reduced in number and density in ASD^{1,6}. Moreover, a critical role for PCs in autism has been demonstrated in PC-specific conditional mouse models lacking the ASD-associated genes *Tsc1*, *Tsc2* and *Shank2*⁷⁻⁹. Together, these and other findings suggest that PC dysfunction during a critical developmental period may contribute to ASD^{3,10}. Remarkably, developmental injury to the human cerebellum constitutes the largest single non-heritable risk of developing ASD³. However, our understanding of the genetic and molecular mechanisms that underpin developmental PC dysfunction and might trigger ASD remains poor.

To gain a better understanding of the genes expressed during PC development and their potential association with ASD, we profiled the transcriptome of developing mouse PCs by deep sequencing across multiple developmental time points. We identified a single cluster of genes whose expression is positively regulated over PC

¹MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, OX1 3PT, United Kingdom. ²Present address: Earlham Institute, Norwich Research Park, Norwich, NR4 7UG, United Kingdom. Correspondence and requests for materials should be addressed to W.H. (email: wilfried.haerty@earlham.ac.uk) or E.B.E.B. (email: esther.becker@dpag.ox.ac.uk)

development and that is enriched for genes associated with ASD. This ASD-associated gene cluster was specific to developing PCs and was not associated with the mouse neocortex during the same developmental period, in which we identified a second temporally regulated ASD gene module. Strikingly, we recognized a significant difference in the composition of the ASD risk genes underlying the clusters in the two brain regions that relates to their association with intellectual disability (ID).

Results

Increasing evidence points to a role for PC dysfunction during a critical developmental period in the causation of ASD^{3,10}. The human cerebellum develops over a protracted period of time, ranging from 4 weeks of gestational age until 20 months of postnatal age^{11,12}. The third trimester of pregnancy is characterized by a highly dynamic period for cerebellar development¹³, during which time the cerebellum is extremely vulnerable to insults that are strongly associated with autism^{3,14}.

We set out to survey the intrinsic genetic program that drives PC development over this highly sensitive and critical period and to correlate this with the expression of ASD genes. To do so, we analyzed PC transcriptome dynamics in the developing mouse cerebellum during the first three postnatal weeks, a key developmental period that is equivalent to the third trimester in human cerebellar development¹⁵. PCs undergo extensive developmental changes during this period, including elaborate dendritic outgrowth and synaptogenesis (Fig. 1A)^{16,17}.

PCs constitute considerably less than 1% of the total cerebellar cell population¹⁸. To identify PC-specific transcriptome dynamics that might otherwise be masked using a whole tissue approach, we employed laser capture microdissection to isolate individual PCs. Deep sequencing was performed on RNA isolated from 1000 PCs at five developmental time points (postnatal days P0, P4, P8, P14 and P21) in triplicate. Importantly, time points were separated in a principal component analysis (PCA) (Fig. 1B), demonstrating that PCs were effectively captured at independent developmental stages.

We next confirmed that PCs were isolated successfully and contributed the majority of the captured RNA. To do so, we compared the transcript abundance (RPKM values) of genes considered to be markers of PCs or other cerebellar cell populations, including cerebellar granule neurons, Basket and Stellate cells, Golgi cells and glia. The captured cells showed very high enrichment for PCs, with the PC-specific markers *Calb1* and *Itpr1* displaying strong expression that increased over postnatal development (Fig. 1C; Supplementary Fig. S1B). Markers for other cerebellar cell types demonstrated little-to-no expression. We nevertheless identified minor expression of a glial cell marker (*Gdf10*), suggesting a potential contamination by this cell type, likely due to the intimate physical association of Bergmann glia with PCs¹⁹. Together, these results demonstrate the successful isolation of a highly enriched population of PCs and the capture of gene expression in PCs across all time points.

We next investigated gene expression patterning across development using weighted gene co-expression network analysis (WGCNA)²⁰. We defined 17 gene clusters (Supplementary Fig. S2) and then investigated each of these modules for both functional terms and disease-associated gene enrichment. Only two clusters (the aquamarine and the blue module) yielded significant functional annotation enrichments (Fig. 2A). The first cluster (WGCNA_{neg}) contained 5,052 genes (680 with module membership >0.8) that were negatively correlated with PC maturation, and was significantly enriched for genes involved in RNA processing (Supplementary Fig. S3). The second cluster (WGCNA_{pos}) was composed of 4,226 genes (1,084 with module membership >0.8) whose expression positively correlates with PC maturation. This cluster was significantly enriched in terms associated with neuronal development. Specifically, WGCNA_{pos} was enriched in genes whose disruption leads to long-term potentiation phenotypes, and abnormal nervous system electrophysiology and PC morphology (Supplementary Fig. S3). These findings are consistent with an enrichment of genes critical for postnatal PC development. Interestingly, abnormal PC development and early PC dysfunction are emerging mechanisms underlying many cerebellar ataxias^{21,22}. Our data support this as we found a strong enrichment for the KEGG ataxia pathway in the WGCNA_{pos} gene cluster (Supplementary Fig. S3). Overall, these results confirm ontologies and pathways expected in PC development and suggest that we can exploit these data to explore the genetic correlates of ASD in developing PCs.

As the relatively low number of replicates per time point may affect the robustness of WGCNA clusters, we performed the same analyses using DESeq²³ on genes that were identified as being significantly differentially expressed across time and correlated with PC maturation. We identified 2,511 and 1,917 significantly differentially expressed genes across time that were negatively (DESeq 2_{neg}) and positively correlated (DESeq 2_{pos}) with PC maturation, respectively (Supplementary Fig. S4). The WGCNA and DESeq 2 approaches were largely concordant: 82.5% and 97.0% of the genes within the negatively and positively regulated WGCNA clusters (module membership >0.8) were also found within the two DESeq 2 clusters, and gene set enrichments were similar (Supplementary Figs S3 and S4). We will refer to the positively correlated gene clusters WGCNA_{pos} and DESeq 2_{pos} as “PC development clusters”.

Given the association of the cerebellum with various neurological diseases including ataxia but also cognitive disorders such as ASD²⁴, we next investigated whether genes associated with these disorders were enriched in any of the identified gene clusters. Of the 17 WGCNA clusters, only the WGCNA_{pos} cluster returned significant enrichments for human genes causing ataxia (4.3-fold, $q = 2.27 \times 10^{-5}$) as well as genes that when disrupted in mice cause ataxia-like phenotypes (8.8-fold, $q = 1.19 \times 10^{-7}$) (Fig. 3A, Supplementary Table S4). The difference in enrichment observed between human and mouse likely stems from the fact that many genes in mouse do not have a one-to-one orthologous gene in human leading to a reduction in power.

We next interrogated the gene clusters for an enrichment for ASD-associated genes from the Simons Foundation Autism Research Initiative (SFARI) database²⁵. The SFARI human database contains more than 800 human genes associated with ASD, while the mouse ASD-associated gene database contains a subset of these (229), for which genetic models have been generated. Interestingly, the WGCNA_{pos} cluster was the only cluster with a significant enrichment, for which we observed a 2.4-fold enrichment ($q = 1.10 \times 10^{-2}$) for mouse

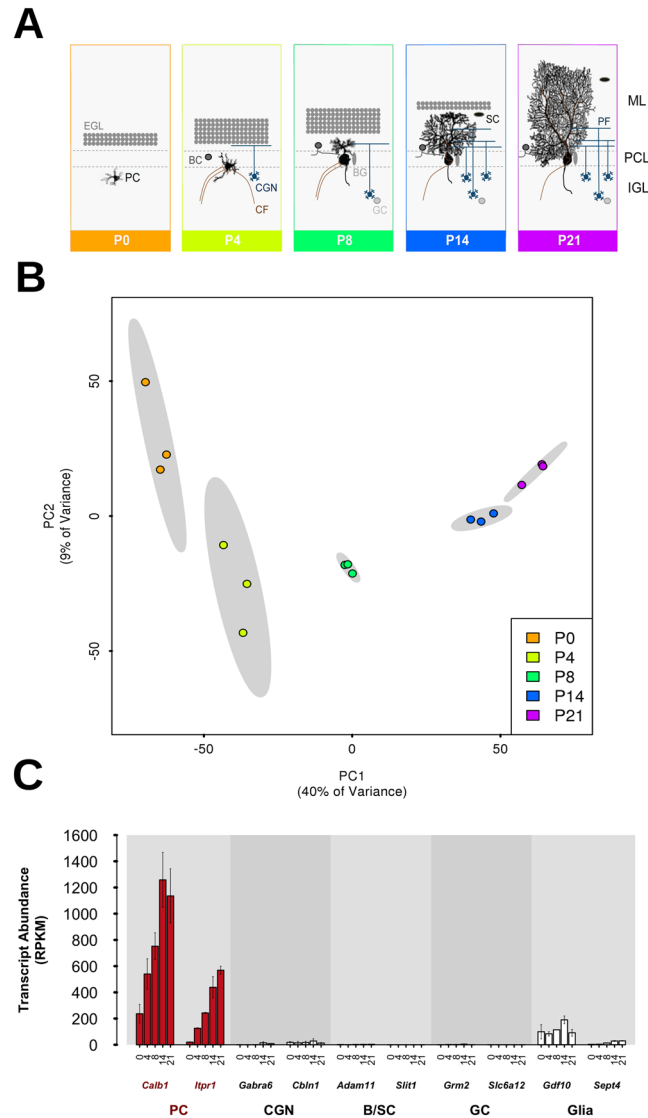


Figure 1. Gene expression data cluster according to developmental time point and confirm successful Purkinje cell (PC) capture. **(A)** Schematic diagram of mouse PC development from postnatal day 0 (P0) until P21. This developmental period is characterized by extensive morphological changes and the formation of synaptic connections with climbing fibers (CF), Basket cells (BC) and parallel fibers (PF) of cerebellar granule neurons (CGN). BG: Bergmann glia; EGL: external granular layer; GC: Golgi cells; IGL: internal granular layer; ML: molecular layer; PCL: Purkinje cell layer; SC: Stellate cells. **(B)** PCs were effectively captured at independent developmental stages. Scatter plot of the first two principal components (PC1/2) for all samples, displaying significant groupings according to postnatal day. 95% confidence ellipses for these groups are shaded in grey, displaying full separation. **(C)** Predominant capture of PC-specific transcripts. Gene expression data (RPKM values) across developmental time points (X-axis) for two cell-specific markers for each cell-type. CGN: cerebellar granule neurons; B/SC: basket/stellate cells; GC: Golgi cells; BG: Bergmann glia. See also Supplementary Fig. S1.

ASD-associated genes (Fig. 3A; Supplementary Table S5). Similar conclusions were reached for the cluster identified with DESeq 2 (Fig. 3B; Supplementary Table S5). Importantly, we also found a significant enrichment of orthologues for human genes associated with ASD within the DESeq 2_{pos} cluster (1.6-fold enriched, $q = 6.34 \times 10^{-3}$). We also used a recently published set of *de novo* variants associated with autism²⁶ (Fig. 3B; Supplementary Table S5). We found no significant enrichment for these variants after multiple test correction in the clusters identified with WGCNA and with DESeq 2 (Fig. 3).

The difference between the WGCNA and DESeq 2 clusters likely stems from the greater number of genes within the latter, resulting in its greater analytical power. There was no significant enrichment for genes associated with schizophrenia in either cluster, suggesting that the observed enrichment is specific to ASD but not another related neurological disorder, namely schizophrenia (Fig. 3).

In addition to the ASD-associated genes from the SFARI database, we also investigated 842 genes whose transcripts were previously identified as targets of the Fragile X mental retardation protein (FMRP)²⁷ and significantly

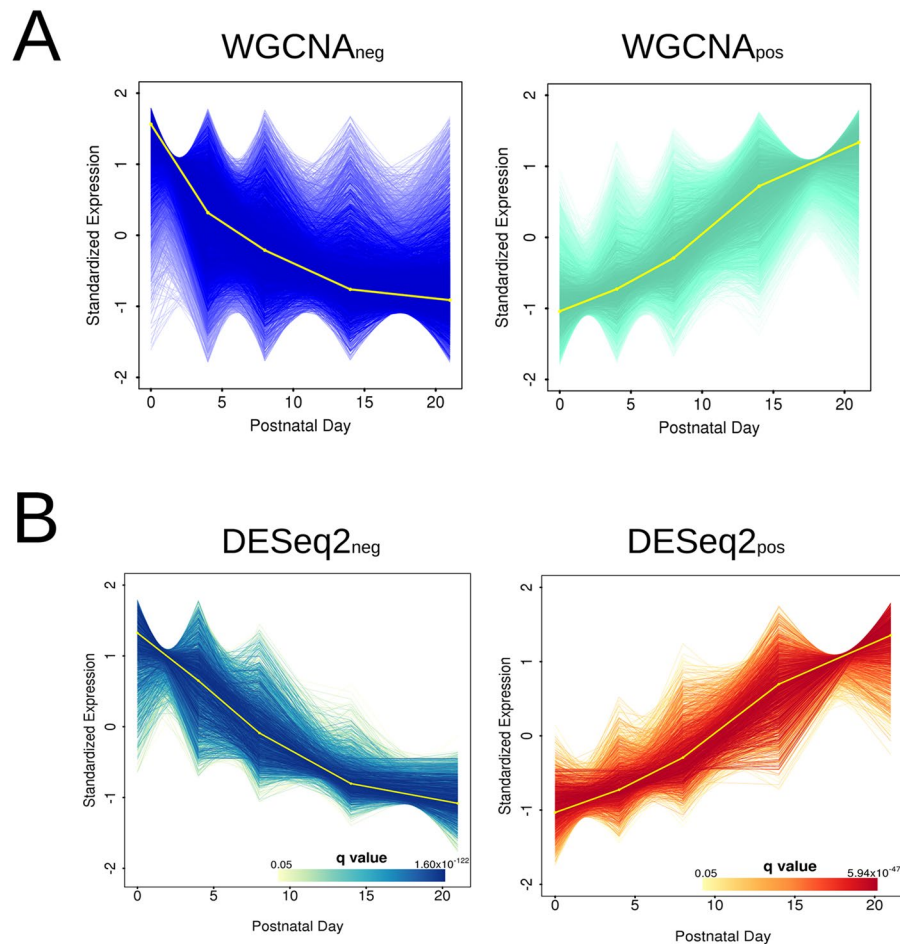


Figure 2. Co-expression network analysis reveals temporal gene clusters that are correlated with PC development. **(A)** Genes assigned to the two WGCNA modules with statistically significant functional annotations ($q < 0.05$), expressing patterns of negative ($WGCNA_{neg}$) and positive ($WGCNA_{pos}$) fold-change over time, respectively. $WGCNA_{neg}$ is significantly enriched for ontology terms associated with RNA processing, while $WGCNA_{pos}$ is significantly enriched for ontology terms associated with neuron formation and neurological disorders. See also Supplementary Figs S2 and S3. Expression values are normalized and standardized to a mean of zero and standard deviation of 1. The eigengene of the cluster is displayed in yellow. **(B)** Statistically significant differentially expressed genes over time, separated in clusters of negative ($DESeq2_{neg}$) and positive ($DESeq2_{pos}$) fold-change, respectively. The DESeq2 clusters show strong intersection with those identified by WGCNA and provide robustness for downstream analyses. Both WGCNA and DESeq2 clusters contained similarly enriched ontology terms. See also Supplementary Figure S4 and Supplementary Table S3. Each gene within the module is plotted as a line of increasing colour intensity with decreasing q value, as displayed in the colour scale. Expression values are normalized and standardized to a mean of zero and standard deviation of 1. The eigengene of the cluster is displayed in yellow.

enriched for ASD-associated genes^{28,29}. FMRP targets were only enriched in the $WGCNA_{pos}$ cluster (2.7-fold enrichment, $q = 7.34 \times 10^{-21}$) and both $DESeq2_{pos}$ and $DESeq2_{neg}$ clusters (2.3-fold and 1.4-fold enrichments; $q = 2.78 \times 10^{-30}$ and 1.02×10^{-6}). These findings provide evidence that FMRP targets are enriched in developmentally regulated PC genes.

We next investigated the specificity of these results with regards to PC development by undertaking an equivalent analysis on published transcriptomes from the developing mouse neocortex at the same postnatal time period³⁰. For this analysis we limited our investigations to the sequencing libraries from layer 4, as it was the only layer fully isolated from other cortical layers³⁰. Twelve WGCNA gene clusters were defined from these data (Supplementary Fig. S5), of which only one (the darkgoldenrod module) was significantly enriched for both mouse and human SFARI ASD genes (categories 1-4 and syndromic) (3.0-fold and 2.9-fold enrichments; $q = 0.018$ and 3.61×10^{-4} , for the mouse and human gene sets, respectively; Supplementary Fig. S6). The enrichment analyses performed on category 1 or 2 genes did not yield significant results likely due to loss of statistical power associated with the low number of genes involved. In contrast to the gene clusters identified in developing PCs, this ASD-enriched neocortex cluster does not show positive differential expression over development (Fig. 4A). When using the variants identified by Krumm *et al.*²⁶ we observed a low p-value ($p = 0.029$), but the enrichment was no longer significant after multiple test correction ($q > 0.05$). We identified a significant

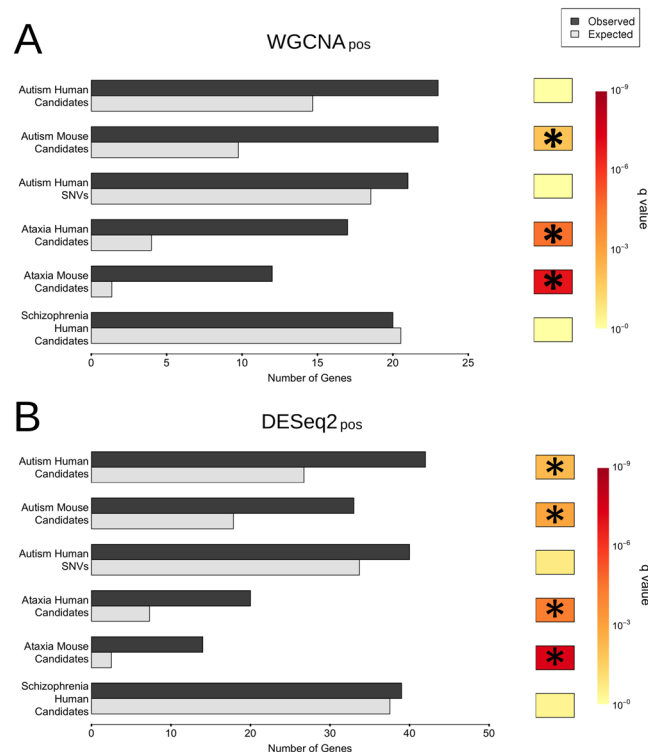


Figure 3. The developmental PC cluster positively correlated with timing contains unexpectedly high numbers of ataxia- or autism-associated genes. **(A)** Observed and randomly expected numbers of genes present in the WGCNA_{pos} PC cluster (module membership >0.8), for genes and SNVs associated with ataxia, autism, and schizophrenia (including both mouse model and human candidate lists for the former two disorders). *q*-Values are indicated using a Hinton plot (right), with significant results ($q < 0.05$) marked by an asterisk. See also Supplementary Table S4. **(B)** Analogous testing within the corresponding DESeq2_{pos} cluster. This returned similar results, with significant enrichments for both autism gene lists (human and mouse, respectively). *q*-Values are indicated using a Hinton plot (right), with significant results ($q < 0.05$) marked by an asterisk.

enrichment of FMRP target genes in two clusters. These included the ASD-enriched neocortex cluster (3.7-fold enrichment, $q = 3.12 \times 10^{-8}$) and a cluster positively correlated with developmental timing (1.85-fold enrichment, $q = 3.96 \times 10^{-19}$). Interestingly, we found that ASD-associated genes in the PC and neocortex clusters showed distinct spatiotemporal expression levels. Genes in the neocortical gene cluster exhibited low coherency and preservation over PC development, and *vice versa* (Fig. 4B). These results indicate that the two clusters are distinct and not maintained across development in different brain regions.

To test the degree of overlap between PC and neocortex cluster genes, we performed an intersection analysis of the ASD-associated genes within each identified cluster. Of 182 genes that occur in either the human or mouse ASD gene lists, 39 were present in the neocortex cluster and 59 occurred in the PC cluster (Fig. 4C). We found only a single gene (*Maosa*) that was present in both sets, which represents a statistically significant depletion ($p = 5.8 \times 10^{-4}$; Fisher's exact test). The unexpectedly low overlap suggests that the observed enrichment of ASD genes in developing PCs is specific to this neuronal subpopulation at the investigated developmental time period.

The identification of the two spatiotemporally distinct ASD gene clusters in PCs and neocortex, respectively (Fig. 4B,C), allowed us to ask the intriguing question of whether genes within these cell type-specific modules might contribute to distinct ASD endophenotypes³¹. A stratification of the broad autism phenotype into clinically, genetically and biologically meaningful categories has thus far been lacking. One of the ASD endophenotypes that has been proposed to define clinical subgroups is intelligence quotient^{32–34}. We therefore considered whether the two different gene clusters reflect differences in intellectual disability (ID) among ASD individuals. To do so, using the SFARI annotations, we split the human candidate ASD gene list into those associated with ID and those with no evidence for ID, and tested for their enrichment within either the PC or neocortex modules. Comparison of the odds ratios from these four contingency tables revealed that the PC gene cluster contains a significantly greater proportion of ID-free ASD-associated genes than the neocortex gene cluster. In contrast, this pattern is inverted in the neocortex cluster, with it containing a greater proportion of ID-associated ASD genes ($p = 4.2 \times 10^{-2}$; permutation test; Fig. 4D). By repeating this analysis on the syndromic SFARI genes separately, this inversion of patterns was found to be significantly pronounced ($q = 6.8 \times 10^{-4}$; permutation test; Fig. 4D) and the trend was maintained, albeit at a non-significant level, for categories 1–4 ($q = 2.4 \times 10^{-1}$; permutation test; data not shown). Of the 30 syndromic genes associated with ID, 8 were identified in the neocortex cluster, and only one in the PC cluster. Conversely, of the 11 syndromic genes with no reported association with ID,

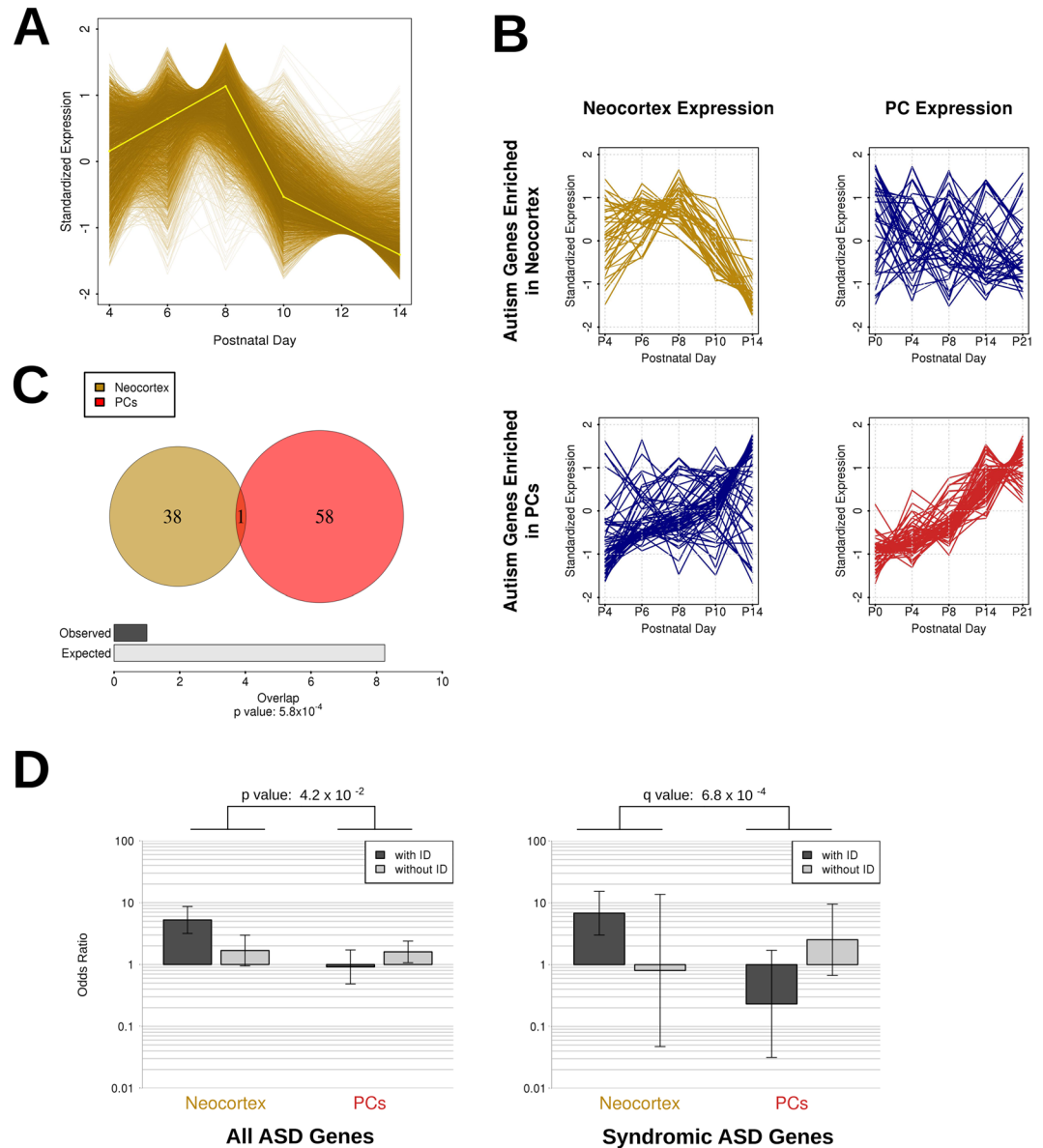


Figure 4. Comparisons between PC ASD (DESeq 2_{pos}) and neocortex ASD gene clusters reveal differential ID co-morbidity. **(A)** The neocortex cluster exhibits statistically significant enrichment for autism genes (for either mouse models or human candidates – see also Supplementary Table S5). Each gene within the module is plotted as a line whose colour intensity is determined by module membership. Expression values are normalized and standardized to a mean of zero and standard deviation of 1. The eigengene is displayed in yellow. **(B)** The expression pattern of the autism-associated genes over development in either the neocortex or PCs shows a lack of coherency of each of the two modules for expression in the other tissue. Each gene within the module is plotted as a line, with expression values normalized and standardized to a mean of zero and standard deviation of 1. **(C)** Significant lack of overlap between the 182 autism-associated genes within the neocortex and PC clusters (Fisher’s exact test). **(D)** Intellectual disability (ID)-associated autism genes are significantly differently partitioned between the PC and neocortex gene clusters. Two sets of odds ratios from four contingency tables, including ID-associated and ID-free autism gene lists and their occurrences in both the neocortex cluster and the PC cluster. The two panels represent testing on autism gene lists for all ASD genes (syndromic genes and category 1–4 genes) and syndromic genes only. The p- and q-values indicate the estimated probability of the difference between the ratios of odds ratios for the two clusters.

none were found in the neocortex cluster and 3 were found in the PC cluster. These results indicate that there is a significant difference in the composition of the ASD genes underlying the clusters in the two brain regions with relation to their association with ID. These differences suggest the existence of genetically and spatiotemporally distinct endophenotypes of ASD.

Discussion

The analysis of gene expression within disease-relevant tissues and cell types is a powerful approach to identify genes and regulatory networks whose disruption is associated with disease. Given the wide heterogeneity of the ASD phenotype, it is expected that multiple brain regions, cell types and critical time periods contribute to the disease. Previous analyses indicated that ASD-associated genes expressed in the prefrontal cortex show a significant association with developmental patterning^{29,35}. However, little attention has been paid to the ASD gene expression profiles in the cerebellum, a key brain region in ASD. Here, we report gene expression changes across key postnatal developmental stages within a population of laser-captured cerebellar cells that are highly enriched for Purkinje cells. Interestingly, we have identified a cluster of ASD genes that is enriched among all genes whose expression significantly increases over PC development. Importantly, the ASD genes enriched in the developing PC cluster are associated with a distinct disease endophenotype, namely the absence of ID co-morbidity. This is in contrast with a single, independent gene cluster in the mouse neocortex at the same developmental time point that is enriched for ASD genes with ID association. To our knowledge, these findings are the first to indicate a relationship between the spatiotemporal expression pattern of ASD genes and genetically distinct ASD endophenotypes.

A potential limitation of our study might stem from potential differences in developmental processes and developmental timing in the neocortex and the cerebellum owing to, for instance, neurogenesis occurring earlier in the neocortex. Such differences might partly explain the difference in the correlation between gene expression and development for the ASD gene-enriched neocortex and PC cluster, respectively. Furthermore, differential developmental gene expression might explain why some of the identified transcripts appear only in the PC cluster or neocortex cluster, despite their known widespread expression patterns.

Our identification of an ASD-gene enriched expression cluster that is specific to developing PCs provides support for the emerging concept that this neuron population is highly vulnerable in ASD^{1–3,10}. Moreover, our findings inform future functional studies of the identified specific ASD candidate genes that should be carried out in developing PCs to obtain a deeper understanding of the pathophysiological mechanisms underlying ASD. It is important to note that we found some evidence of glial markers in our samples and thus cannot rule out the presence of some contaminating glia cells, most likely Bergmann glia that are present in the PC layer. However, given the very high abundance of PC markers compared to glial markers, the effect of Bergmann glia should be marginal and does not affect our conclusions that the developing cerebellum has an ASD signature. Other transcriptomic studies have identified cerebellar gene networks enriched in ASD genes that show partial overlap with our data. A correlation analysis of ASD genes with images from the Allen Mouse Brain Atlas³⁶ found two co-expression modules that were significantly overexpressed in the cerebellar cortex including *Aldh5a1*, *Astn2*, *Auts2*, *Dpp10* and *Sez6l2*, which are also present in our developing PC gene cluster. Interestingly, a recent transcriptomic analysis in human post-mortem cerebellum from ASD and control individuals identified three co-expression modules significantly associated with ASD that include several genes present in the PC development cluster identified in our study, including *NDUFA5* as a hub gene in one of the downregulated clusters³⁷. In addition, co-expression network analysis with human BrainSpan data identified an ASD gene enrichment including *ANK2* in thalamus and cerebellum during postnatal human development³⁸. These studies support our findings but also suggest that gene networks in cerebellar cell populations other than PCs and across different developmental epochs are relevant to ASD. In future studies, it will be important to employ a similar approach to ours to all other subpopulations of the cerebellum and other brain regions across the entire developmental period in mouse but also human tissue. Advances in single-cell sequencing experiments across development are poised to identify and compare cell- and stage-specific expression signatures.

The identification of spatiotemporally distinct ASD gene clusters in the developing cerebellum and cortex that correlate with ID suggest genetically distinct ASD endophenotypes. These results raise the interesting prospect of genes and pathways associated with these modules being of use in predicting, upon initial diagnosis, eventual ASD phenotype severity, patient stratification and, ultimately, targets for therapeutic interventions.

Methods

Animals. Male C57/BL6 mice were used in this study. All animal work was approved by the University of Oxford Ethics Panel and in accordance with UK Home Office regulations.

Laser Capture Microdissection and RNA Extraction. Laser capture microdissection was performed as previously described³⁹. For the time points P4–P21, 1000 individual PC soma were collected randomly from 2–3 parasagittal tissue sections per mouse. For P0 cerebella, where it was difficult to isolate individual cells, clusters of PCs were isolated from 6 parasagittal tissue sections. Three biological replicates per time point were collected. Total RNA was extracted using the RNeasy Micro Kit (Qiagen) according to the Fibrous Tissues protocol. RNA quality was assessed on a 2100 BioAnalyzer using the RNA 6000 Pico Assay (Agilent Technologies). All RNA samples used for deep sequencing had an RNA Integrity Number (RIN) of ≥ 5 . No principal components of the data were significantly correlated to the RIN values ($q < 0.05$).

Library Preparation and Sequencing. cDNA libraries were prepared using the SMARTer[®] Ultra[™] Low RNA Kit for Illumina[®] sequencing (Clontech), followed by NEBNext[®] DNA Library Prep Master Mix Set for Illumina[®] (New England Biolabs) according to manufacturers' instructions except for the use of own custom indexes⁴⁰. Sequencing of multiplexed, 100-bp paired-end libraries was done on an Illumina[®] HiSeq system using TruSeq SBS v3 chemistry. We acquired, on average, 61 million (range 38–100 million) paired-end 100-bp reads (Supplementary Table S2).

Read Processing and Alignment. The same quality control and alignment procedures were followed for all RNA-sequencing data used in this study. Data quality was visualized through FastQC v0.9.2 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>), and reads were trimmed using FASTX Toolkit v0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit/index.html) and Cutadapt v1.2.1⁴¹ for adapter trimming and removal of other over-represented sequences. Reads were also trimmed when quality scores dropped below 20. For the 5' paired-end Purkinje cell (PC) data and the single-end neocortex data, the reads were trimmed and retained, unless the resulting reads were less than 30 bp in length. Due to heavy adapter contamination in the 3' paired-end PC data this was run in two iterations, firstly with more strict settings on the 3' reads, followed by a second and less stringent run on the 5' reads (seeking to maintain longer reads for guiding the alignment).

Alignment to the mouse reference genome GRCm38/mm10 was executed with the Genomic Short-read Nucleotide Alignment Program (GSNAP) of the Genomic Mapping and Alignment Program for mRNA and EST sequences (GMAP) package version 2012-07-20⁴². GSNAP was run for the paired-end PC data, with an estimated expected insert size of 220 bp and with an estimated deviation of 50 bp. GSNAP was run for the single-end read neocortex data with the same options (with no need for insert size parameters). Following this, for both datasets, only uniquely mapping reads were retained, and count tables were produced using HTSeq version 0.5.3p9⁴³ with intersection-strict settings (a conservative measure only counting reads fully contained within features). All subsequent analyses in R were run within R version 3.1.1, and Bioconductor (biobase) version 2.26.0⁴⁴.

Investigation into read quality returned evidence of substantial SMARTer adapter contamination⁴⁵, from which initial trimming led to a mean loss of 4.89% of 5' reads and 42.23% of 3' reads. The contamination and resultant removal/orphaning of nearly half of all read pairs were most likely due to the low levels of RNA used for sequencing, with a mean concentration of 357 pg/μl and mean RIN value of 6.75. To circumvent this issue, heavier trimming was applied to the 3' reads (to a minimum of 17 bp before discarding) while original thresholds were maintained for the 5', therefore providing guidance during alignment with a specified estimated insert size of 220 bp (and allowed deviation of 50 bp), resulting in an average removal of 15.34% of the 3' reads.

An equivalent procedure was followed for the neocortex data, using the data for the central layer (layer 4; the only layer completely isolated in this study). Please refer to Fertuzinhos *et al.*³⁰ for more information on these neocortex libraries. The read trimming we applied based on quality scores resulted in a mean discarding of 0.31% of reads (0.26–0.38%), and read trimming due to contaminant removal resulted in a mean discarding of 3.75% of reads (2.58–6.35%). Of these remaining reads there was a mean of 98.38% (98.24–98.56%) of reads successfully mapped, with 59.95% (56.76–62.23%) having been uniquely mapped and retained for analysis.

Gene expression was assessed after read counting using HTSeq, and the normalization performed using DESeq 2²³. RPKM were manually computed for each gene based on the number of read mapped. All analyses were performed on the normalized counts. Across all samples, a total of 20,928 expressed genes (coding and non-coding) with a minority of missing values across all samples were retained for analysis. We observed the mean correlation coefficients of gene expression values to be considerably higher between samples within time points ($r^2 = 0.93$; range 0.86–0.98) than across time points ($r^2 = 0.65$; range 0.34–0.95; Supplementary Fig. S1A). Picard metrics are reported in Supplementary Table S6.

Weighted Gene Co-expression Network Analysis (WGCNA). Using variance stabilized values from DESeq 2 we applied the WGCNA R package version 1.41.1²⁰. Genes with greater than 50% missing values and/or zero variance were removed from analysis. Soft thresholding power value was set at 3 and 4 for the Purkinje cells and neocortex respectively, based on the scale-free topology fit (signed) and mean connectivity of the network. We set a minimum module size of 30 and the initial modules were merged based on eigengene identity using a dendrogram height of 0.5. Eigengenes were produced for each module by calculating their first principal components, thereby explaining the maximum amount of variation of expression levels. See Supplementary Table S7 and Supplementary Figs S2 and S5 for further details.

GO, KEGG, and MGI and Disease Enrichment. All genes used in enrichment testing (gene-lists, clusters/modules, and backgrounds) were reduced to protein coding genes only. All enrichment testing was performed with Fisher's exact tests providing q-values (Benjamini-Hochberg FDR corrected p-values), which were then corrected for gene-length bias using tools within the GOSec R package version 1.18.0⁴⁶. We used all genes with non-zero read count and non-zero variance within the Purkinje cells as a background gene set for the enrichment analyses. Gene Ontology enrichment was performed with GOSec provided GO terms. GO terms were filtered through REVIGO (Reduce and Visualize Gene Ontology)⁴⁷, which labels terms exhibiting semantic similarity as redundant, with strict settings of allowed similarity at 0.5. The KEGG (Kyoto Encyclopedia of Genes and Genomes) database⁴⁸ and MGI (Mouse Genome Informatics) database⁴⁹ enrichment tests were performed with pathway and phenotype information accessed through KEGGREST version 1.6.0 and BioMART version 2.22.0, respectively.

For disease enrichment testing with human gene lists, these were translated to their corresponding orthologs in mice, and all counts were reduced further to those with 1:1 orthologs (between human and mice) only. Enrichment for ataxia-associated genes was tested with candidate lists obtained through literature searches (Supplementary Table S1). Testing with autism-associated genes used both human candidate and mouse model genes obtained from the SFARI database (accessed 2nd Dec 2014)²⁵. The SFARI resource annotates genes by degrees of confidence of their associations with ASD (1–6, decreasing in confidence), and by whether they are associated with syndromic forms of ASD. Only genes belonging to SFARI categories 4 to 1 (“minimal evidence”, “suggestive evidence”, “strong candidate” and “high confidence”) were used to conduct enrichment analyses. Testing with schizophrenia-associated genes used a human candidate gene list⁵⁰. We used genes found to be expressed in the tissue of interest (PCs or neocortex) and with one-to-one orthologs in human and mouse as the background gene set.

Differential Expression Analysis. Differential expression analysis was performed with the R package DESeq 2 version 1.6.2²³. This provided a platform for applying a likelihood ratio test, suitable for time-series analysis with the aim of comparing models with and without time factor. Significant results were determined as those with a q-value < 0.05, and these were subsequently separated into positive and negative fold-change, providing two clusters. DESeq 2 statistics are reported in Supplemental Table S8.

Permutation Testing. Permutation testing was performed to statistically quantify the differing trend in odds ratios with an empirical p-value. For each permutation, all four contingency tables and subsequent odds ratios were simulated (with matching sampling distribution) using Patefield's algorithm, and those tables containing zeroes were adjusted with Haldane's correction. From these, the log Ratio of Odds Ratio (log₁₀ ROR) was calculated for each tissue, before taking the absolute difference between these. For each test, these tables were permuted 10⁶ times. The test was run twice on stratified candidate lists, once with syndromic ASD human candidates, and then with all other ASD human candidates. The p-values from the two stratified runs of permutation testing were adjusted for multiple testing, with significant results determined as those FDR < 0.05.

Accession Codes. Sequencing data have been deposited into the GEO repository (accession number GSE86824).

References

1. la Torre-Ubieta de, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* **22**, 345–361 (2016).
2. Becker, E. B. E. & Stoodley, C. J. Autism spectrum disorder and the cerebellum. *Int Rev Neurobiol* **113**, 1–34 (2013).
3. Wang, S. S. H., Kloth, A. D. & Badura, A. The Cerebellum, Sensitive Periods, and Autism. *Neuron* **83**, 518–532 (2014).
4. Stoodley, C. J. Distinct regions of the cerebellum show gray matter decreases in autism, ADHD, and developmental dyslexia. *Front. Syst. Neurosci.* **8**, 92 (2014).
5. D'Mello, A. M., Crocetti, D., Mostofsky, S. H. & Stoodley, C. J. Cerebellar gray matter and lobular volumes correlate with core autism symptoms. *Neuroimage Clin* **7**, 631–639 (2015).
6. Wegiel, J. *et al.* Brain-region-specific alterations of the trajectories of neuronal volume growth throughout the lifespan in autism. *Acta Neuropathol Commun* **2**, 28 (2014).
7. Brinke Ten, M. M. *et al.* Dysfunctional cerebellar Purkinje cells contribute to autism-like behaviour in Shank2-deficient mice. *Nat Commun* **7**, 1–14 (2016).
8. Tsai, P. T. *et al.* Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. *Nature* **488**, 647–651 (2012).
9. Reith, R. M. *et al.* Loss of Tsc2 in Purkinje cells is associated with autistic-like behavior in a mouse model of tuberous sclerosis complex. *Neurobiol. Dis.* **51**, 93–103 (2013).
10. Kern, J. K. Purkinje cell vulnerability and autism: a possible etiological connection. *Brain Dev* **25**, 377–382 (2003).
11. Donkelaar ten, H. J., Lammens, M., Wesseling, P., Thijssen, H. O. M. & Renier, W. O. Development and developmental disorders of the human cerebellum. *J Neurol* **250**, 1025–1036 (2003).
12. Limperopoulos, C. & Plessis, D. A. J. Disorders of cerebellar growth and development. *Curr. Opin. Pediatr.* **18**, 621–627 (2006).
13. Volpe, J. J. Cerebellum of the Premature Infant: Rapidly Developing, Vulnerable, Clinically Important. *J Child Neurol* **24**, 1085–1104 (2009).
14. Limperopoulos, C. *et al.* Does cerebellar injury in premature infants contribute to the high prevalence of long-term cognitive, learning, and behavioral disability in survivors? *Pediatrics* **120**, 584–593 (2007).
15. Biran, V., Verney, C. & Ferriero, D. M. Perinatal cerebellar injury in human and animal models. *Neurol Res Int* **2012**, 858929–9 (2012).
16. Lohof, A. M., Letellier, M., Mariani, J. & Sherrard, R. M. In *Handbook of the Cerebellum and Cerebellar Disorders* (eds Manto, M., Schmähmann, J. D., Rossi, F., Gruol, D. L. & Koibuchi, N.) 257–279 (Springer Netherlands, 2013).
17. Kapfhammer, J. Cellular and molecular control of dendritic growth and development of cerebellar Purkinje cells. *Prog Histochem Cytochem* **39**, 131–182 (2004).
18. Altman, J. & Bayer, S. A. *Development of the cerebellar system* (CRC Press, 1997).
19. Yamada, K. & Watanabe, M. Cytodifferentiation of Bergmann glia and its relationship with Purkinje cells. *Anat Sci Int* **77**, 94–108 (2002).
20. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
21. Chopra, R. & Shakkottai, V. G. Translating cerebellar Purkinje neuron physiology to progress in dominantly inherited ataxia. *Future Neurol* **9**, 187–196 (2014).
22. Leto, K. *et al.* Consensus Paper: Cerebellar Development. *Cerebellum* **15**, 789–828 (2016).
23. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq 2. *Genome Biol.* **15**, 550 (2014).
24. Reeber, S. L., Otis, T. S. & Sillitoe, R. V. New roles for the cerebellum in health and disease. *Front. Syst. Neurosci.* **7**, 83 (2013).
25. Basu, S. N., Kollu, R. & Banerjee-Basu, S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* **37**, D832–6 (2009).
26. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nature Genetics* **47**, 582–588 (2015).
27. Darnell, J. C. *et al.* FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* **146**, 247–261 (2011).
28. Iossifov, I. *et al.* De Novo Gene Disruptions in Children on the Autistic Spectrum. *Neuron* **74**, 285–299 (2012).
29. Steinberg, J. & Webber, C. The Roles of FMRP-Regulated Genes in Autism Spectrum Disorder: Single- and Multiple-Hit Genetic Etiologies. *Am J Hum Genet* **93**, 825–839 (2013).
30. Fertuzinhos, S. *et al.* Laminar and Temporal Expression Dynamics of Coding and Noncoding RNAs in the Mouse Neocortex. *Cell Reports* **6**, 938–950 (2014).
31. Jeste, S. S. & Geschwind, D. H. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol* **10**, 74–81 (2014).
32. Liu, X.-Q., Paterson, A. D. & Szatmari, P. Autism Genome Project Consortium. Genome-wide linkage analyses of quantitative and categorical autism subphenotypes. *Biol Psychiatry* **64**, 561–570 (2008).
33. Vieland, V. J. *et al.* Novel method for combined linkage and genome-wide association analysis finds evidence of distinct genetic architecture for two subtypes of autism. *J Neurodev Disord* **3**, 113–123 (2011).
34. Chaste, P. *et al.* A Genome-wide Association Study of Autism Using the Simons Simplex Collection: Does Reducing Phenotypic Heterogeneity in Autism Increase Genetic Homogeneity? *Biol Psychiatry* **77**, 775–784 (2015).
35. Parikshak, N. N. *et al.* Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell* **155**, 1008–1021 (2013).

36. Menashe, I., Grange, P., Larsen, E. C., Banerjee-Basu, S. & Mitra, P. P. Co-expression Profiling of Autism Genes in the Mouse Brain. *PLoS Comput. Biol.* **9**, e1003128 (2013).
37. Parikshak, N. N. *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
38. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
39. Dulneva, A. *et al.* The mutant Moonwalker TRPC3 channel links calcium signaling to lipid metabolism in the developing cerebellum. *Hum Mol Genet* **24**, 4114–4125 (2015).
40. Lample, S. *et al.* Improved workflows for high throughput library preparation using the transposome-based nextera system. *BMC Biotechnology* **13**, 104 (2013).
41. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**, 10–12 (2011).
42. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
43. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
44. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
45. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology* **30**, 777–782 (2012).
46. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14 (2010).
47. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800 (2011).
48. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
49. Blake, J. A. *et al.* The Mouse Genome Database: integration of and access to knowledge about the laboratory mouse. *Nucleic Acids Res.* **42**, D810–7 (2014).
50. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

Acknowledgements

We thank S. Lee and A. Heger for technical and analytical support and T.G. Belgard for critical comments on this manuscript. We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z and MRC Hub grant G0900747 91070) for the generation of the sequencing data. This work was supported by the Royal Society and the UK Medical Research Council. The authors declare no conflict of interest.

Author Contributions

H.C., C.P.P., W.H. and E.B.E.B. designed experiments. H.C., A.D. and W.H. performed experiments and data analysis. H.C., C.P.P., W.H. and E.B.E.B. wrote the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-37284-1>.

Competing Interests: The authors declare no competing interests.

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 4.0 International License, which permits use, sharing, adaptation, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019