Atypical neurogenesis in induced pluripotent stem cell (iPSC) from autistic individuals

Dwaipayan Adhya[1,3,*], Vivek Swarup[2,*], Roland Nagy[3], Lucia Dutan[3],Carole Shum[3], Kamila Maria Jozwik[4], Maria Andreina Mendez[5], Jamie Horder[5], Paulina Nowosiad[3], Irene Lee[6], David Skuse[6], Eva Loth[5], Deirdre Howley[5], Frances A. Flinter[7], Declan Murphy[5], Grainne McAlonan[5], Daniel H. Geschwind[2,9], Jack Price[3,8], Jason Carroll[4], Deepak P. Srivastava[3,8]§, & Simon Baron-Cohen[1]§

[1]Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, CB2 8AH UK.

[2]Program in Neurogenetics, Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA.

[3]Department of Basic and Clinical Neuroscience, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, SE5 9NU, UK.

[4]Cancer Research UK Cambridge Institute, Cambridge CB2 0RE, UK.

[5]Department of Forensic and Neurodevelopmental Sciences, Sackler Institute for Translational Neurodevelopment, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK.

[6]Behavioural and Brain Sciences Unit, Population Policy Practice Programme, Great Ormond Street Institute of Child Health, University College London, London WC1N 1EH, UK.

[7]Department of Clinical Genetics, Guy's & St Thomas' NHS Foundation Trust, London, UK.

[8]MRC Centre for Neurodevelopmental Disorders, King's College London, London, UK.

[9]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA.

§ Joint senior authors

* Joint first authors

Short title: Atypical neurogenesis in autism iPSC-derived neurons

Abstract

Autism is a heterogenous collection of disorders with a complex molecular underpinning. Evidence from post mortem studies using adult brains have identified atypical co-expression of genes which may occur during the prenatal period. Recent studies using induced pluripotent stem cells (iPSCs) generated from autistic individuals have suggested that prenatal development is a critical period for the emergence of pathophysiology associated with this condition. However, how early during development these differences emerge and whether such alterations can be seen across the development of multiple brain regions is unclear. In this study we investigated whether early prenatal stages of neurodevelopment differ between iPSCs generated from typically developing and autistic individuals. We specifically selected autistic individuals unrelated to each other with a heterogeneous genetic background and no known comorbidities, to probe for common molecular phenotypes. Differentiation of iPSCs towards a cortical lineage revealed abnormal cell fate acquisition from an early stage of development. Interestingly, abnormal differentiation occurred in the absence of alteration in cell proliferation during cortical differentiation, differing from previous studies. Moreover, these effects appeared specific for the acquisition of a cortical fate, as differentiation of iPSCs towards a midbrain lineage was not accompanied by differences in neurogenesis between typically developing and autism iPSC lines. RNA-sequencing on a subset of our cohort further revealed autism-specific signatures during cortical differentiation similar to that observed in post mortem studies, indicating a potential common biological mechanism. Together, these data suggest unique developmental differences associated with autism may establish at an early prenatal stage.

Introduction

Autism spectrum conditions (henceforth autism) are a genetically heterogeneous spectrum of neurodevelopmental conditions[1-3] diagnosed on the basis of impaired social-communication, alongside unusually narrow and repetitive interests and activities[4]. It is known to affect several brain regions such as the primary sensory cortex, association and frontal cortex, and parietal-occipital circuits[5, 6], as well as the medial prefrontal cortex, superior temporal sulcus, temporoparietal junction, amygdala, and fusiform gyrus[7, 8]. Based on clinical criteria, autism is classified into syndromic and non-syndromic forms. Individuals carrying single gene mutations, copy number variations and/or chromosomal abnormalities, in addition to an autism diagnosis are usually classified as 'syndromic'[9]. Non-syndromic autism is characterized by individuals with a primary diagnosis of autism that is not associated with a single mutation in a high risk genetic variant[9]. The symptoms of autism cannot be detecting until twelve to eighteen months of age[10], and apart from elevated steroidogenic activity in amniotic fluid previously reported by our group[11, 12] there is a paucity of prenatal markers reliably associated with an autism diagnosis. This has resulted in the understanding that the autistic phenotype might start to appear after birth. However, an increasing number of pre-clinical studies indicate that perturbation during critical periods of development maybe key for the emergence of autism[13]. Moreover, autism post mortem studies using adult brains have identified putative prenatal gene expression pathways[14], further suggesting that abnormal molecular and cellular events during gestation may contribute to autism.

Recent methodological advances in the field of induced pluripotent stem cell (iPSC) technology[15-20] has made it possible to study prenatal cellular behaviour in autism in detail, something that was not possible using post mortem brains. IPSCs have similar abilities to

embryonic stem cells to generate any tissue of the body. These can then be differentiated into neurons of various lineages[15-17, 19, 20]. As the neurons contain the same genetic information as the individual from whom it was derived, typical or autistic, its cellular behaviours is influenced by its genetic background. Using these methods, studies have shown significant anomalies in cellular/molecular behaviour during prenatal-equivalent periods of development in autistic individuals[21-23]. One such study generated iPSCs from autistic individuals who were comorbid for macrocephaly, and demonstrated: (1) atypical cortical differentiation and increased cell proliferation of cortical neural precursor cells (NPCs) from iPSCs, and (2) an imbalance in excitatory (glutamate-producing) and inhibitory (GABA-producing) receptor activity[21]. More recently, using the same collection of iPSCs, an acceleration in neuronal maturation was found to be dependent on early cortical NPC development, as circumventing this stage of cortical development by direct conversion of iPSCs into maturing neurons did not recapitulate these effects[23]. In another study, 3D neural cultures were generated from iPSCs generated from autistic participants also with macrocephaly; here an overproduction of GABAergic neurons was observed[22]. From these studies it is becoming clear that cellular and molecular phenotypes associated with autism likely start before birth, and possibly at a very early stage of brain development[23], and that some phenotypes may be associated with macrocephaly. The main phenotypes that appeared in these studies suggested in an acceleration in neuronal maturation as well as anomaly in excitatory and inhibitory neuron population and function. This latter phenotype has also independently reported using magnetic resonance spectroscopy (MRS) of autistic individuals, where abnormalities in levels of excitatory glutamate and inhibitory GABA metabolites were found[24, 25]. However, it is of note that in these iPSC studies atypical neural differentiation in autism was associated with higher cell proliferation[21-23]. As autistic participants in these studies also had macrocephaly it is yet to be determined whether the observed abnormal development is due to this comorbidity.

Nevertheless, altered gene expression network dynamics were also identified at early stages of development in these iPSCs[23]. However, as macrocephaly is present only in a subset of autistic individuals[26] it is not known if abnormal development can be generalised to atypically developing individuals without macrocephaly. Moreover, whether the alteration in gene expression networks at an early stage are also accompanied by a difference in precursor population composition is unknown. Finally, as these studies have predominantly focused on the development of forebrain/cortical neurons, it is yet to be tested whether abnormal development can also be observed in neural precursors fated towards a different lineage.

In this study, we have selected a cohort of autistic individuals without macrocephaly and with heterogeneous genetic backgrounds to represent the wider autism population. In this collection of iPSCs, we have observed atypical cortical differentiation in autism lines, including a divergent cellular identity starting from early stages of development (equivalent to neural tube formation at ~4 weeks gestation). Interestingly autism iPSCs did not display any abnormalities in cell proliferation. Furthermore, iPSCs were not compromised in their ability to generate precursors fated to become midbrain neural cells. Finally, we investigated whether atypical cortical differentiation was autism-specific by comparing it with observations made in post mortem studies, and further checked that the phenotypes we observed were not a result of non-specific effects.

Study participants, Materials and Methods

Induced pluripotent stem cells

IPSCs were produced from keratinocytes from plucked hair follicles using a previously established protocol[27] (**Supplementary Figure S1**). 9 autistic individuals – including six non-

syndromic autistic individuals (ASDM1, 004ASM, 010ASM, 026ASM, 132ASM, 289ASM), one individual with 3p deletion syndrome (245ASM) and two autistic individuals with a mutation in the NRXN1 gene (109NXM, 092NXF), and 6 typically developing individuals were chosen (**Supplementary Table S1, S2**). All autistic individuals had a primary diagnosis of autism. For clinical diagnosis of autistic individuals included in our RNA-sequencing based gene network analyses, see Table S1. Participants were recruited for this study under approval by NHS Research Ethics Committee (REC No 13/LO/1218); informed consent and methods were carried out in accordance to REC No 13/LO/1218.

Neuronal differentiation

We differentiated iPSC lines into cortical neurons using a well-established method based on dual SMAD inhibition; this results in the recapitulation of key hallmarks of corticogenesis and the generation of cortical neurons[17, 28]. iPSCs were differentiated till early neuron stage – day 35 (**Figure 1A**) (see extended experimental procedures). iPSCs were also differentiated to midbrain floorplate precursors using established protocols[19, 20].

Immunocytochemistry

Cortical differentiation of all aforementioned autism and control iPSCs were characterised using immunocytochemistry. iPSCs were differentiated towards a cortical lineage till day 9, day 21 and day 35 and tagged with antibodies of appropriate markers associated with each developmental stage (see extended experimental procedures). iPSCs were differentiated towards midbrain floor plate lineage till day 11. Nuclei were stained using DAPI, and imaging was performed using a 40× objective on a Leica SP5 confocal microscope (**Figure 1B**). High throughput confocal imaging was performed at day 9, day 21 and day 35 of cortical differentiation and at day 11 of midbrain differentiation on the Opera Phenix High-Content

Screening System (Perkin Elmer), and cell type analysis was performed using the Harmony High Content Imaging and Analysis Software (Perkin Elmer).

EdU labelling

IPSCs at day 0, day 9, day 21 and day 35 stages of differentiation were labelled with EdU (5-ethynyl-2´-deoxyuridine) using the Click-iT EdU Assay (Invitrogen). Cells were incubated with EdU for 4 hours at 37°C, then an additional 4 hours with EdU-free media. After incubation, labelled cells were fixed and prepared for detection using the Click-iT reaction cocktail. Nuclei were labelled using Hoechst 33342. EdU was detected with Alexa Fluor 488. Number of EdU-labelled cells were recorded as a percentage over total number of live nuclei. Imaging and analysis were done using the Opera Phenix High-Content Screening System (Perkin Elmer) and the Harmony High Content Imaging and Analysis Software (Perkin Elmer) respectively.

Statistics

Quantification of cell types was performed using the Harmony High Content Imaging and Analysis Software (Perkin Elmer). Percentage of cells positive for desired marker (probed for Pax6, Tuj1, EMX1 or Gad67) versus total number of live cells (probed using DAPI) was calculated. To take into consideration variability associated with iPSC differentiation, 8 independent experimental replicates of 2 clones per individual was used at every stage. Care was taken to ensure that fluorescence intensity was calculated only from known intracellular location of markers (Pax6, EMX1: nucleus; Tuj1, Gad67: cytoplasm). Appropriate threshold of fluorescent intensity was selected, and maintained throughout the experiment for each probe (**Supplementary Table S3**). Independent 2-group t-test was used to check significant difference between autism and control using p-value ≤ 0.05. One way ANOVA was performed to investigate in-group variance. All statistical analysis was performed on R statistical software.

RNA-sequencing

RNA-sequencing was performed from a subset of our cohort, on 2 clones from each participant (ASDM1, 004ASM, 010ASM, CTRM1, CTRM2, CTRM3), and each clone had 2 technical replicates. Starting with 500ng of total RNA, poly(A) containing mRNA was purified and libraries were prepared using TruSeq Stranded mRNA kit (Illumina). Unstranded libraries with a mean fragment size of 150bp (range 100-300bp) were constructed, and underwent 50bp single ended sequencing on an Illumina HiSeq 2500 machine. Reads were mapped to the human genome GRCh37.75 (UCSC version hg19) using STAR: RNA-seq aligner[29]. Quality control was performed using Picard tools (Broad Institute) and QoRTs[30]. Gene expression levels were quantified using an union exon model with HTSeq[31].

Differential gene expression (DGE)

DGE analysis was performed using R statistical packages[32] with gene expression levels adjusted for gene length, library size, and G+C content (henceforth referred to as "Normalized FPKM"). A linear mixed effects model framework was used to assess differential expression in $\log_2$(Normalized FPKM). Autism diagnosis was treated as a fixed effect, while also using technical covariates accounting for RNA quality, library preparation, and batch effects as fixed effects in this model.

Weighted gene coexpression network analysis

The R package weighted gene coexpression network analysis (WGCNA)[33] was used to construct coexpression networks as previously shown[14]. Biweight midcorrelation was used to assess correlations between $\log_2$(Normalized FPKM). For module-trait analysis, 1st principal component of each module (eigengene) was related to an autism diagnosis in a linear mixed effects framework as above, replacing the expression values of each gene with the eigengene.

Gene sets

A SFARI autism associated gene set was compiled using the online SFARI gene database, AutDB, using "Gene Score" as shown previously[14]. We obtained dev_asdM2, dev_asdM3, dev_asdM13, dev_asdM16 and dev_asdM17 modules from an independent transcriptome analysis study using RNA-sequencing data from post mortem early developing brains[14]. Modules asdM12 and asdM16 were obtained from an autism post mortem gene expression study[34]. We obtained another three autism-associated modules: ACP_asdM5, dev_asdM13, ACP_asdM14 from an independent gene expression study profiling dysregulated cortical patterning genes in autism post mortem brain[35]. These gene sets were used to undertake enrichment analysis. Gene modules from schizophrenia post mortem studies were used as non-autism neuropsychiatric gene modules[36, 37]. All previously published gene modules were established using WGCNA. The Cancer Gene Census (CGC) was used to establish non-neuropsychiatric gene enrichment.

Gene set overrepresentation analysis

Enrichment analyses were performed either with logistic regression. All GO enrichment analysis to characterize gene modules was performed using GO Elite[38] with 10,000 permutations. Molecular function and biological process terms were used for display purposes.

Code availability

Computer code used in our analyses is available from the authors upon request.

Results

Participant overview

Of the nine autistic participants in this study, eight were male, with one female (**Supplementary Table S1**). Six participants were diagnosed as having non-syndromic autism, while three participants were diagnosed with syndromic autism. All the autistic individuals had genetically heterogeneous backgrounds and were also genetically unrelated. Two non-syndromic participants had deletion type CNVs in the 1p21.3 and 8q21.12 regions respectively, with *DYPD* and *PTBP2* being autism-associated genes affected in the former, while the latter also having a deletion in the *AXL* gene (**Supplementary Table S2**). We also detected a stop-gain mutation in the *SHANK3* gene of another non-syndromic participant from exome analysis (**Supplementary Table S2**). Of the three syndromic participants, two syndromic participants had deletion type CNVs in the 2p16.3 region, affecting the *NRXN1* gene, a well-established autism-associated gene, while the third carried a deletion in the 3p chromosomal region, which includes the *CONTACTIN4* gene (*CNTN4*)[39] (**Supplementary Table S2**). IPSCs were generated from follicular keratinocytes from each participant and reprogrammed using established methods[18].

Autism iPSCs diverge from typical development at precursor cell stages during cortical differentiation.

Multiple studies have implicated anomalies in cortical regions associated with autism[5-8]. Adult post-mortem brain studies have indicated altered cortical development[14, 34, 35]. Studies using iPSCs from autistic individuals with macrocephaly have suggested atypical cortical differentiation in autism associated with altered cell proliferation[21, 22, 40]. As the cerebral cortex is the most strongly identified region of the brain affected in autism, we differentiated our iPSCs towards a cortical lineage using a previously established protocol[17, 28]. This differentiation protocol recapitulates formation of neuroectoderm and directs neural precursors

to form primarily dorsal forebrain cells[17, 28]. As we were also interested in determining divergence between control and autism iPSCs, we took advantage of this protocol to focus on three distinct developmental stages (**Figure 1A**): (1) Day 9: early neural precursor stage, when stem cells form new precursor cells which self-organise into neural tube-like structures known as neural rosettes with a directional apical-basal arrangement; (2) Day 21: late neural precursor stage, a period during which neural precursor cells begin forming layers from the apical surface and are primed for differentiation into neurons as they move outwards; and (3) Day 35: immature cortical neurons, a stage at which precursors become post-mitotic and adopt a deep layer neuronal identity (**Figure 1B)**.

Although there is evidence of divergence between control and autism iPSCs during development, it has been suggested that this divergence was a consequence of altered cell proliferation associated with macrocephaly[21]. In order to determine if these divergences still hold true in the wider autism spectrum, we selected our cohort with a heterogeneous genetic background and not having macrocephaly. First, we asked whether iPSCs from control and autistic participants displayed differences in the expression of neuronal differentiation markers Pax6 and Tuj1 in early and late neural precursor cell stages. Pax6 is a robust marker for neural precursors of cortical lineage[41], while Tuj1 is a robust pan-neuronal and neural precursor marker[42]. It is to be noted that the cortical differentiation protocol used in this study produces variability in expression of neuronal markers. This phenomenon has been associated with stochastic fluctuations in activation of key transcription factors during cortical differentiation[43]. To account for this, we generated 8 independent experimental replicates from 2 clones per participant at every stage and recorded the average effect for each participant. In control precursor cells on day 9, Pax6 (**Pax6** Control: 90%) and Tuj1 (**Tuj1** Control: 59%) was expressed in majority of cells. On day 21, both markers remained highly expressed (**Pax6**; Control: 91%, **Tuj1**; Control: 78%) which was consistent with cortical differentiation from

control iPSCs (**Figure 1C-E**, **Table 1**). In contrast, examination of autism precursor cells on day 9 revealed significantly lower Pax6 and Tuj1 expression compared to controls (**Pax6** Autism: 34%. **Tuj1** Autism: 20%. $p \leq 0.05$) (**Figure 1C-E**, **Table 1**). Moreover, at day 21, Pax6 and Tuj1 expression in autism precursor cells was still significantly lower than controls; however, the extent to which these cells were different to control precursors was not as pronounced as at day 9 (**Pax6** Control: 91%, Autism: 72%; $p \leq 0.05$. **Tuj1**; Control: 78%, Autism: 64%. $p \leq 0.05$) (**Figure 1C-E, Table 1**). As expected, variability was observed throughout the differentiation protocol between experimental replicates. However, this variability was more pronounced in the autism iPSCs. ANOVA analysis revealed greater overall spread of data points and higher F-values in a majority of observed parameters during autism iPSC differentiation (**Supplementary Figure S2B, S2C**). Taken together, these data suggest that autism and control iPSCs show divergent neural cell identity during cortical differentiation which manifests during early stages but remains significant even at later stages of differentiation.

Control and autism iPSCs show similar proliferative capacity during cortical differentiation and negligible differences while generating midbrain floorplate precursor cells.

Altered proliferation of neural precursors associated with macrocephaly in autistic individuals has been linked to divergent cortical differentiation in autism iPSCs[21, 22, 40]. To investigate if differences we observed were associated with atypical cell proliferation, we EdU (5-ethynyl-2′-deoxyuridine; a nucleoside analogue of thymidine) labelled[44] developing control and autism iPSCs and examined rate of proliferation during cortical differentiation. Specifically, we examined proliferation in iPSC (day 0), early (day 9) and late (day 21) precursors as well as in day 35 immature neurons (**Figure 2A**). Control iPSCs displayed an

overall reduction in cell proliferation as their mitotic capacity was reduced during differentiation, a phenomenon predicted to occur during the process of cellular differentiation from stem cells[45] (**Figure 2A**). Intriguingly, the proliferation capacities of autism iPSCs was similarly reduced at each stage of cortical differentiation examined (**Figure 2A**). These data indicate that the observed differences in the expression of Pax6 and Tuj1 were not due to altered levels of proliferation between control and autism iPSCs.

Next, we wanted to rule out the possibility that the cellular identity differences observed during cortical differentiation were not due to any inherent abnormality associated with our autism iPSCs. We reasoned that no significant differences would be observed between control and autism iPSCs when differentiated towards a midbrain neuronal lineage, a region of the brain not known to be directly linked to the cognitive deficits associated with autism[46, 47]. Therefore, we differentiated control and autism iPSCs towards a midbrain floorplate lineage using established protocols[19, 20]. Here we differentiated iPSCs for 11 days, a stage at which majority of cells take on a midbrain floorplate precursor (mFPP) identity. Interestingly, nearly 100% of mFPPs from both control and autism iPSCs were positive for LMX1A an essential transcription factor required for defining a midbrain identity[48] (**Figure 2B, C**). Similarly, expression of the transcription factor FOXA2, which positively regulates neurogenic factors in dopaminergic precursor cells[49], also did not differ significantly between control and autism mFPPs. Taken together, these data indicated that autism iPSCs did not show midbrain developmental differences, and that their atypical behaviour was specific to cortical differentiation.

As our data indicates that autism iPSCs show abnormality in neural identity specifically of a cortical lineage, we wanted to know if there was a difference in the capacity of these iPSCs to generate neurons. We differentiated the iPSCs up until day 35, when immature cortical

neurons start to appear. Cells at this stage primarily express TBR1, a transcription factor expressed in early born neurons[17, 50]. In our study, we found similar levels of TBR1 positive cells in both control and autism groups (**Figure 2D, E**). In addition, cells from the autism group did not demonstrate greater variability in TBR1 expression compared to the controls (F-values: control = 53.68, autism = 30.08; $p<0.05$). These data suggest that whilst autism precursor cells displayed differences in cortical-fate proteins such as Pax6 and Tuj1, this did not translate into significant differences in their capacity to differentiate into immature neurons.

Altered development of forebrain precursor lineages in autism iPSCs

Early differences in cortical fate protein expression did not appear to affect cellular identity of immature neurons from autism iPSCs. Thus, while cortical differentiation generated precursors of the cortical lineage from both groups, we wanted to find out whether the precursors had identical neuronal identities or if they generated different cortical subtypes. We were particularly interested in investigating the development of the two major neuronal subtypes; namely dorsal forebrain and ventral forebrain neurons which are known to give rise to excitatory and inhibitory neurons respectively in the mature state[51, 52]. Specifically, dorsal forebrain precursors are known to give rise to cortical excitatory, glutamatergic neurons whereas ventral forebrain precursors give rise into inhibitory GABAergic neurons[51, 52]. The appearance of excitatory-inhibitory imbalance is a widely reported phenotype associated with autism[24, 25]. Studies by both Mariani et al., (2015) and Marchetto et al., (2017) have also demonstrated an imbalance in the ratio of excitatory and inhibitory neural precursors and neurons in 2D and 3D cultures differentiated from iPSCs from autistic individuals. However, in their studies the autistic participants also had macrocephaly and this imbalance was attributed to alterations in proliferative capacity associated with macrocephaly. As we did not

observe an alteration in the proliferation in neural precursors from our autism-iPSCs, we hypothesised that the appearance of excitatory-inhibitory imbalance could also be a result of an early divergence in precursor development. We looked at Emx1 and Gad67 expression in neural precursors at day 9 and day 21, and immature neurons at day 35 of differentiation (**Figure 3**, **Table 1**). Emx1 is expressed in dorsal forebrain precursors[52-54], while Gad67 is the rate limiting enzyme in the GABA synthesis pathway and expressed in ventral forebrain precursors[55, 56]. At day 9, Emx1 expression was significantly higher in control compared to autism neural precursors, even though it was highly expressed in both controls as well as autism precursors (**Emx1**; Control: 95%, Autism: 80%; $p \leq 0.05$) (**Figure 3B, C**). At day 21, Emx1 expression in both groups appeared to remain stable, with only a slight reduction in control precursors compared to day 9. At this stage in the control group expression of Emx1 was still significantly higher (**Emx1**; Control: 90%, Autism: 81%; $p \leq 0.05$) (**Figure 3B, C**). In day 35 immature neurons, Emx1 expression in both control and autism neurons was reduced compared to day 9 and day 21 precursors; however the reduction was significantly more acute in the autism group (**Emx1**; Control: 72%, Autism: 50%; $p \leq 0.05$) (**Figure 3B, C**). Gad67 expression in autism and controls followed a different trajectory. At day 9, the expression of Gad67 was significantly higher in the control precursors, while there was negligible expression in the autism precursors (**Gad67**; Control: 35%, Autism: 4%; $p \leq 0.05$) (**Figure 3B, C**). At day 21, there was reduction of Gad67 expression in the control group, but significant increase of Gad67 in the autism group (**Gad67**; Control: 26%, Autism: 27%; $p > 0.05$) (**Figure 3B, C**). At this stage, both control and autism precursors had similar Gad67 expression with no statistically significant differences between groups. By day 35, Gad67 expression in autism precursors overtook Gad67 expression in controls, and its expression in control precursors was further reduced (**Gad67**; Control: 17%, Autism: 48%; $p > 0.05$) (**Figure 3B, C**). Similar to what we observed with Pax6 and Tuj1 expressing cells, Emx1 and Gad67 expression also showed

conspicuous variability. Again, ANOVA revealed greater variability in majority of the parameters in the autism group (**Supplementary Figure S2A, S2C**). When syndromic and nonsyndromic autistic individuals were analysed separately, both groups still appeared to follow similar trends when compared to the control group (**Supplementary Figure S3**). Taken together, these data showed significant differences in the determination of neuronal subtype identity of cortical lineage, from control versus autism iPSCs. It was evident that autism precursor cells had a greater propensity to differentiate into neural cells of ventral forebrain lineage compared to controls, and that this effect was not a result of atypical cell proliferation. Overall, these data also showed increased variability in expression of key cortical differentiation genes in autism iPSCs. This phenomenon that was considerably reduced during midbrain differentiation. However, the variability did not appreciably affect the mean group differences between control and autism (**Figure 1E, 3C**).

Neurodevelopmental gene expression signatures in autism-iPSC-derived neurons

Based on our cellular analyses, we found two distinct phenotypes associated with cortical differentiation of autism iPSCs: *(1)* they show atypical neural differentiation of cortical lineage cells, *(2)* atypical differentiation into cortical precursor subtypes of dorsal forebrain and ventral forebrain fates. To assess whether the observed atypical neural differentiation was specific to autism, we undertook RNA-sequencing on day 35 cortically differentiated immature neurons derived from a subset of our autism- and control-iPSC cohort. Our aim was to identify global gene expression patterns and compare them with gene expression patterns reported in post-mortem brains. We undertook RNA-sequencing, then identified gene co-expression pathways which were then used to estimate preservation in gene co-expression pathways in post-mortem brains. This gave us a reliable estimate of autism-specific gene expression in our

iPSC-derived neurons and showed recapitulation of gene expression patterns in iPSC-derived neural cells. To our knowledge, a similar comparison had not been conducted before. We ensured that our analysis pipeline was consistent with those used to investigate transcriptomic changes in autism post-mortem brain studies (**Figure 4A, Figure S4**). Using these analysis methods, we established: (i) networks of genes dysregulated in autism iPSC-immature neurons and, (ii) the degree to which gene networks were dysregulated in post-mortem samples. Differential expression from RNA-sequencing data revealed control and autism samples clustering hierarchically into two separate groups based on gene expression patterns and principal component analysis (**Figure 4B**, **Supplementary Figure S5A, S5B**). To identify co-expressed genes, signed weighted gene coexpression network analysis (WGCNA) was performed. We identified 11 co-expression modules significantly associated with autism (labelled according to R-assigned colours, e.g., salmon, **Figure 4C**). We ranked the modules according to their module eigengene values (ME; the first principal component of the module) (**Figure 4C, D Supplementary Figure S6A**). Of the 11 modules, 5 modules were positively correlated, and 6 modules were negatively correlated in autism-iPSC immature neurons. The modules were assigned consensus functions based on gene ontology (GO) terms (**Supplementary Figure S7A**). The top 3 positively correlated modules having higher MEs in autism were, 'steelblue' (Cellular Metabolic Processes), 'lightgreen' (Neural Development) and 'white' (Immune Activation), and the top 3 negatively correlated modules having lower MEs in autism were, 'grey60' (Epigenetic Regulation), 'salmon' (Gene Regulation) and 'sienna3' (Chromosome Organisation) (**Figure 4E, F**). The 'steelblue' module was enriched for GO terms for metabolic functions associated with atypical cell proliferation (**Figure 4G**). The 'lightgreen' module was enriched for neural development functions including regulation of cell-cell adhesion, cognition, calcium mediated signalling and regulation of dendrite maturation associated with neural development. The 'white' module was enriched for non-

neuronal functions such as cytokine binding, regulation of DNA damage response, positive regulation of apoptosis and negative regulation of neuronal death (**Figure 4I**). The 'salmon' module was enriched for epigenetic gene regulatory functions such as RNA methyltransferase activity, and s-adenosylmethionine-dependant methyltransferase activity (**Figure 4J**). Finally, the 'sienna3' module was enriched for cellular gene regulatory functions such as nucleic acid binding, regulation of RNA metabolic process and regulation of gene expression (**Figure 4K**), while the 'grey60' module was enriched for chromosome organisation functions such as regulation of histone H3-K4 methylation, DNA binding and chromosome organisation (**Figure 4L**).

Gene co-expression modules in autism-iPSC immature neurons show significant correlation to co-expression modules from autism post-mortem brains

We used autism post-mortem gene modules from three studies to test autism gene network preservation in iPSC neurons[14, 34, 35]. Additionally, we used a set of 155 autism associated candidate genes from a previous study[14] using the Simons Foundation Autism Research Initiative (SFARI) database to identify effect of high impact autism associated genes in our modules. The SFARI list of autism genes is a database of genes collated according to the type of genetic variations from whole genome sequencing studies, rare genetic mutations and mutations causing syndromic forms of autism. It was first published in 2009[57] and an up-to-date reference for all known associated genes can be found at: https://gene.sfari.org/autdb/HG_Home.do. We mapped the SFARI autism associated genes with our gene modules and found them to be enriched in the negatively correlated 'salmon' (epigenetic gene regulation) module (p=0.002; odds ratio [OR] = 1.5) (**Figure 5A**). High impact SFARI genes such as cytoskeletal ANK2 (gene score: 1) and regulatory ARID1B (gene

score: 1; syndromic) were negatively enriched in this module and could be potential loss-of-function risk genes in our study. We then mapped 5 autism-associated developmental gene modules dysregulated in post mortem brains (APMB), from Parikshak et al., (2013) (dev_asdM2, dev_asdM3, dev_asdM13, dev_asdM16, dev_asdM17)[14], shown in Figure 5A. Of these 5 sets, dev_asdM2 and dev_asdM3 were downregulated in autism and represent DNA-binding and transcriptional regulation, while dev_asdM13, dev_asdM16 and dev_asdM17 were upregulated in autism and represent later phase neuronal functions and development of synaptic structure. The downregulated dev_asdM2 set was enriched in the top downregulated genes ('Top –ve DE', $p = 2\times10^{-4}$; OR = 1.8), as well as 'grey60' (p = 0.004; OR = 1.6) and 'sienna3' ($p = 10^{-5}$; OR = 2) which were downregulated modules. Similarly, the dev_asdM3 set was enriched in the top downregulated genes ('Top –ve DE', p = 0.008; OR = 1.5), and 'grey60' ($p = 3\times10^{-14}$; OR = 2.5) and 'sienna3' ($p = 4\times10^{-4}$; OR = 1.7) modules. The upregulated dev_asdM13 set was enriched in the top upregulated genes ('Top +ve DE', $p = 10^{-6}$; OR = 2.1), and 'lightgreen' ($p = 3\times10^{-9}$; OR = 3.1) and 'white' ($p = 10^{-6}$; OR = 2.3) which were upregulated modules. Similarly, the dev_asdM16 set was enriched in the 'lightgreen' module ($p = 10^{-4}$; OR = 2.6), and, the dev_asdM17 set was enriched in the top upregulated genes ('Top +ve DE', p = 0.002; OR = 1.7) and the 'lightgreen' module (p = 0.002; OR = 1.9). We then mapped two gene modules upregulated in the temporal and frontal cortex of the adult autism brain – APMB_asdM12 (a synaptic function module) and APMB_asdM16 (an immune module) from Voineagu et al., (2011) [34] (**Figure 5A**). The APMB_asdM12 module was enriched in the upregulated 'white' module (p = 0.04; OR = 1.8), while the APMB_asdM16 was enriched in the top upregulated genes ('Top +ve DE', $p = 5\times10^{-6}$; OR = 2.6) and the 'white' module ($p = 6\times10^{-5}$; OR = 2.7) (**Figure 5A**). Gene sets associated with attenuated cortical patterning or ACP [35] were also mapped (**Figure 5A**) and suggested greater prediction of ACP in autism iPSC neural cells. One additional aspect of this analysis was that it indicated

specificity of our module enrichment. Only the downregulated modules were enriched in downregulated post-mortem modules, while the upregulated modules only enriched in the upregulated post-mortem modules. This mutual exclusive clustering of enrichment suggested recapitulation of autism-associated gene expression pathways in our cohort (**Figure 5A**). We found even more evidence of specificity our gene expression modules. When correlated to schizophrenia and cancer gene modules, enrichment was poor and non-specific (**Figure 5B**). A module preservation analysis also revealed moderate to high preservation of our gene modules in similarly labelled gene modules from autism iPSC studies[21, 22] (**Supplementary Figure S7B**). However, we did not find significant correlation of differential gene expression with single nucleotide polymorphism (SNP) load in the exomes of the individual participants (**Supplementary Figure S8A, S8B**).

Discussion

The overarching aim of this study was to identify cellular identities during early stages of cortical development from control and autism iPSCs, and investigate specificity of these phenotypes in autism. We wanted to study this in a heterogeneous cohort of autistic individuals who did not have macrocephaly. To this end, here we have selected a cohort comprised of 6 autistic individuals with uncharacterised genetic background and 3 autistic individuals with known CNVs in high autism risk loci. First we found that during cortical differentiation control and autism iPSCs showed significant differences in development of early neural precursor cells. This effect persisted at a late precursor cell stage although to a lesser degree. Interestingly, we did not find proliferative capacities between control and autism to be significantly altered. This observation was contrary to previous reports that atypical proliferation was the primary cause for altered neurogenesis from autism iPSCs[21-23]. This also suggested a potentially

different mechanism was involved in producing the divergent precursor populations. When looking at cortical neuron subtypes, we found a divergence in the development of dorsal forebrain or excitatory precursors and ventral forebrain or inhibitory precursors from an early stage of development. On comparing transcriptomes between control and autism cortical neurons, we found that transcriptomic signatures in autism iPSC-derived neurons were highly correlated to those identified in post mortem brain studies. This showed autistic iPSCs when cortically differentiated, was significantly different from control iPSCs, and that based on gene expression signatures, these differences were specific to autism.

Landmark autism iPSC studies have demonstrated imbalance in excitatory versus inhibitory neural signalling in neurons[21] and cerebral organoids[22]. More recent work using these autism lines have highlighted early neurodevelopment as being key for the convergence of molecular pathologies in autism[23]. Building on this body of evidence we examined whether there were differences in cell identity during early neural differentiation starting from a stage equivalent to neural tube formation. When iPSCs were differentiated towards a cortical fate, we found significant differences in the number of cells expressing Pax6 and Tuj1 in autism compared to controls, with both Pax6 and Tuj1 levels being lower in autism neural precursors. As Pax6 is an essential transcription factor that determines the cortical identity of neural precursors[58], fewer Pax6 expressing cells from autism iPSCs may signify divergent differentiating behaviour. Fewer number of Tuj1 expressing cells were also observed during autism cortical differentiation, and as Tuj1 is expressed in all neurons and precursors this was consistent with the Pax6 expression. The differences in Pax6 and Tuj1 expression was more pronounced at the early precursor stage (day 9) than at the later precursor stage (day 21) when Pax6 and Tuj1 levels in autism cells appeared to be ‘catching up’. Interestingly, at day 35 immature neuron stage, there was negligible differences between control and autism neurons expressing TBR1, a transcription factor expressed in early-born post-mitotic neurons[17, 59]. This

implies that despite differences in the expression of Pax6 and Tuj1 in early and late precursor cells, there appears to be no alteration in the number of cortical neurons being generated during differentiation.

The iPSC studies that showed altered neurodevelopment to be associated with autism, selected autistic participants based on a comorbid diagnosis of macrocephaly[21-23, 40]. According to these studies, increased cell proliferation associated with macrocephaly was the cause for dysregulation of neurodevelopment and GABAergic versus glutamatergic neuron balance[21, 22]. However, not all autistic individuals have macrocephaly, and we wanted to study a more heterogeneous cohort of autistic individuals to find out if the assertions made in previous studies still held true in the wider autistic population. As a result we did not use macrocephaly as a criteria for recruitment. As expected, we did not find any differences in proliferative capacities between control and autism iPSCs in our study. This suggested that divergent cortical differentiation could be result of altered cell fate rather than altered proliferation of certain cell types in autism.

Due to Pax6 and Tuj1 expression in differentiating autism cells ‘catching up’ at day 21, and negligible differences in TBR1 expression at day 35, we wanted to investigate whether early differences in cortical differentiation had any effect on the neural subtype identities of cortical precursors. To this end, we looked at the appearance of precursors expressing markers for two major types of cortical neurons, EMX1 (expressed in precursors of the dorsal forebrain, and adult excitatory neurons) and Gad67 (expressed in precursors of the ventral forebrain, and adult inhibitory neurons). We found significant differences between control and autism in cells expressing EMX1 and Gad67. We also found cells expressing Gad67 to be decreasing over time in controls while increasing in autism. Higher levels of Gad67 cells in autism day 35 immature neurons in our study was similar to higher levels of Gad1/Gad67 cells observed in

autism iPSC-derived organoids[22]. The atypical alteration in Emx1 and Gad67 expressing cells could have consequences on ratio of dorsal and ventral forebrain neurons at adulthood when they have an excitatory and inhibitory function respectively, and could be a possible prenatal origin of the excitatory/inhibitory imbalance phenotype of autism[21, 22, 24, 25]. As we did not observe any differences in proliferative capacity between control and autism iPSCs, it is possible that the physiological differences associated with altered excitatory-inhibitory cell types were independent of the proliferative capacity of iPSCs, and instead a result of divergent cell fate at early precursor stages.

An important question that arose in our study was whether atypical maturation of autism iPSCs was specific to cortically differentiating cells. To investigate if this was a more general, widespread, aberration impacting the development of multiple brain regions, we differentiated control and autism iPSCs into midbrain precursors. The different neuronal lineage of midbrain floor plate precursors (mFPPs) meant iPSCs were exposed to a different set of patterning factors[19, 20]. This also allowed us to determine whether there was any inherent abnormality in the way our autism iPSCs were responding to patterning factors used in neural differentiation protocols. Surprisingly, differentiating iPSCs into mFPPs revealed negligible differences in between control and autism iPSCs. This was consistent with the fact that it is cortical brain regions such as the prefrontal cortex that are primarily affected in autism causing the deficits of cognitive functions such as social skills and communication, stereotyped and repetitive behaviour – the phenotypes which form the basis for its diagnosis[47], while lower brain regions though partially associated with autism spectrum conditions, their roles are not as well defined[46, 47].

Through our studies we also tried to account for a major caveat of working with iPSCs as a research model, i.e., variability. Indeed it is important to acknowledge the reported variance

in cellular behaviour which is characteristic during cortical differentiation of iPSCs[43]. Such variance could mask the observed phenotypes or be indicative of an issue with the iPSCs. In order to account for these possibilities, we have taken a number of technical steps to ameliorate this. First, we have used 2 iPSC clones per participant which were generated from 9 atypically developing autistic individuals in all cellular assays to account for both intra- and inter-donor variability. Second, we undertook 8 experimental replicates from each iPSC clone to account for experimental variance due to batch differences. We analysed the variance seen within, in addition to across, all lines, in order to assess whether cellular phenotypes observed were robust for each participant. We found that overall variance was greater in the autism group in the majority of cellular parameters measured compared to controls. As this variance was within the expected variability seen during cortical differentiation, this suggested the observed cellular phenotype were associated with autism and not due to variability in the protocol. Importantly, differentiation of the same iPSCs into mFPPs was not accompanied by increased variability as seen during cortical differentiation. To understand the relevance of this variability, this phenomenon will need to be studied further. Interestingly, it has been reported that during cortical differentiation precursor cells undergo through a multitude of 'microstates' which might be a result of stochastic fluctuations in activation of key transcription factors[43]. It would be interesting to determine whether the activity of these transcription factors are altered in autistic iPSCs, introducing abnormalities during the transition between microstates.

The aberrations we observed in our heterogeneous pool of autism iPSCs appeared to suggest a common set of prenatal cellular mechanisms associated with the condition. We then investigated signatures in autism iPSC immature neurons that were specific to autism, and quantify these signatures with the help of those reported in post mortem studies[14, 34, 35], which were performed on large cohorts having heterogeneous genetic backgrounds. In our study, we selected a sub-cohort of autistic and control participants to differentiate into neural cells and

undertake a similarly designed gene expression analysis as the post mortem studies. We first identified both positively and negatively expressed gene modules, then ran enrichment analyses with positively and negatively gene modules reported in the post-mortem studies. We found positive modules in our study to be enriched in positive modules in post-mortem brains, while negative modules to be enriched in negative modules in post-mortem brains. This meant that gene networks that were being dysregulated in autism iPSC-derived immature neurons were also being dysregulated in autism post-mortem brains. We found even more evidence that our gene expression modules were autism-specific, when we found enrichment analyses with post mortem schizophrenia gene modules and the cancer gene census did not yield disease-specific enrichment with our gene modules.

To conclude, the present study demonstrates significant differences between cortical differentiation of iPSCs from autistic individuals and typical controls. Not only did we find differences in development of precursor cells fated towards cortical neurons, we also found differences in development of dorsal and ventral forebrain precursors and immature neurons. However, no significant differences were found in development of midbrain precursors, indicating that autism associated atypical neural identity was limited to the cortical lineage. In addition, we did not find altered cellular proliferation in autism iPSCs, suggesting atypical neural identity was due to cell fate mechanisms rather than proliferation. To our knowledge, this is the first study to show aberrations specifically in cortical differentiation occurring at a developmental stage as early as neural tube formation (equivalent to ~4 weeks of gestation). We also observed high variability within experimental replicates during cortical differentiation but not midbrain differentiation, with variability increasing in cortically differentiating autism iPSCs. RNA-sequencing revealed cortical differentiation from autism iPSCs had autism-specific gene expression signatures. Future studies will reveal if the phenotypes we observed are a result of dysregulation of pathways responsible for early cortical development, such as

Wnt and Notch signalling[21], and will reveal if stabilising these pathways at an early stage of differentiation can help recover autism-associated dorsal/ventral forebrain precursor fates. If precursor fates are responsible for later imbalances in excitatory and inhibitory cell types in the brain, early interventions to stabilise signalling pathways might be able to ameliorate this effect.

Acknowledgments

We gratefully acknowledge the participants in this study. This study was supported by grants from the European Autism Interventions (EU-AIMS) and AIMS-2-TRIALS; the Wellcome Trust ISSF Grant (No. 097819) and the King's Health Partners Research and Development Challenge Fund – a fund administered on behalf of King's Health Partners by Guy's and St Thomas' Charity (Grant R130587) awarded to DPS; an Independent Investigator's Award from the Brain and Behavior Foundation (formally National Alliance for Research on Schizophrenia and Depression (NARSAD); Grant No. 25957), and Seed funding from Medical Research Council, UK (MR/N026063/1) awarded to DPS; the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115300, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution (JP, SBC, DPS, DM, GM); the European Union's Seventh Framework Programme (FP7-HEALTH-603016) (DPS, JP); the Mortimer D Sackler Foundation; the Autism Research Trust, the Chinese University of Hong Kong, and a doctoral fellowship from the Jawaharlal Nehru Memorial Trust awarded to D.A. SBC was funded by the Autism Research Trust, the Wellcome Trust, the Templeton World Charitable Foundation, and the NIHR Biomedical Research Centre in Cambridge, during the period of this work. The funding organizations had no role in the design and conduct of the study, in the collection,

management, analysis and interpretation of the data, or in the preparation, review or approval of the manuscript. We are grateful to Debbie Spain and Suzanne Coghlan for participant recruitment, to Rosy Watkins, Hema Pramod, Rupert Faraway, Pooja Raval, Kate Sellers, Michael Deans and Rodrigo Rafagnin for assistance during the study, and to Aicha Massrali, Arkoprovo Paul, Bhismadev Chakrabarti, Michael Lombardo, Rick Livesey and Mark Kotter for valuable discussions. We thank the Wohl Cellular Imaging Centre (WCIC) at the IoPPN, Kings College, London for help with microscopy.

Ethics, consent and permissions

Informed consent from participants have been taken before recruitment: Patient iPSCs for Neurodevelopmental Disorders (PiNDs) study' (REC No 13/LO/1218).

Consent to publish

We have obtained consent to publish from the participant to report individual patient data.

Availability of data and materials

Sequence data have been uploaded on synapse.org. Synapse ID: syn8118403, DOI: doi:10.7303/syn8118403

Authors' contribution

DA, JP, JC, DPS, SBC conceived the study and wrote the first draft. VS, DHG conceived and developed bioinformatics analysis framework and analysis. DA, RN, LD, PN, CS, KJ responsible for sample preparation and data analysis. GM was responsible for ethics application. GM, MAM, JH, IL, DS, EL, DH, FAF and DM responsible for recruiting and collecting hair samples from individuals with autism and controls. DPS and SBC oversaw the running of the project. All co-authors contributed to study concept, design, and writing of the manuscript. All authors read and approved the final manuscript.

Figure legends

Figure 1. Atypical cortical differentiation of autism iPSCs. (A) iPSCs were reprogrammed from autistic individuals and typical controls, followed by cortical differentiation to early neural precursors (day 9), late neural precursors (day 21) and immature neurons (day 35). (B) Representative images of immunofluorescence of iPSCs showed morphological changes during cortical differentiation. Day 9 early neural precursors expressing Ki67 and Nestin (scale bar: 10µm), day 21 late neural precursors expressing Pax6 (scale bar: 10µm), and day 35 neural cells expressing TBR1 and MAP2 (scale bar: 10µm). (C) At day 9 and day 21 precursor cell stages, both control and autism iPSCs expressed Pax6 and Tuj1. (D) Quantification of $Pax6^{+}$ and $Tuj1^{+}$ cells of individual participants (% cells positive per experimental replicate) showed significant differences between autism and control. (D) Histogram shows normal distribution of experimental data points, and demonstrates variability between control and autism. Mean values of % positive cells over time show significant difference between control and autism at both day 9 and day 21. (LV: Lentivirus reprogramming method used for generating these iPSCs; s: Participants with syndromic autism)

Figure 2. Control and autism iPSCs have similar proliferative capacities, similar midbrain differentiation capacities and are able to produce similar levels of early cortical neurons. (A) EdU labelling revealed no significant differences in cell proliferation during cortical differentiation of control or autism iPSCs. Dashed lines: control participants (inset plot: mean values; blue: control, red: autism). (B) Both control and autism iPSCs expressed LMX1A and FOXA2 when differentiated towards a midbrain floor plate precursor (mFPP) lineage, and some precursors were also starting to express Tuj1. (C) No significant group differences were found between control or autism iPSCs expressing LMX1A or FOXA2. (D-E) Control and autism iPSCs also expressed TBR1 at day 35 of cortical differentiation, and there was no significant difference in TBR1 expression across both groups, although variation within experimental replicates was high, when compared to midbrain differentiation. (LV: Lentivirus reprogramming method used for generating these iPSCs; s: Participants with syndromic autism)

Figure 3. Atypical differentiation into dorsal and ventral forebrain precursors in autism. (A) EMX1 was expressed at day 9, day 21 and day 35 in both control and autism groups. Gad67 expression in both groups was time dependant, it decreased over time in in controls, while increased over time in autism. (B) Quantification of $EMX1^{+}$ and $Gad67^{+}$ cells (% cells positive per experimental replicate) showed significant differences between autism and control. (C) Histogram shows distribution of experimental data points, demonstrating variability between control and autism, but there was clear difference in distribution of data points between groups. Mean values of % positive cells over time show significant difference between control and autism at every time point, except for Gad67 at day 21 precursor stage. (LV: Lentivirus reprogramming method used for generating these iPSCs; s: Participants with syndromic autism)

Figure 4. Transcriptome-wide gene co-expression network analysis in autism and control neurons. (A) Schematic of RNAseq data collection and analyses. (B) Gene expression and hierarchical clustering in control and autism iPSC neural cells (day 35) show significant gene expression differences between autism and control, with autism and control samples clustering in separate groups. Top 50 differentially expressed genes shown here. (C) 11 gene modules were correlated to autism. Modules with positive values indicated increased expression in autism iPSC-derived neural cells, while modules with negative values indicated decreased expression in autism iPSC-derived neural cells. Red dotted lines indicate Benjamini-Hochberg corrected p-values ($p<0.05$). R package designated module colour names were used. (D) Correlation network dendogram of significant gene modules and their relationship with autism (red denotes a positive relationship) revealed six top-correlated modules which were functionally relevant. (E) Module eigengene values of positively correlated gene modules revealed high positive correlation with autism, while negatively correlated gene modules revealed high negative correlation with autism.

Figure 5. High autism-specific enrichment of gene expression modules. (A) Positively correlated and negatively correlated gene modules were enriched in corresponding positively correlated and negatively correlated gene modules from post mortem studies. Negatively correlated salmon module was also enriched in SFARI list of high risk genes. Only OR>1.5 and $p\leq0.05$ has been reported (p-value in parenthesis). (B) Module enrichment in schizophrenia gene modules was non-specific, other than Gilman et al SCZ-Cluster I. Module enrichment in cancer gene census was also non-specific.

Supplementary Info

1. Supplementary methods
2. Supplementary Figure S1
3. Supplementary Figure S2
4. Supplementary Figure S3
5. Supplementary Figure S4
6. Supplementary Figure S5
7. Supplementary Figure S6
8. Supplementary Figure S7
9. Supplementary Figure S8
10. Supplementary Figure S9
11. Supplementary Figure S10
12. Supplementary figure legends
13. Supplementary table S1
14. Supplementary table S2
15. Supplementary table S3

Keywords:

Autism, iPSC, precursors, neural cells, cortical differentiation, midbrain differentiation, neurodevelopment, atypical neurogenesis, post mortem brain, transcriptome, functional genomics.

Supplementary methods

Study participants and neuronal differentiation

Keratinocytes were collected from autistic participants, and typical controls without an autism diagnosis (Ethics approved, 13/LO/1218) as part of a larger European studies (EU-AIMS, STEMBANCC). All participants were Caucasian, while controls were selected if they did not have diagnosis of any psychiatric conditions. These were reprogrammed into iPSCs using previously described methods [18, 27]. IPS cells were cultured in E8 medium (Life Technologies) with E8 supplement (Life Technologies). Cell type quantification and proliferation assays were set up on 96-well plates. 2 clones were selected from each participant, and each clone had 8 technical replicates. RNA-sequencing was performed on 2 clones from each participant, and each clone had 2 technical replicates. This design was maintained at all stages of neural differentiation recorded. Induction of neurons of cortical lineage was established using a modified dual SMADi protocol[17]. Once the cell culture reached 95% confluence, neural induction was initiated by changing the culture medium to support neural induction, neurogenesis and neuronal differentiation. A combination of N2- and B27-containing media with additives was used, henceforth called 'neuralising medium'. N2 medium consisted of DMEM/F12 (Sigma), N2 (Gibco). B27 medium consisted of Neurobasal (Invitrogen), B27 (Gibco). Neuralising medium was supplemented with 'dual SMADi' 1 μM Dorsomorphin (Tocris), 500 ng/ml human Noggin-CF chimera (R&D Systems) – inhibitors of WNT pathway, BMPs and SMAD, and 10 μM SB431542 (Tocris) – inhibitor of TGFβ signaling. Noggin and dorsomorphin supresses embryonic development thereby inducing neural differentiation pathways, while SB431542 mediates loss of pluripotency. Midbrain floorplate precursors were differentiated from all participants till day 11 using previously established protocols[19, 20].

Immunocytochemistry

Cultures were fixed in 4% formaldehyde followed by ice-cold 100% methanol and processed for immunofluorescence staining, confocal microscopy and high throughput imaging. Secondary antibodies used for primary antibody detection were species-specific Alexa-dye conjugates (Invitrogen). We used the following primary antibodies to Ki67 (Thermo Fisher PA5-16785), Nestin (Millipore MAB5326), Pax6 (BioLegend 901301), TBR1 (Abcam ab31940), MAP2 (Abcam ab92434), Emx1 (ThermoFisher PA5-35373), Gad67 (Abcam ab26116), Tuj1 (BioLegend 801201), CD44 (R&D Systems MAB7045), LMX1A (Abcam ab139726), FOXA2 (Invitrogen 701698). Quantification was performed on the Perkin Elmer Harmony Software v4.9, which is based on the CellProfiler high throughput image analysis system[60]. Cell nuclei were first identified based on DAPI staining. For nuclear proteins, only the nuclear area was selected. For cytoplasmic protein, the area around the nucleus was selected. Thresholds of fluorescent intensity was selected after background subtraction. Threshold for each probe remained unchanged for every sample imaged. Antibodies, dilutions used and fluorescence threshold information in **Supplementary Table S3**.

RNA isolation and sequencing

RNA was extracted using TRIzol (Thermo Fischer) and 1-bromo-3-chloropropane (BCP; Sigma). To remove genomic DNA during processing, turbo DNase (Thermo Fischer) was used. RNA concentration was quantified using Ribogreen assay (Invitrogen).

Starting with 500ng of total RNA, poly(A) containing mRNA was purified and libraries were prepared using TruSeq Stranded mRNA kit (Illumina). Unstranded libraries were constructed and underwent 50bp single ended sequencing on an Illumina HiSeq 2500 machine. To analyse iPSC mRNA-seq data, the raw reads were mapped to the human genome GRCh37.75 (UCSC version hg19) using STAR: RNA-seq aligner[29]. Aligned reads were sorted using samtools[61], while biases were removed using Picard tools (Broad Institute). Quality

control was performed using Picard tools (Broad Institute) and QoRTs [30] (**Supplementary Figure S9**). Gene expression levels were quantified using an union exon model with HTSeq [31], which uses uniquely aligned reads. Only the genes with >10 reads and expressed in 80% of the samples, were kept. The resulting read counts were log2 transformed and GC content, gene length, and library size normalised using the cqn package [62] in R.

mRNA weighted co-expression network analysis

Co-expression network analysis was performed using the R library, WGCNA[33]. We wanted to investigate autism-specific iPSC-neuronal culture co-expressed genes (or modules). Biweighted mid-correlations were calculated for all pairs of genes, then a signed similarity matrix was created. In the signed network, the similarity between genes reflects the sign of the correlation of their expression profiles. The signed similarity matrix was then raised to power β to emphasize strong correlations on an exponential scale. The resulting matrix (known as adjacency matrix) was then transformed into a topological overlap matrix. Since we are primarily interested in exploring co-expressed genes conserved across our cohort, we created consensus networks correlated to autism as previously published [14, 35]. After scaling for each individual network (consensus scaling quantile = 0.2), a soft thresholding power of 14 was chosen (as it was the smallest threshold that resulted in a scale-free R^2 fit of 0.8) (**Supplementary Figure S10**). The consensus network was created by using a topological overlap matrix (TOM) to calculate the component-wise minimum values for topological overlap. Using dissTOM = 1 – TOM as distance measure, genes were hierarchically clustered. Modules were then assigned using a dynamic tree-cutting algorithm (cutreeHybrid, using default parameters except deepSplit = 4, cutHeight = 0.999, minModulesize = 100, dthresh=0.1 and pamStage = FALSE).

Resulting modules of co-expressed genes were used to calculate module eigengenes (MEs; or 1st principal component of the module). MEs were correlated to biological traits, in this case autism, to find disease-specific modules. Module hubs were defined by calculating module membership (kME) values which are the Pearson correlation between each gene and corresponding ME, and genes with kME < 0.7 were removed from the module. Network visualisation was done using iGraph package in R [63].

Module preservation analysis

Module preservation analysis was performed to validate co-expression in a previous autism-iPSC study [22]. Module values from autism-iPSC network analysis were used as reference, to calculate the $Z_{summary}$ statistic for each module. This measure combines module density and intramodular connectivity metrics to give a composite statistic where Z > 2 suggests moderate preservation and Z > 10 suggests high preservation [64].

Transcription factor binding site enrichment

The top 200 genes in each module (ranked kME) were used for transcription factor binding site (TFBS) enrichment analysis (**Supplementary Figure S7C**).

Enrichment analysis for gene sets

Two types of gene set enrichments were performed. For autism-correlated module enrichment, logistic regression was performed using already published gene modules [14, 34, 35] to control for gene length and gene expression level. A two-sided Fisher exact test with 95% confidence interval was performed for cell-type enrichment analysis using published human brain dataset [65].

Module genes were characterised using GO Elite (version 1.2.5) [38] using total expressed genes as background. GO Elite uses a Z-score approximation of hypergeometric distribution to assess term enrichment, and removes redundant GO or KEGG terms to give a concise output. 10,000 permutations were used, and required at least 10 genes to be enriched in a given pathway at a Z-score of at least 2. Only biological process and molecular function categories are reported.

Quantitative real time polymerase chain reaction (QPCR)

cDNA was isolated, and QPCR was performed on RNA samples used for RNA-sequencing to show relative quantification of EMX1 (F: CCCTCTCCATTTCTACC, R: ACGTAGTGGTTCTTCTC), Gad67 (F: GTTACCGAGGAGCTAAA, R: CAATGACTCTGCTACTATTT) and GABRA4 (F: CATCTACTGACTTCTTTCTC, R: CATTCACTCATCCATTCC) (**Supplementary Figure S6C**). PowerUp SYBR Green Master Mix (Thermo Scientific) was used to run QPCR, on QuantStudio 7. ΔΔCt method was used to analyse differential expression.

Genomic DNA isolation and exome sequencing

Genomic DNA from autism and control iPSCs was isolated using Promega ReliaPrep™ gDNA Tissue Miniprep System (Promega). RNase A treatment was performed to digest contaminating RNA, and proteinase K to digest proteins. DNA concentration was quantified using Picogreen assay (Invitrogen).

10ng/µl of genomic DNA was used for library preparation and exome enrichment using Nextera Rapid Capture Exomes (Illumina). Paired end libraries were constructed and sequenced using Illumina HiSeq 2500 machine. Sample data was aligned to human genome GRCh37.75 (UCSC version hg19), using the Burrows Wheeler Aligner (BWA) [66]. Aligned

reads were sorted according to chromosome number. Duplicate reads usually created during sequencing were identified and removed using Picard tools (Broad Institute). Quality scores were assigned to individual bases, then adjusted to reduce systematic errors using genome analysis toolkit (GATK, Broad Institute). SNPs (single nucleotide polymorphisms) and indels (insertions-deletions) were identified using GATK, by local de-novo assembly of haplotypes (haploid genotype). The SNPs and indels were then recalibrated to check that they were true genetic variants and not artefacts. Next, these variants were evaluated for ratio of transition mutations to transversions mutations (Ti/Tv), heterozygous/homozygous (het:hom) ratio, and insertion/deletion (indel) ratio. SNPs and indels identified by GATK were annotated using ANNOVAR [67] and variant effector predictor (VEP) [68]. Important annotations include minor allele frequency from the 1000G project, SIFT score [69], PolyPhen 2 score [70], base change, and exonic function. Genotype concordance was performed using GATK to assess validity of SNPs from exome analysis (**Supplementary Figure S5D**).

Supplementary figure legends

Supplementary Figure S1: Quality control images of iPSC lines used in this study showing expression of stem cell markers NANOG and TRA-1-81.

Supplementary Figure S2: One-way ANOVA plots to demonstrate changes in variance of data points from individual participants. (A) Residuals vs Fitted values plots demonstrate spread of values of individual data points across all samples when observing dorsal vs ventral forebrain differentiation. (B) Residuals vs Fitted values plots when observing cortical differentiation. (C) F-values show degree of variance within the control and autism groups. All parameters measured show significant variance ($p < 0.05$) across both groups, and 7 out of 10 parameters show greater variance in the autism group.

Supplementary Figure S3: Trajectory of key developmental genes in syndromic vs non-syndromic vs control groups. Error bars: SD.

Supplementary Figure S4: RNA-sequencing sample preparation and bioinformatics pipeline.

Supplementary Figure S5: Differential gene expression analysis plots. (A) Volcano plot shows significant differentially expressed gene (in red). (B) Principal component analysis shows separate clustering of autism and control samples. (C) Linear modelling pipeline for differential gene expression used in this study showed significant correlation with DESeq. (D) High genotype concordance.

Supplementary Figure S6: Autism correlated gene expression modules with tissue differentiation and housekeeping roles. (A) Correlation network dendogram with consensus function of significant gene modules and their relationship with autism (red denotes a positive relationship). (B) Module eigengene values of correlated gene modules revealed significant differences, but magnitude of differences between groups is small. (C) D35 QPCR of EMX1, Gad67 and GABRA4 genes.

Supplementary Figure S7: Hub genes from significant gene modules and gene module preservation in autism iPSC RNA-seq studies. (A) Coexpression network plot and GO term enrichment of the ‘steelblue’, ‘lightgreen’, ‘white’, ‘salmon’, ‘seinna3’, ‘grey60’ modules. Top 25 genes are shown, with 10 hub genes depicted in the inner circle of each module. (B) Preservation of gene modules with modules from Mariani et al 2015 study using cerebral organoids from autistic individuals and with modules from Marchetto et al 2016 study using neuroprogenitor cells.

Supplementary Figure S8: Relationship of gene modules with exome variants and transcription factors driving each gene module. (A) Exome variants in top differentially expressed genes.

(B) Autism rare and unique variants enrichment (p-value in brackets). (C) Transcription factor binding site analysis of WGCNA gene modules.

Supplementary Figure S9: RNA sequencing quality control plots.

Supplementary Figure S10: Soft thresholding used for exploring co-expressed genes conserved across our cohort.

References

1. Berg JM, Geschwind DH. Autism genetics: searching for specificity and convergence. *Genome Biol* 2012; **13**(7)**:** 247.

2. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci* 2015; **16**(9)**:** 551-563.

3. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al.* Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**(7397)**:** 246-250.

4. APA. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub2013.

5. Minshew NJ, Keller TA. The nature of brain dysfunction in autism: functional brain imaging studies. *Curr Opin Neurol* 2010; **23**(2)**:** 124-130.

6. Samson F, Mottron L, Soulieres I, Zeffiro TA. Enhanced visual functioning in autism: an ALE meta-analysis. *Hum Brain Mapp* 2012; **33**(7)**:** 1553-1581.

7. Dichter GS. Functional magnetic resonance imaging of autism spectrum disorders. *Dialogues Clin Neurosci* 2012; **14**(3)**:** 319-351.

8. Philip RC, Dauvermann MR, Whalley HC, Baynham K, Lawrie SM, Stanfield AC. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neurosci Biobehav Rev* 2012; **36**(2)**:** 901-942.

9. Sztainberg Y, Zoghbi HY. Lessons learned from studying syndromic autism spectrum disorders. *Nat Neurosci* 2016; **19**(11)**:** 1408-1417.

10. Johnson CP, Myers SM, American Academy of Pediatrics Council on Children With D. Identification and evaluation of children with autism spectrum disorders. *Pediatrics* 2007; **120**(5)**:** 1183-1215.

11. Baron-Cohen S, Auyeung B, Norgaard-Pedersen B, Hougaard DM, Abdallah MW, Melgaard L *et al.* Elevated fetal steroidogenic activity in autism. *Mol Psychiatry* 2015; **20**(3)**:** 369-376.

12. Baron-Cohen S, Tsompanidis A, Auyeung B, Norgaard-Pedersen B, Hougaard DM, Abdallah M *et al.* Foetal oestrogens and autism. *Mol Psychiatry* 2019.

13. Schubert D, Martens GJ, Kolk SM. Molecular underpinnings of prefrontal cortex development in rodents provide insights into the etiology of neurodevelopmental disorders. *Mol Psychiatry* 2015; **20**(7)**:** 795-809.

14. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 2013; **155**(5)**:** 1008-1021.

15. Lancaster MA, Renner M, Martin CA, Wenzel D, Bicknell LS, Hurles ME *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* 2013; **501**(7467)**:** 373-379.

16. Pasca SP, Portmann T, Voineagu I, Yazawa M, Shcheglovitov A, Pasca AM *et al.* Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat Med* 2011; **17**(12)**:** 1657-1662.

17. Shi Y, Kirwan P, Livesey FJ. Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat Protoc* 2012; **7**(10)**:** 1836-1846.

18. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; **131**(5)**:** 861-872.

19. Fedele S, Collo G, Behr K, Bischofberger J, Muller S, Kunath T *et al.* Expansion of human midbrain floor plate progenitors from induced pluripotent stem cells increases dopaminergic neuron differentiation potential. *Sci Rep* 2017; **7**(1)**:** 6036.

20. Kriks S, Shim JW, Piao J, Ganat YM, Wakeman DR, Xie Z *et al.* Dopamine neurons derived from human ES cells efficiently engraft in animal models of Parkinson's disease. *Nature* 2011; **480**(7378)**:** 547-551.

21. Marchetto MC, Belinson H, Tian Y, Freitas BC, Fu C, Vadodaria KC *et al.* Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry* 2016.

22. Mariani J, Coppola G, Zhang P, Abyzov A, Provini L, Tomasini L *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* 2015; **162**(2)**:** 375-390.

23. Schafer ST, Paquola ACM, Stern S, Gosselin D, Ku M, Pena M *et al.* Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons. *Nat Neurosci* 2019; **22**(2)**:** 243-255.

24. Ajram LA, Horder J, Mendez MA, Galanopoulos A, Brennan LP, Wichers RH *et al.* Shifting brain inhibitory balance and connectivity of the prefrontal cortex of adults with autism spectrum disorder. *Transl Psychiatry* 2017; **7**(5)**:** e1137.

25. Horder J, Petrinovic MM, Mendez MA, Bruns A, Takumi T, Spooren W *et al.* Glutamate and GABA in autism spectrum disorder-a translational magnetic resonance spectroscopy study in man and rodent models. *Transl Psychiatry* 2018; **8**(1)**:** 106.

26. Stevenson RE, Schroer RJ, Skinner C, Fender D, Simensen RJ. Autism and macrocephaly. *Lancet* 1997; **349**(9067)**:** 1744-1745.

27. Aasen T, Izpisua Belmonte JC. Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells. *Nat Protoc* 2010; **5**(2)**:** 371-382.

28. Deans PJM, Raval P, Sellers KJ, Gatford NJF, Halai S, Duarte RRR *et al.* Psychosis Risk Candidate ZNF804A Localizes to Synapses and Regulates Neurite Formation and Dendritic Spine Structure. *Biological psychiatry* 2017; **82**(1)**:** 49-61.

29. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013; **29**(1)**:** 15-21.

30. Hartley SW, Mullikin JC. QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. *BMC Bioinformatics* 2015; **16:** 224.

31. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015; **31**(2)**:** 166-169.

32. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

33. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008; **9:** 559.

34. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 2011; **474**(7351)**:** 380-384.

35. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 2016.

36. Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat Neurosci* 2012; **15**(12)**:** 1723-1728.

37. Torkamani A, Dean B, Schork NJ, Thomas EA. Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res* 2010; **20**(4)**:** 403-412.

38. Zambon AC, Gaj S, Ho I, Hanspers K, Vranizan K, Evelo CT *et al.* GO-Elite: a flexible solution for pathway and ontology over-representation. *Bioinformatics* 2012; **28**(16)**:** 2209-2210.

39. Malmgren H, Sahlen S, Wide K, Lundvall M, Blennow E. Distal 3p deletion syndrome: detailed molecular cytogenetic and clinical characterization of three small distal deletions and review. *Am J Med Genet A* 2007; **143A**(18)**:** 2143-2149.

40. Marchetto MC, Carromeu C, Acab A, Yu D, Yeo GW, Mu Y *et al.* A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. *Cell* 2010; **143**(4)**:** 527-539.

41. Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W *et al.* Dissecting neural differentiation regulatory networks through epigenetic footprinting. *Nature* 2015; **518**(7539)**:** 355-359.

42. Tischfield MA, Baris HN, Wu C, Rudolph G, Van Maldergem L, He W *et al.* Human TUBB3 mutations perturb microtubule dynamics, kinesin interactions, and axon guidance. *Cell* 2010; **140**(1)**:** 74-87.

43. Stumpf PS, Smith RCG, Lenz M, Schuppert A, Muller FJ, Babtie A *et al.* Stem Cell Differentiation as a Non-Markov Stochastic Process. *Cell Syst* 2017; **5**(3)**:** 268-282 e267.

44. Salic A, Mitchison TJ. A chemical method for fast and sensitive detection of DNA synthesis in vivo. *Proc Natl Acad Sci U S A* 2008; **105**(7)**:** 2415-2420.

45. Soufi A, Dalton S. Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. *Development* 2016; **143**(23)**:** 4301-4311.

46. Dadalko OI, Travers BG. Evidence for Brainstem Contributions to Autism Spectrum Disorders. *Front Integr Neurosci* 2018; **12:** 47.

47. Rogers TD, McKimm E, Dickson PE, Goldowitz D, Blaha CD, Mittleman G. Is autism a disease of the cerebellum? An integration of clinical and pre-clinical research. *Front Syst Neurosci* 2013; **7:** 15.

48. Deng Q, Andersson E, Hedlund E, Alekseenko Z, Coppola E, Panman L *et al.* Specific and integrated roles of Lmx1a, Lmx1b and Phox2a in ventral midbrain development. *Development* 2011; **138**(16)**:** 3399-3408.

49. Ferri AL, Lin W, Mavromatakis YE, Wang JC, Sasaki H, Whitsett JA *et al.* Foxa1 and Foxa2 regulate multiple phases of midbrain dopaminergic neuron development in a dosage-dependent manner. *Development* 2007; **134**(15)**:** 2761-2769.

50. Gaspard N, Bouschet T, Hourez R, Dimidschstein J, Naeije G, van den Ameele J *et al.* An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* 2008; **455**(7211)**:** 351-357.

51. Corbin JG, Rutlin M, Gaiano N, Fishell G. Combinatorial function of the homeodomain proteins Nkx2.1 and Gsh2 in ventral telencephalic patterning. *Development* 2003; **130**(20)**:** 4895-4906.

52. Costa MR, Muller U. Specification of excitatory neurons in the developing cerebral cortex: progenitor diversity and environmental influences. *Front Cell Neurosci* 2014; **8:** 449.

53. Zhang W, Peterson M, Beyer B, Frankel WN, Zhang ZW. Loss of MeCP2 from forebrain excitatory neurons leads to cortical hyperexcitation and seizures. *J Neurosci* 2014; **34**(7)**:** 2754-2763.

54. Gorski JA, Talley T, Qiu M, Puelles L, Rubenstein JL, Jones KR. Cortical excitatory neurons and glia, but not GABAergic neurons, are produced in the Emx1-expressing lineage. *J Neurosci* 2002; **22**(15)**:** 6309-6314.

55. Lazarus MS, Krishnan K, Huang ZJ. GAD67 deficiency in parvalbumin interneurons produces deficits in inhibitory transmission and network disinhibition in mouse prefrontal cortex. *Cereb Cortex* 2015; **25**(5)**:** 1290-1296.

56. Azim E, Jabaudon D, Fame RM, Macklis JD. SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nat Neurosci* 2009; **12**(10)**:** 1238-1247.

57. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic Acids Res* 2009; **37**(Database issue)**:** D832-836.

58. Elsen GE, Bedogni F, Hodge RD, Bammler TK, MacDonald JW, Lindtner S *et al.* The Epigenetic Factor Landscape of Developing Neocortex Is Regulated by Transcription Factors Pax6--> Tbr2--> Tbr1. *Front Neurosci* 2018; **12:** 571.

59. Bedogni F, Hodge RD, Elsen GE, Nelson BR, Daza RA, Beyer RP *et al.* Tbr1 regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc Natl Acad Sci U S A* 2010; **107**(29)**:** 13129-13134.

60. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006; **7**(10)**:** R100.

61. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**(16)**:** 2078-2079.

62. Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 2012; **13**(2)**:** 204-216.

63. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006; **1695**(5)**:** 1-9.

64. Langfelder P, Luo R, Oldham MC, Horvath S. Is my network module preserved and reproducible? *PLoS Comput Biol* 2011; **7**(1)**:** e1001057.

65. Zhang Y, Sloan SA, Clarke LE, Caneda C, Plaza CA, Blumenthal PD *et al.* Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* 2016; **89**(1)**:** 37-53.

66. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**(14)**:** 1754-1760.

67. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* 2015; **10**(10)**:** 1556-1566.

68. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 2010; **26**(16)**:** 2069-2070.

69. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003; **31**(13)**:** 3812-3814.

70. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* 2002; **30**(17)**:** 3894-3900.

Figure 1

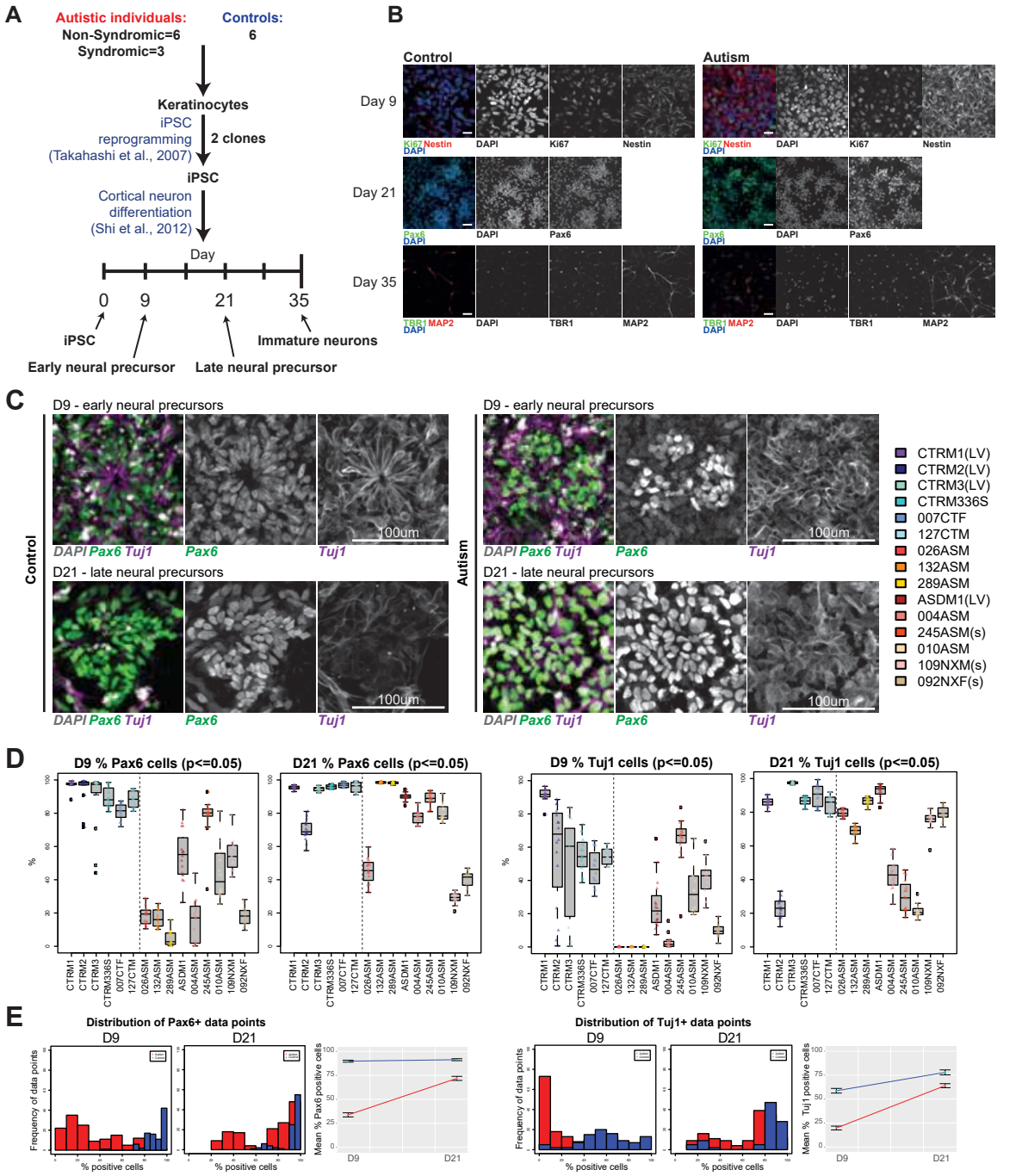

A
Autistic individuals:
Non-Syndromic=6
Syndromic=3
Controls:
6
Keratinocytes
iPSC reprogramming (Takahashi et al., 2007)
2 clones
iPSC
Cortical neuron differentiation (Shi et al., 2012)
Day
0
9
21
35
iPSC
Early neural precursor
Late neural precursor
Immature neurons
B
Control
Autism
Day 9
Day 21
Day 35
Ki67 Nestin DAPI
DAPI
Ki67
Nestin
Pax6 DAPI
Pax6
TBR1 MAP2 DAPI
TBR1
MAP2
C
D9 - early neural precursors
D21 - late neural precursors
Control
Autism
DAPI Pax6 Tuj1
Pax6
Tuj1
100um
CTRM1(LV)
CTRM2(LV)
CTRM3(LV)
CTRM336S
007CTF
127CTM
026ASM
132ASM
289ASM
ASDM1(LV)
004ASM
245ASM(s)
010ASM
109NXM(s)
092NXF(s)
D
D9 % Pax6 cells (p<=0.05)
D21 % Pax6 cells (p<=0.05)
D9 % Tuj1 cells (p<=0.05)
D21 % Tuj1 cells (p<=0.05)
%
CTRM1
CTRM2
CTRM3
CTRM336S
007CTF
127CTM
026ASM
132ASM
289ASM
ASDM1
004ASM
245ASM
010ASM
109NXM
092NXF
E
Distribution of Pax6+ data points
Distribution of Tuj1+ data points
D9
D21
Frequency of data points
% positive cells
Mean % Pax6 positive cells
Mean % Tuj1 positive cells

Figure 2

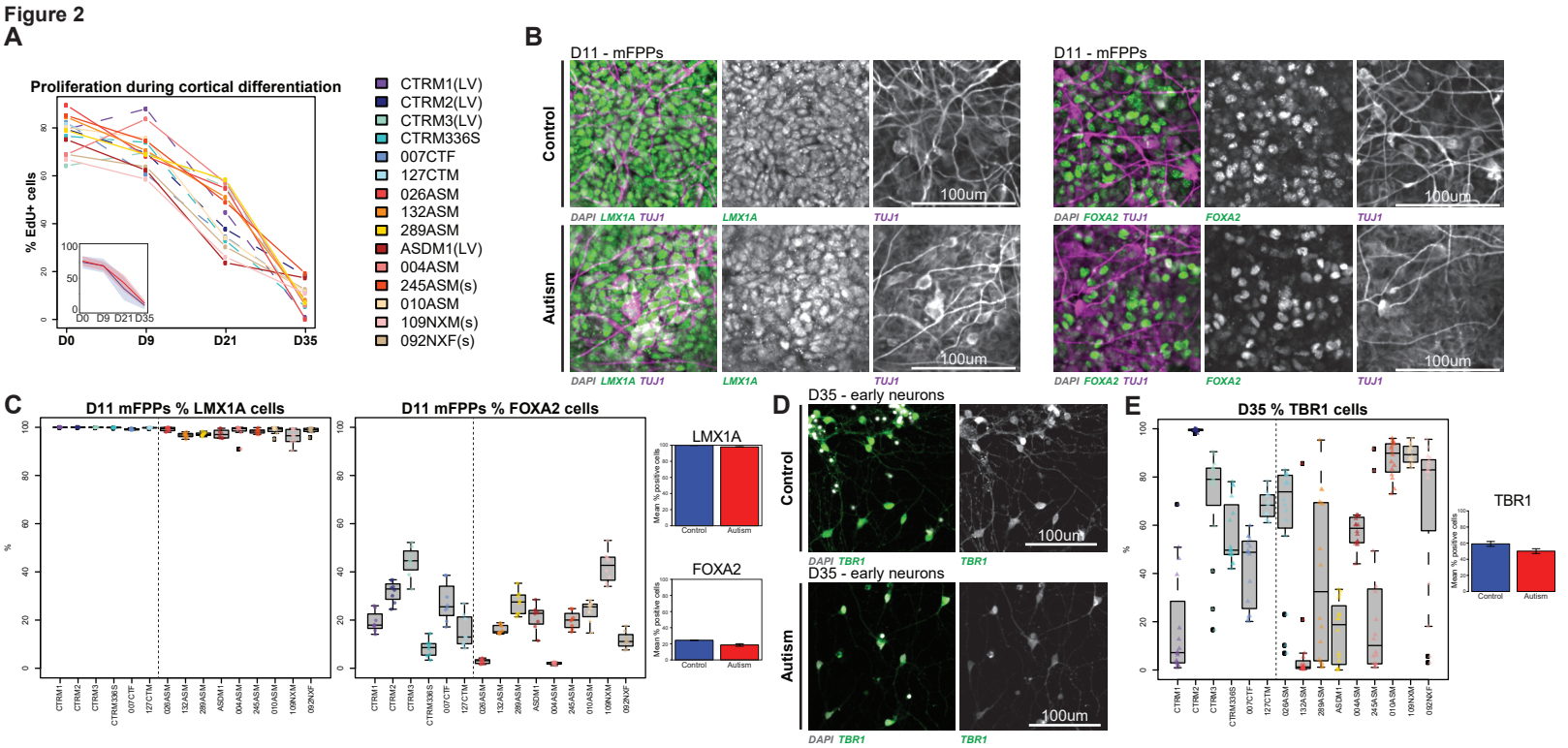

A
Proliferation during cortical differentiation
% EdU+ cells
D0
D9
D21
D35
CTRM1(LV)
CTRM2(LV)
CTRM3(LV)
CTRM336S
007CTF
127CTM
026ASM
132ASM
289ASM
ASDM1(LV)
004ASM
245ASM(s)
010ASM
109NXM(s)
092NXF(s)
B
D11 - mFPPs
Control
Autism
DAPI LMX1A TUJ1
LMX1A
TUJ1
DAPI FOXA2 TUJ1
FOXA2
100um
C
D11 mFPPs % LMX1A cells
D11 mFPPs % FOXA2 cells
LMX1A
FOXA2
Mean % positive cells
Control
Autism
D
D35 - early neurons
DAPI TBR1
TBR1
E
D35 % TBR1 cells
TBR1

Figure 3

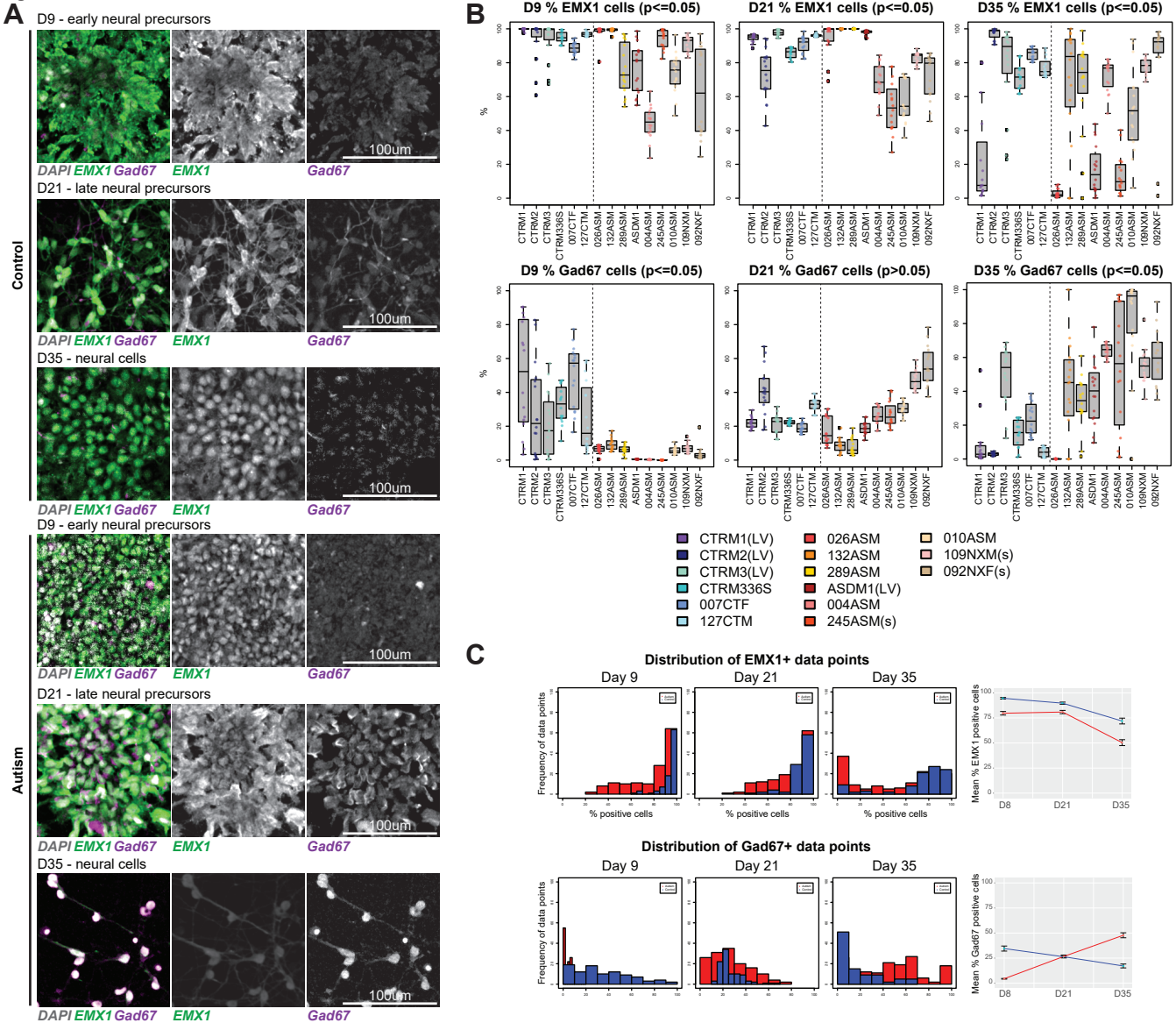

A
Control
Autism
D9 - early neural precursors
D21 - late neural precursors
D35 - neural cells
DAPI EMX1 Gad67
EMX1
Gad67
100um
B
D9 % EMX1 cells (p<=0.05)
D21 % EMX1 cells (p<=0.05)
D35 % EMX1 cells (p<=0.05)
D9 % Gad67 cells (p<=0.05)
D21 % Gad67 cells (p>0.05)
D35 % Gad67 cells (p<=0.05)
%
CTRM1
CTRM2
CTRM3
CTRM336S
007CTF
127CTM
026ASM
132ASM
289ASM
ASDM1
004ASM
245ASM
010ASM
109NXM
092NXF
CTRM1(LV)
CTRM2(LV)
CTRM3(LV)
CTRM336S
007CTF
127CTM
026ASM
132ASM
289ASM
ASDM1(LV)
004ASM
245ASM(s)
010ASM
109NXM(s)
092NXF(s)
C
Distribution of EMX1+ data points
Day 9
Day 21
Day 35
Frequency of data points
% positive cells
Mean % EMX1 positive cells
D8
D21
D35
Distribution of Gad67+ data points
Mean % Gad67 positive cells

Figure 4

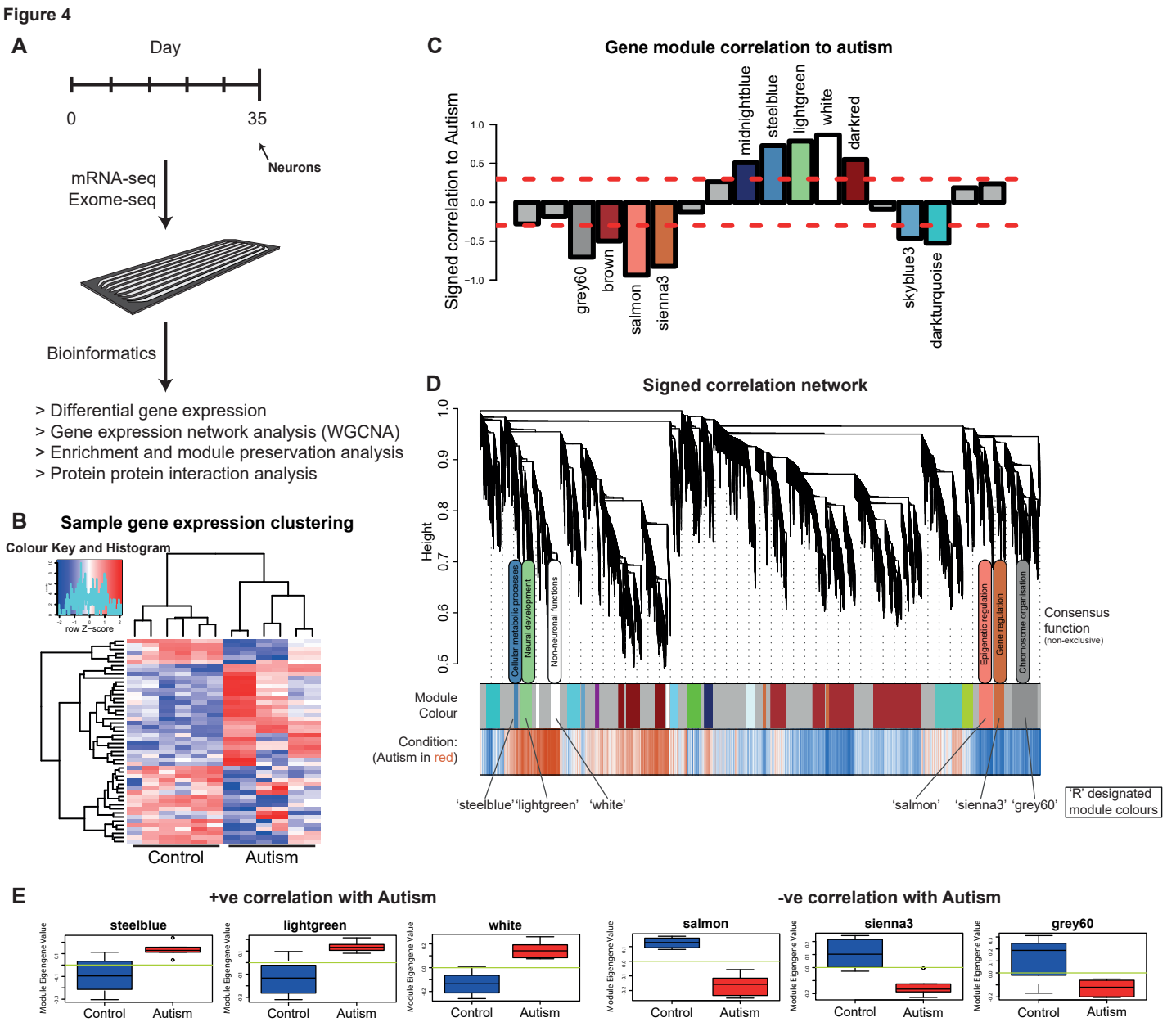

A
Day
0
35
Neurons
mRNA-seq
Exome-seq
Bioinformatics
> Differential gene expression
> Gene expression network analysis (WGCNA)
> Enrichment and module preservation analysis
> Protein protein interaction analysis
B
Sample gene expression clustering
Colour Key and Histogram
row Z-score
Control
Autism
C
Gene module correlation to autism
Signed correlation to Autism
1.0
0.5
0.0
−0.5
−1.0
midnightblue
steelblue
lightgreen
white
darkred
grey60
brown
salmon
sienna3
skyblue3
darkturquoise
D
Signed correlation network
Height
1.0
0.9
0.8
0.7
0.6
0.5
Cellular metabolic processes
Neural development
Non-neuronal functions
Epigenetic regulation
Gene regulation
Chromosome organisation
Consensus function
(non-exclusive)
Module Colour
Condition:
(Autism in red)
'steelblue' 'lightgreen' 'white'
'salmon' 'sienna3' 'grey60'
'R' designated module colours
E
+ve correlation with Autism
-ve correlation with Autism
steelblue
lightgreen
white
salmon
sienna3
grey60
Module Eigengene Value
Control
Autism

Figure 5

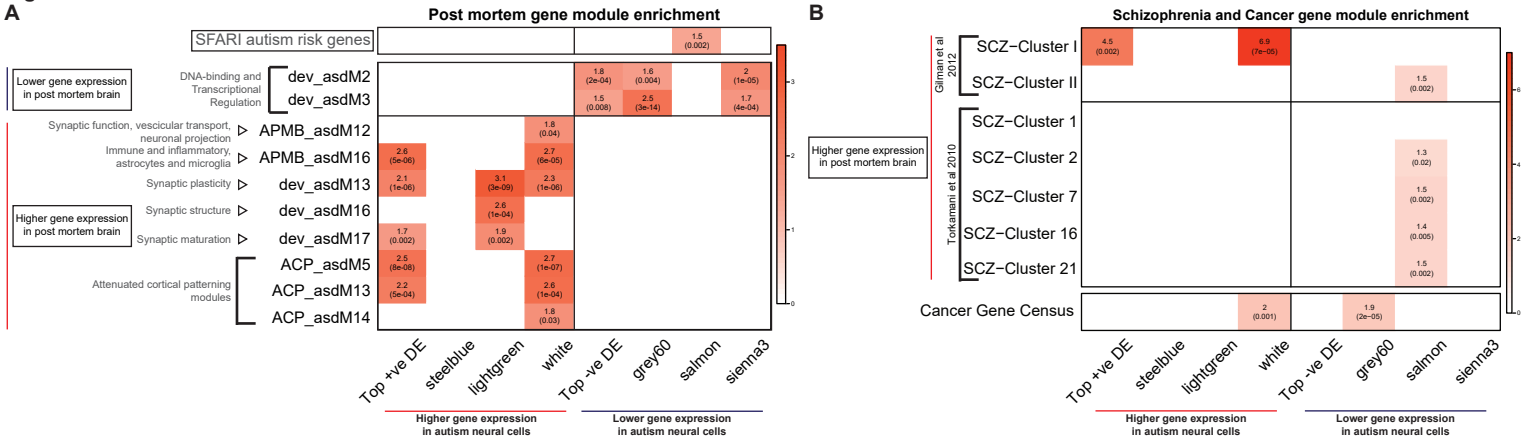
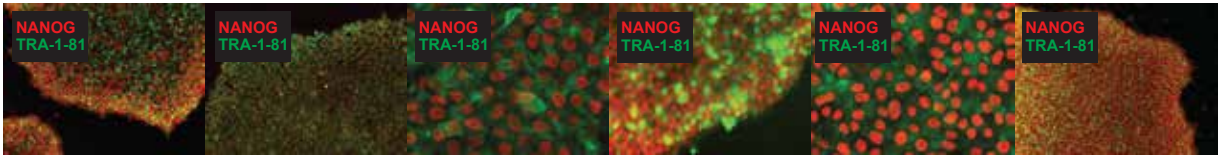

A
Post mortem gene module enrichment
SFARI autism risk genes
1.5 (0.002)
Lower gene expression in post mortem brain
DNA-binding and Transcriptional Regulation
dev_asdM2
dev_asdM3
1.8 (2e-04)
1.6 (0.004)
2 (1e-05)
1.5 (0.008)
2.5 (3e-14)
1.7 (4e-04)
Synaptic function, vescicular transport, neuronal projection
APMB_asdM12
1.8 (0.04)
Immune and inflammatory, astrocytes and microglia
APMB_asdM16
2.6 (5e-06)
2.7 (6e-05)
Synaptic plasticity
dev_asdM13
2.1 (1e-06)
3.1 (3e-09)
2.3 (1e-06)
Higher gene expression in post mortem brain
Synaptic structure
dev_asdM16
2.6 (1e-04)
Synaptic maturation
dev_asdM17
1.7 (0.002)
1.9 (0.002)
Attenuated cortical patterning modules
ACP_asdM5
2.5 (8e-08)
2.7 (1e-07)
ACP_asdM13
2.2 (5e-04)
2.6 (1e-04)
ACP_asdM14
1.8 (0.03)
Top +ve DE
steelblue
lightgreen
white
Top -ve DE
grey60
salmon
sienna3
Higher gene expression in autism neural cells
Lower gene expression in autism neural cells
B
Schizophrenia and Cancer gene module enrichment
Gilman et al 2012
SCZ−Cluster I
4.5 (0.002)
6.9 (7e−05)
SCZ−Cluster II
1.5 (0.002)
Higher gene expression in post mortem brain
Torkamani et al 2010
SCZ−Cluster 1
SCZ−Cluster 2
1.3 (0.02)
SCZ−Cluster 7
1.5 (0.002)
SCZ−Cluster 16
1.4 (0.005)
SCZ−Cluster 21
1.5 (0.002)
Cancer Gene Census
2 (0.001)
1.9 (2e−05)
Top +ve DE
steelblue
lightgreen
white
Top -ve DE
grey60
salmon
sienna3
Higher gene expression in autism neural cells
Lower gene expression in autism neural cells

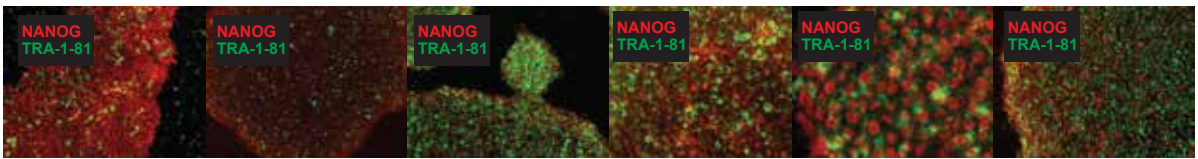
Figure S1

A Control iPSC lines

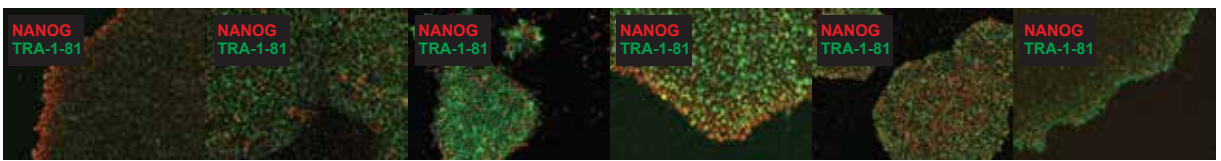


CTR.M1.04 CTR.M3.22 CTR.M2.42 CTR.M1.11 CTR.M3.15 CTR.M2.05

B Autism iPSC lines

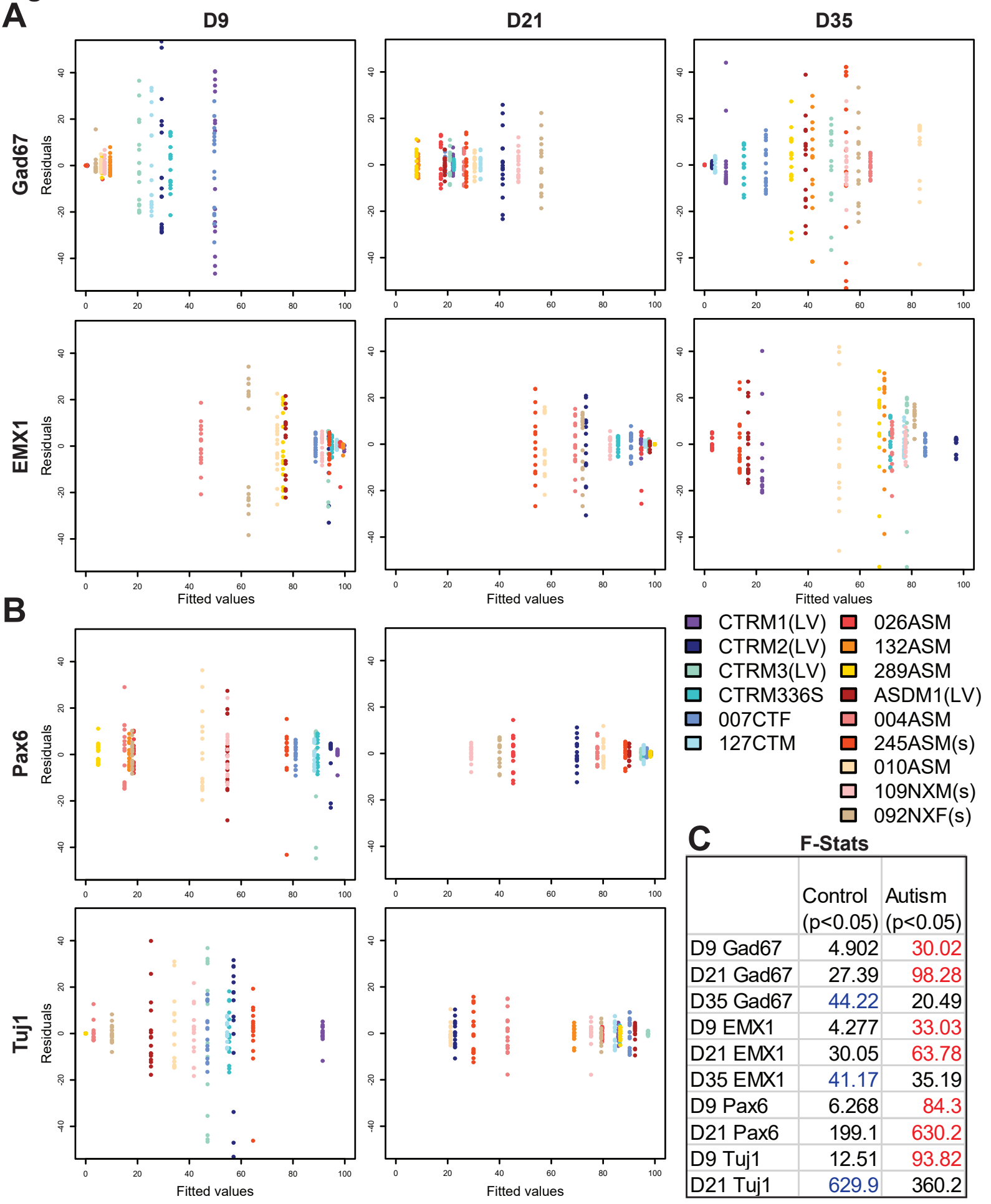


010.ASM.06 ASD.M1.08 004.ASM.09 010.ASM.05 ASD.M1.01 004.ASM.01



026.ASM.03 289.ASM.03 245.ASM.02 132.ASM.02 109.NXM.04 092.NXF.04

Figure S2



C

F-Stats

	Control (p<0.05)	Autism (p<0.05)
D9 Gad67	4.902	30.02
D21 Gad67	27.39	98.28
D35 Gad67	44.22	20.49
D9 EMX1	4.277	33.03
D21 EMX1	30.05	63.78
D35 EMX1	41.17	35.19
D9 Pax6	6.268	84.3
D21 Pax6	199.1	630.2
D9 Tuj1	12.51	93.82
D21 Tuj1	629.9	360.2

Figure S3

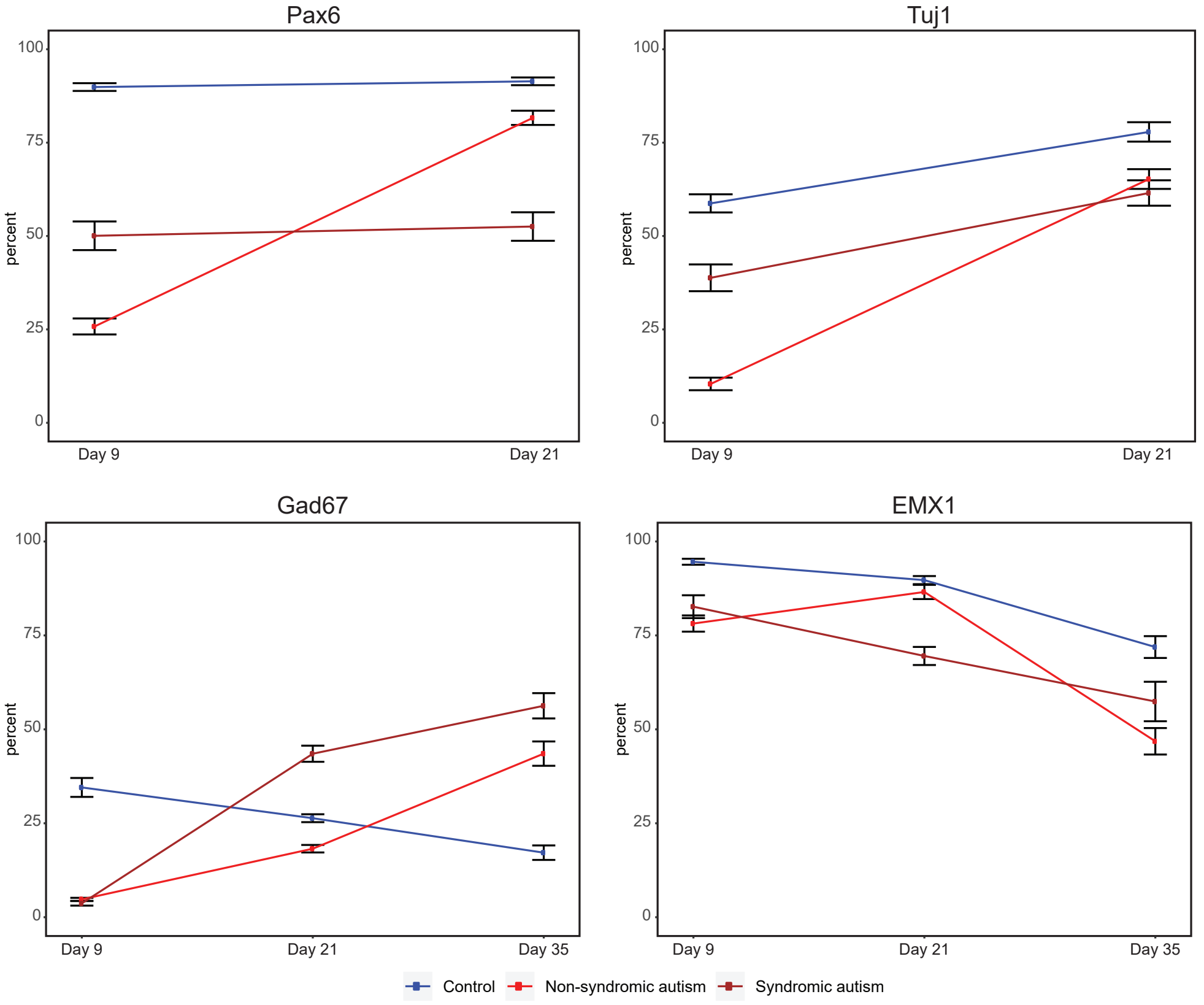

Pax6
Tuj1
Gad67
EMX1
percent
100
75
50
25
0
Day 9
Day 21
Day 35
Control
Non-syndromic autism
Syndromic autism

Figure S4

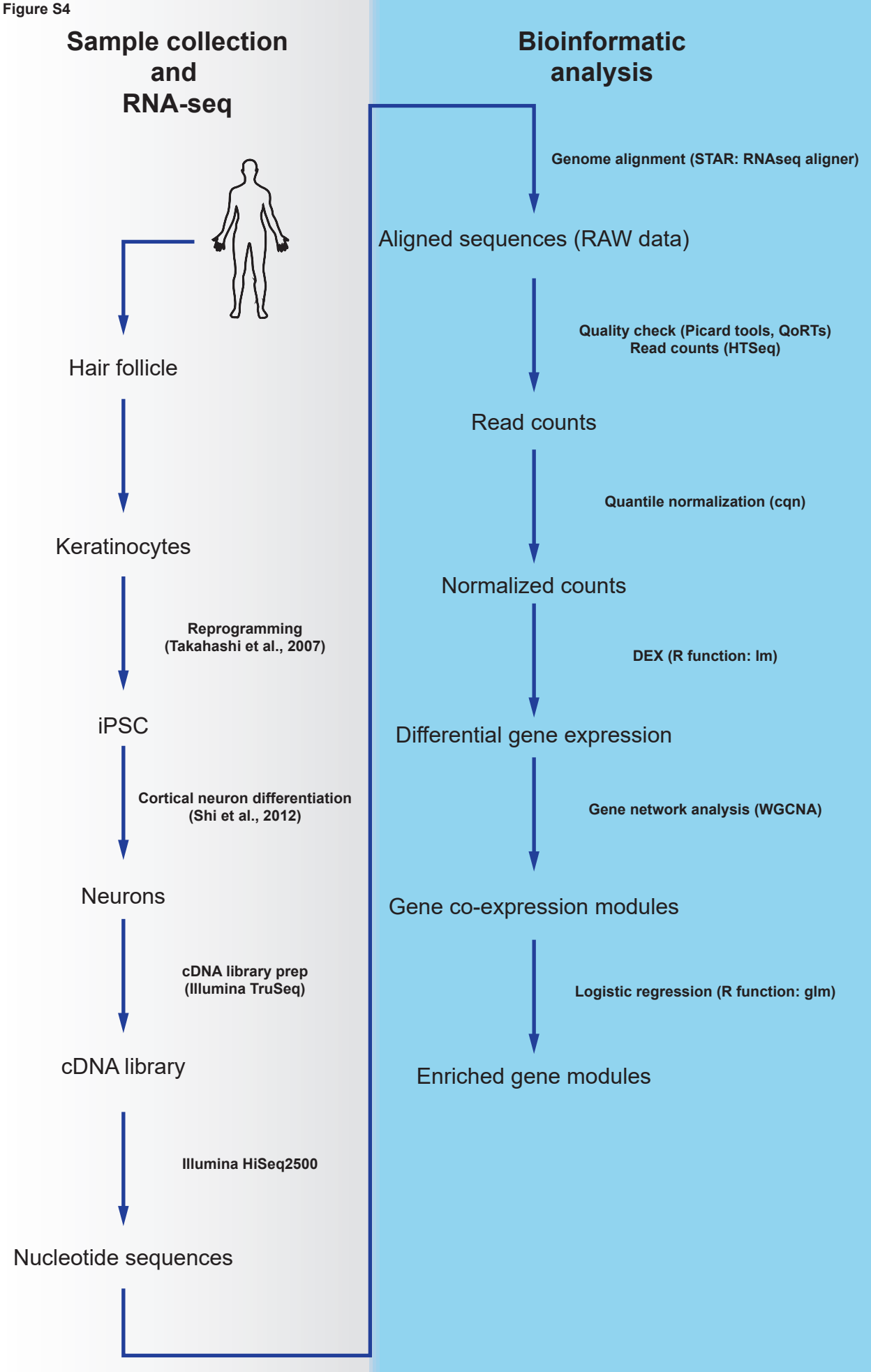

Sample collection and RNA-seq
Bioinformatic analysis
Hair follicle
Keratinocytes
Reprogramming (Takahashi et al., 2007)
iPSC
Cortical neuron differentiation (Shi et al., 2012)
Neurons
cDNA library prep (Illumina TruSeq)
cDNA library
Illumina HiSeq2500
Nucleotide sequences
Genome alignment (STAR: RNAseq aligner)
Aligned sequences (RAW data)
Quality check (Picard tools, QoRTs)
Read counts (HTSeq)
Read counts
Quantile normalization (cqn)
Normalized counts
DEX (R function: lm)
Differential gene expression
Gene network analysis (WGCNA)
Gene co-expression modules
Logistic regression (R function: glm)
Enriched gene modules

Figure S5

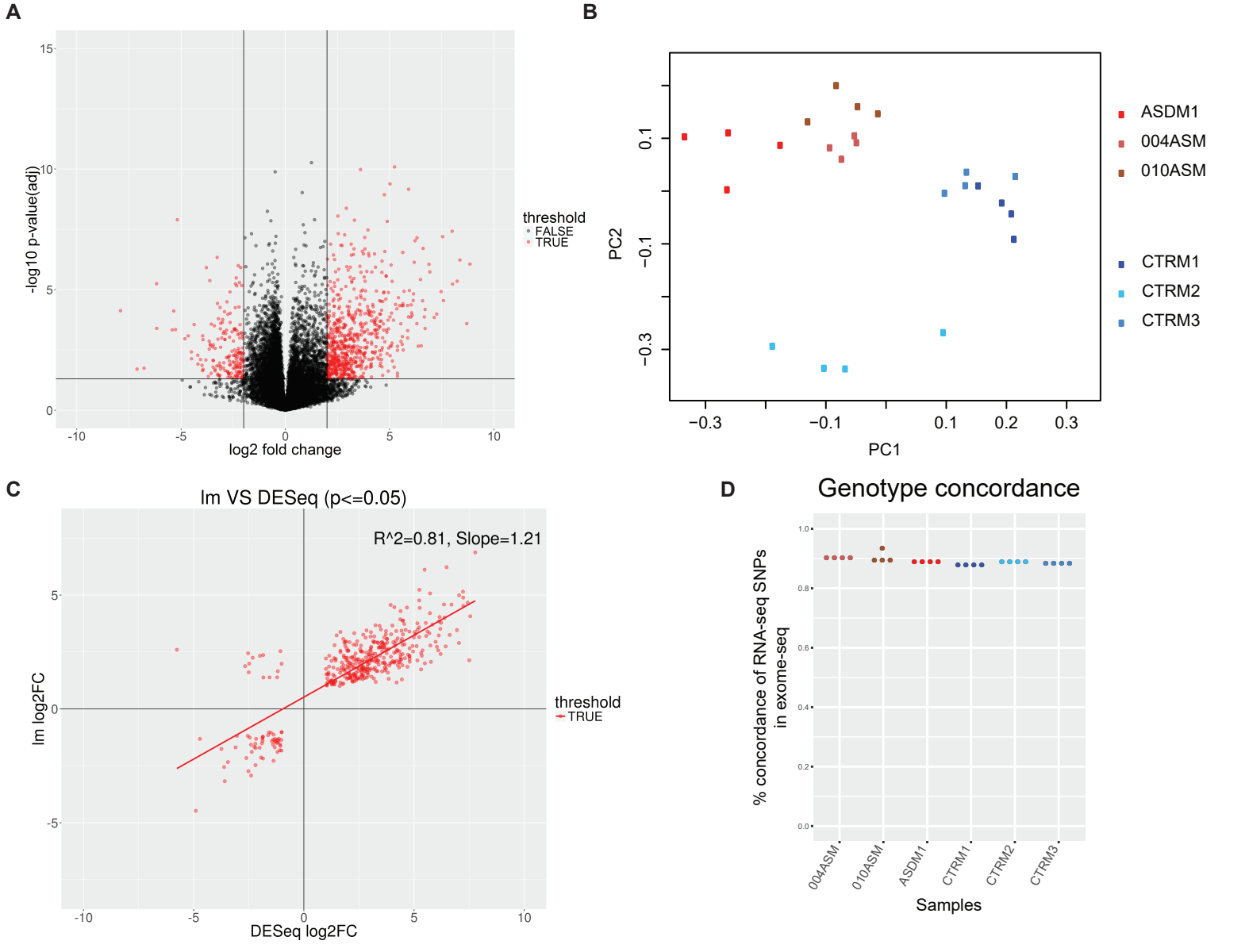

A
-log10 p-value(adj)
log2 fold change
threshold
FALSE
TRUE
B
PC2
PC1
ASDM1
004ASM
010ASM
CTRM1
CTRM2
CTRM3
C
lm VS DESeq (p<=0.05)
R^2=0.81, Slope=1.21
lm log2FC
DESeq log2FC
threshold
TRUE
D
Genotype concordance
% concordance of RNA-seq SNPs in exome-seq
Samples
004ASM
010ASM
ASDM1
CTRM1
CTRM2
CTRM3

Figure S6

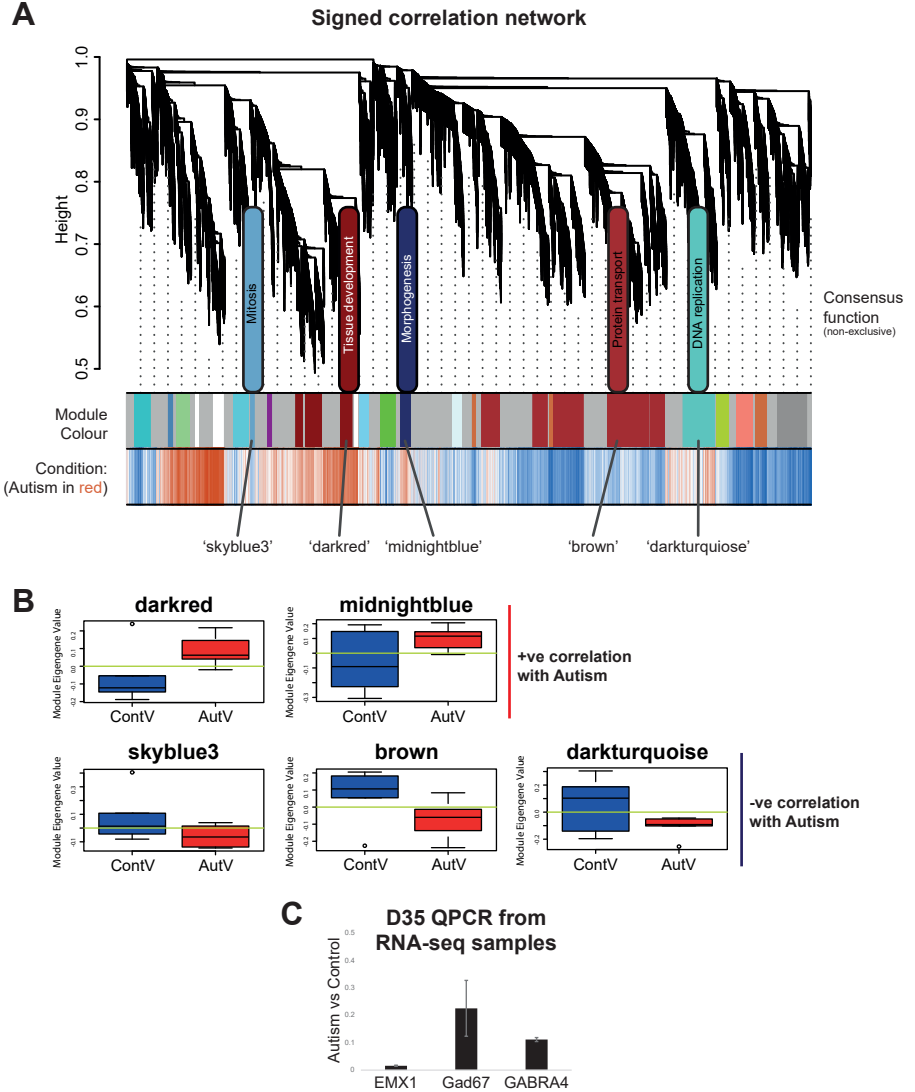

A
Signed correlation network
Height
1.0
0.9
0.8
0.7
0.6
0.5
Mitosis
Tissue development
Morphogenesis
Protein transport
DNA replication
Consensus function
(non-exclusive)
Module Colour
Condition:
(Autism in red)
'skyblue3'
'darkred'
'midnightblue'
'brown'
'darkturquiose'
B
darkred
midnightblue
skyblue3
brown
darkturquoise
Module Eigengene Value
ContV
AutV
+ve correlation with Autism
-ve correlation with Autism
C
D35 QPCR from RNA-seq samples
Autism vs Control
EMX1
Gad67
GABRA4

Figure S7

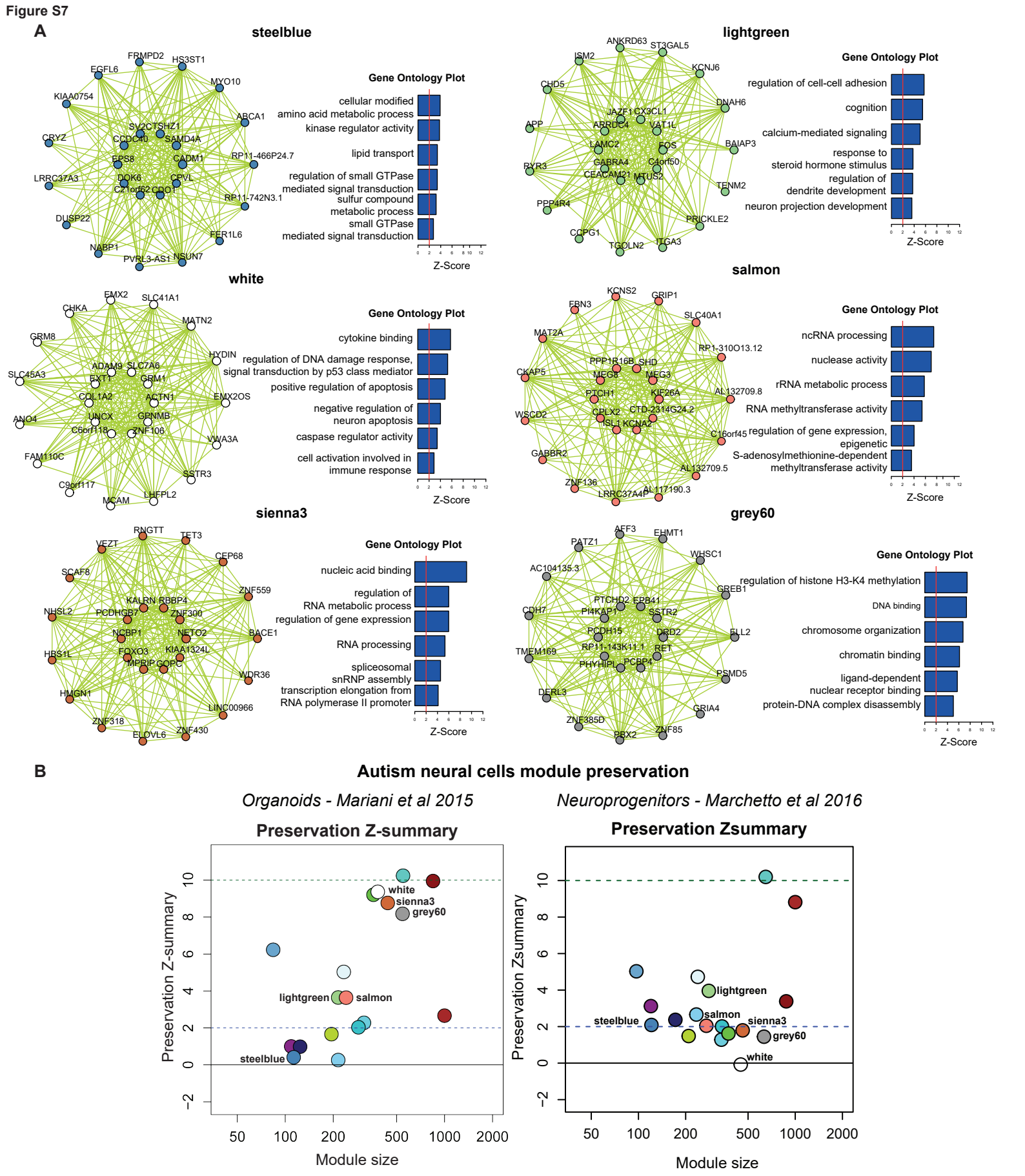
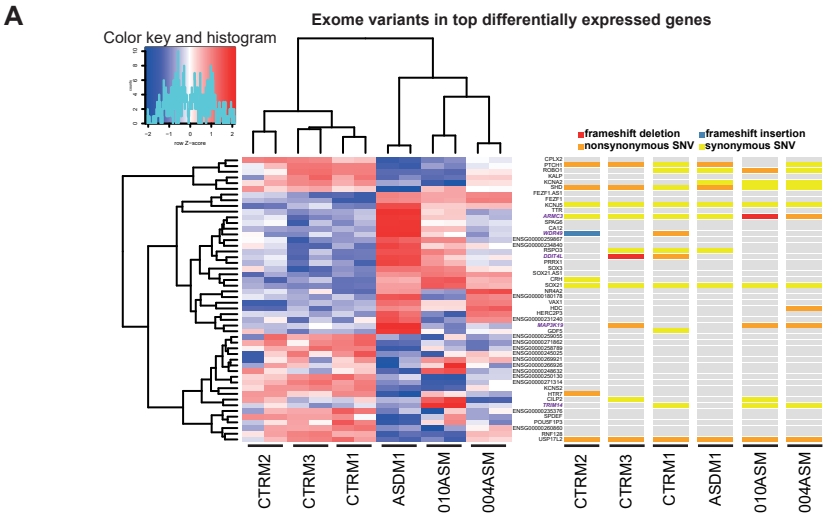
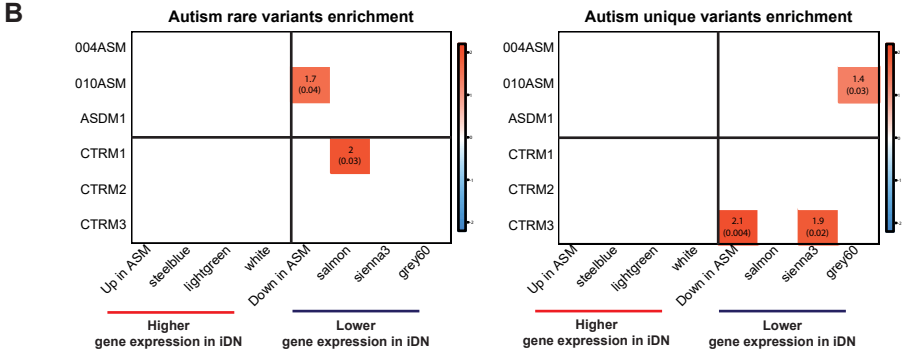

A
steelblue
FRMPD2
HS3ST1
EGFL6
MYO10
KIAA0754
ABCA1
SV2CTSHZ1
CRYZ
CCDC40
SAMD4A
EPS8
CADM1
RP11-466P24.7
LRRC37A3
DOK6
CPVL
C21orf62
CDO1
RP11-742N3.1
DUSP22
FER1L6
NABP1
PVRL3-AS1
NSUN7
Gene Ontology Plot
cellular modified amino acid metabolic process
kinase regulator activity
lipid transport
regulation of small GTPase mediated signal transduction
sulfur compound metabolic process
small GTPase mediated signal transduction
Z-Score
lightgreen
ANKRD63
ST3GAL5
ISM2
KCNJ6
CHD5
DNAH6
JAZF1
CX3CL1
APP
ARRDC4
VAT1L
LAMC2
FOS
BAIAP3
RYR3
GABRA4
C4orf50
CEACAM21
MTUS2
TENM2
PPP4R4
PRICKLE2
CCPG1
TGOLN2
ITGA3
Gene Ontology Plot
regulation of cell-cell adhesion
cognition
calcium-mediated signaling
response to steroid hormone stimulus
regulation of dendrite development
neuron projection development
Z-Score
white
EMX2
SLC41A1
CHKA
MATN2
GRM8
HYDIN
ADAM9
SLC7A6
SLC45A3
EXT1
GRM1
COL1A2
ACTN1
EMX2OS
ANO4
UNCX
GPNMB
C6orf118
ZNF106
VWA3A
FAM110C
SSTR3
C9orf117
MCAM
LHFPL2
Gene Ontology Plot
cytokine binding
regulation of DNA damage response, signal transduction by p53 class mediator
positive regulation of apoptosis
negative regulation of neuron apoptosis
caspase regulator activity
cell activation involved in immune response
Z-Score
salmon
KCNS2
GRIP1
FBN3
SLC40A1
MAT2A
RP1-310O13.12
PPP1R16B
SHD
CKAP5
MEG8
MEG3
PTCH1
KIF26A
AL132709.8
WSCD2
CPLX2
CTD-2314G24.2
SL1
KCNA2
C16orf45
GABBR2
AL132709.5
ZNF136
LRRC37A4P
AL117190.3
Gene Ontology Plot
ncRNA processing
nuclease activity
rRNA metabolic process
RNA methyltransferase activity
regulation of gene expression, epigenetic
S-adenosylmethionine-dependent methyltransferase activity
Z-Score
sienna3
RNGTT
TET3
VEZT
CEP68
SCAF8
ZNF559
KALRN
RBBP4
NHSL2
PCDHGB7
ZNF300
NCBP1
NETO2
BACE1
HBS1L
FOXO3
KIAA1324L
MPRIP
GOPC
WDR36
HMGN1
LINC00966
ZNF318
ZNF430
ELOVL6
Gene Ontology Plot
nucleic acid binding
regulation of RNA metabolic process
regulation of gene expression
RNA processing
spliceosomal snRNP assembly
transcription elongation from RNA polymerase II promoter
Z-Score
grey60
AFF3
EHMT1
PATZ1
WHSC1
AC104135.3
GREB1
PTCHD2
EPB41
CDH7
PI4KAP1
SSTR2
PCDH15
DRD2
ELL2
TMEM169
RP11-143K11.1
RET
PHYHIPL
PCBP4
PSMD5
DERL3
GRIA4
ZNF385D
ZNF85
PEX2
Gene Ontology Plot
regulation of histone H3-K4 methylation
DNA binding
chromosome organization
chromatin binding
ligand-dependent nuclear receptor binding
protein-DNA complex disassembly
Z-Score
B
Autism neural cells module preservation
Organoids - Mariani et al 2015
Preservation Z-summary
Preservation Z-summary
Module size
white
sienna3
grey60
lightgreen
salmon
steelblue
Neuroprogenitors - Marchetto et al 2016
Preservation Zsummary
Preservation Zsummary
Module size
lightgreen
salmon
steelblue
sienna3
grey60
white

Figure S8

A



B



C

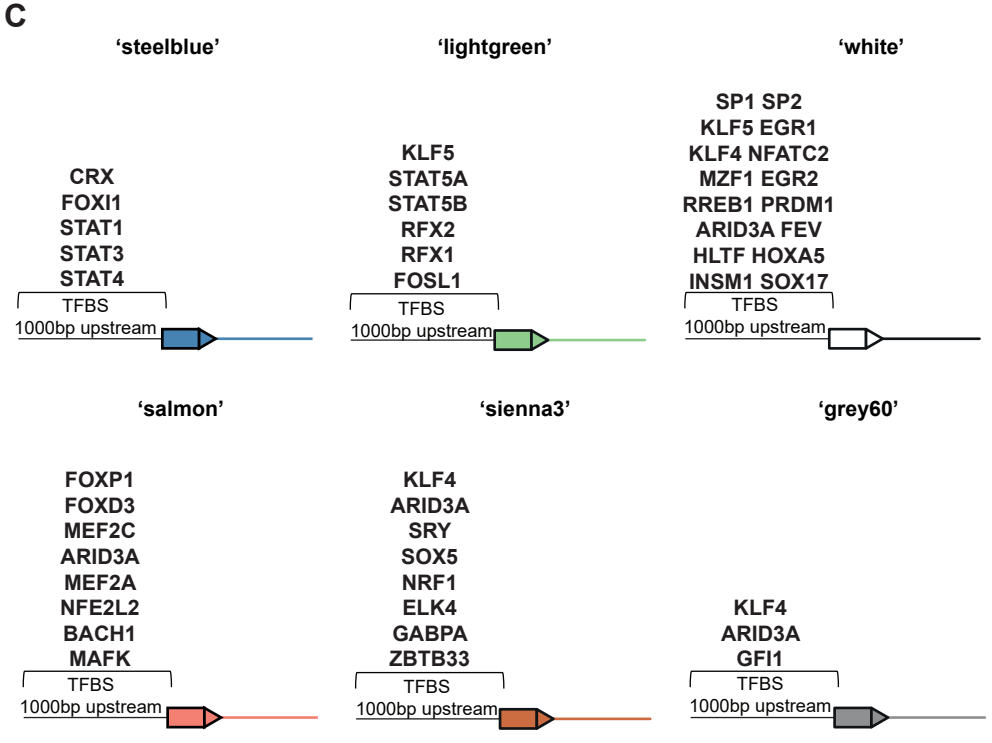


Figure S9

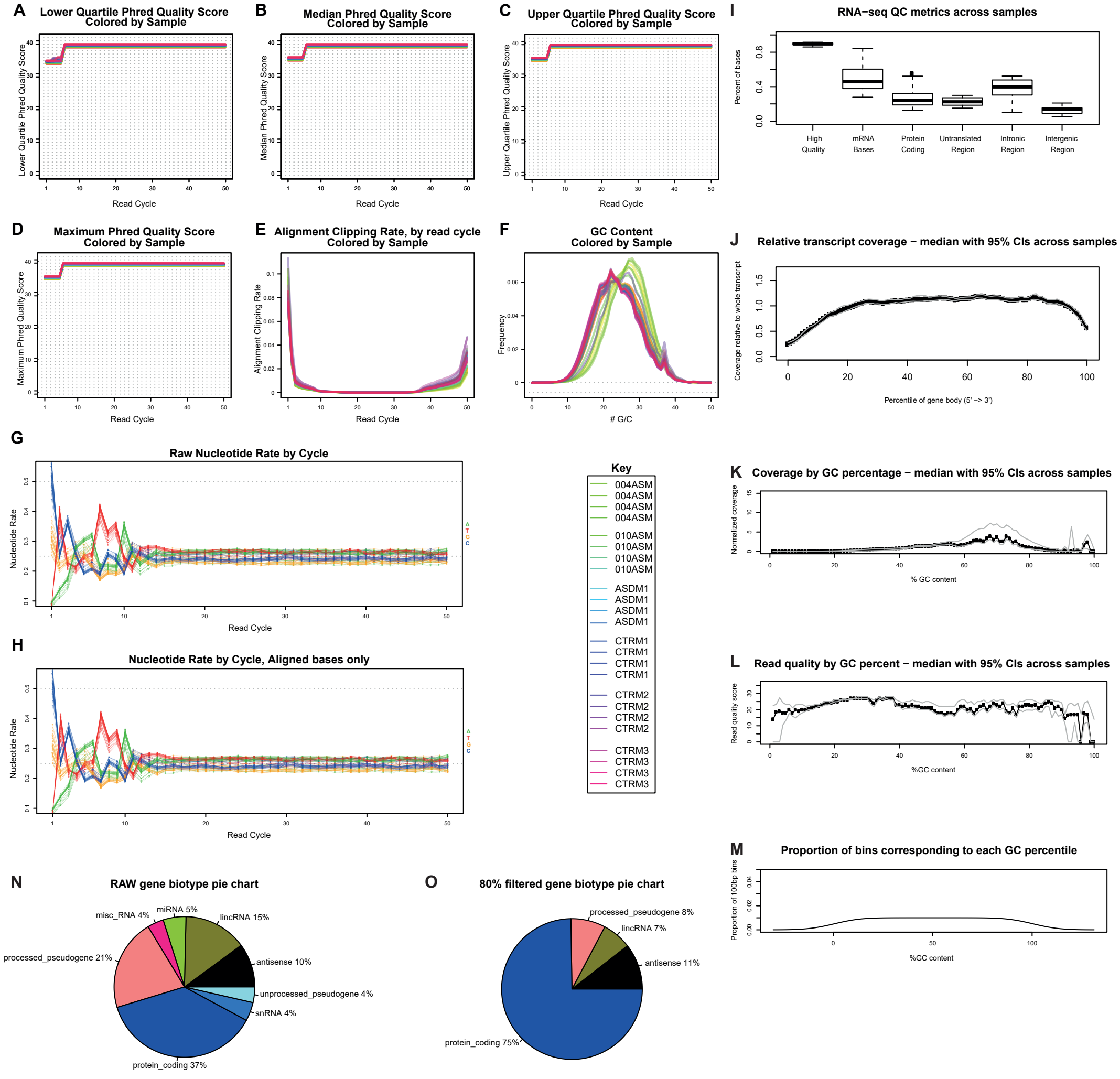
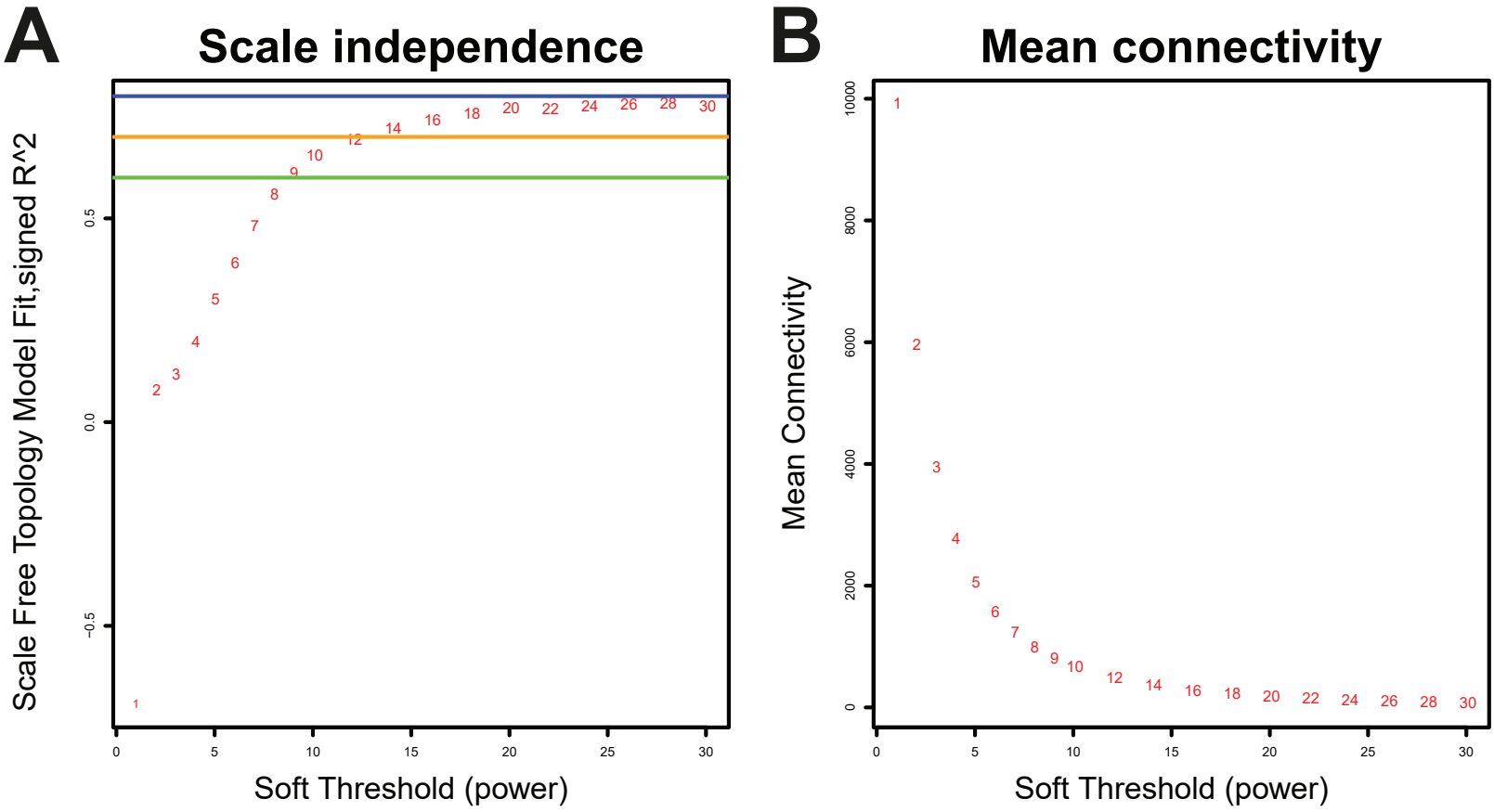

A
Lower Quartile Phred Quality Score Colored by Sample
Lower Quartile Phred Quality Score
Read Cycle
B
Median Phred Quality Score Colored by Sample
Median Phred Quality Score
Read Cycle
C
Upper Quartile Phred Quality Score Colored by Sample
Upper Quartile Phred Quality Score
Read Cycle
I
RNA−seq QC metrics across samples
Percent of bases
High Quality
mRNA Bases
Protein Coding
Untranslated Region
Intronic Region
Intergenic Region
D
Maximum Phred Quality Score Colored by Sample
Maximum Phred Quality Score
Read Cycle
E
Alignment Clipping Rate, by read cycle Colored by Sample
Alignment Clipping Rate
Read Cycle
F
GC Content Colored by Sample
Frequency
G/C
J
Relative transcript coverage − median with 95% CIs across samples
Coverage relative to whole transcript
Percentile of gene body (5' −> 3')
G
Raw Nucleotide Rate by Cycle
Nucleotide Rate
Read Cycle
A
T
G
C
Key
004ASM
004ASM
004ASM
004ASM
010ASM
010ASM
010ASM
010ASM
ASDM1
ASDM1
ASDM1
ASDM1
CTRM1
CTRM1
CTRM1
CTRM1
CTRM2
CTRM2
CTRM2
CTRM2
CTRM3
CTRM3
CTRM3
CTRM3
K
Coverage by GC percentage − median with 95% CIs across samples
Normalized coverage
% GC content
H
Nucleotide Rate by Cycle, Aligned bases only
Nucleotide Rate
Read Cycle
A
T
G
C
L
Read quality by GC percent − median with 95% CIs across samples
Read quality score
%GC content
M
Proportion of bins corresponding to each GC percentile
Proportion of 100bp bins
%GC content
N
RAW gene biotype pie chart
misc_RNA 4%
miRNA 5%
lincRNA 15%
processed_pseudogene 21%
antisense 10%
unprocessed_pseudogene 4%
snRNA 4%
protein_coding 37%
O
80% filtered gene biotype pie chart
processed_pseudogene 8%
lincRNA 7%
antisense 11%
protein_coding 75%

Figure S10


A
Scale independence
Scale Free Topology Model Fit,signed R^2
Soft Threshold (power)
B
Mean connectivity
Mean Connectivity
Soft Threshold (power)