Shared gene co-expression networks in autism from induced pluripotent stem cell (iPSC) neurons

Dwaipayan Adhya[1,2,*], Vivek Swarup[3,*], Paulina Nowosiad[2], Carole Shum[2], Kamila Jozwik[4], Grainne McAlonan[5], Maria Andreina Mendez[5], Jamie Horder[5], Declan Murphy[5], Daniel H. Geschwind[3,7], Jack Price[2,6], Jason Carroll[4], Deepak P. Srivastava[2,6+], & Simon Baron-Cohen[1+]

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

[2]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.

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

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

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

+ Joint senior authors

* Joint first authors

Short title: Transcriptome analysis of iPSC-derived neurons

Abstract

Autism Spectrum Conditions (henceforth, autism) are a diverse set of neurodevelopmental phenotypes with a complex genetic basis. Idiopathic autism, characterized by a diagnosis of autism not caused by a known genetic variant, is associated with hundreds of rare and common genetic variants each of small effect. Functional genomics analyses of post mortem brain tissue have identified convergent atypical gene correlation networks in idiopathic autism. However, post mortem tissue is difficult to obtain and is susceptible to unknown confounding factors related to the cause of death and to storage conditions. To circumvent these limitations, we created induced pluripotent stem cells (iPSCs) from hair follicles of idiopathic autistic individuals and made iPSC-derived neurons, to investigate its usefulness as a substitute for post mortem brain tissue. Plucking hair follicles is a relatively painless and ethical procedure, and hair samples can be obtained from anyone. Functional genomics analyses were used as a replicable analysis pipeline to assess efficacy of iPSC-derived neurons. Gene expression networks, previously identified in adult autism brains, were atypical in the iPSC autism neural cultures in this study. These included those associated with neuronal maturation, synaptic maturation, immune response and inflammation, and gene regulatory mechanisms. In addition, GABRA4, HTR7, ROBO1 and SLITRK5 were atypically expressed among genes previously associated with autism. A drawback of this study was its small sample size, reflecting practical challenges in generating iPSCs from patient cohorts. We conclude that, using rigorous functional genomics analyses, atypical molecular processes seen in the adult autistic post-mortem brain can be modelled in hair follicle iPSC-derived neurons. There is thus potential for scaling up of autism transcriptome studies using an iPSC-based model system.

Introduction

Autism spectrum conditions (henceforth autism) are neurodevelopmental in nature, with a heterogeneous genetic background (Berg; Geschwind, 2012; Bourgeron, 2015; O'Roak; et al., 2012). Autism is diagnosed on the basis of impaired social-communication, alongside unusually narrow and repetitive interests and activities (APA, 2013). Idiopathic autism is characterized by a primary diagnosis of autism not associated with a known genetic variant. Exome sequencing studies and analysis of copy number variation have revealed hundreds of rare and common genetic variants are associated with idiopathic autism (O'Roak; et al., 2011; O'Roak; et al., 2012). Functional convergence among these genetic variants has been revealed through RNA sequencing of autistic post mortem brains (Parikshak; et al., 2013; Voineagu; et al., 2011). However, autistic post mortem brain tissue is a scarce resource and may be susceptible to developing anoxic-ischemic changes based on cause of death and post mortem interval (Lewis, 2002). Conditions and period of storage, as well as cause of death, can also have a bearing on RNA quality (Kretzschmar, 2009). More confounding factors may be introduced in brains of donors with a history of illnesses, seizures and substance abuse (Woolfenden; et al., 2012). There are also considerable ethical issues associated with human organ donation for research (Kretzschmar, 2009).

To tackle these challenges, there has been a shift towards development of an induced pluripotent stem cell (iPSC) model of autism, through the differentiation into brain tissue of iPSCs reprogrammed from somatic cells of autistic individuals (Marchetto; et al., 2010; Pasca; et al., 2011). Using this strategy it is possible to create well-defined brain tissue *in vitro* to undertake controlled experiments, something that has not been afforded by post mortem brain tissue. It is hypothesized that the transcriptome of brain tissue derived in this fashion mimics fundamental characteristics of post mortem brain tissue.

There are different methods of generating iPSC-derived brain tissue: 1) the more recent three-dimensional cerebral organoid method capable of generating heterogeneous human brain tissue with neuronal cells of multiple lineages (Lancaster; et al., 2013); or 2) the classical two-dimensional culture method for generating cortical neurons (Shi; Kirwan; Livesey, 2012). As autism is known to affect cortical grey matter, we used the two-dimensional cortical neuron differentiation method in the current proof-of-principle study. Our study aims to demonstrate the efficacy of iPSC-derived neurons or brain tissue as a proxy for post mortem brain tissue, and propose a rigorous pipeline for reliable scaling up of autism transcriptome studies, without the ethical issues associated with human organ donation.

To compare iPSC-derived neurons (iDN) with post mortem brain tissue, we used comparable gene expression network analysis methods previously used to discover atypical gene expression networks in autism (Parikshak; et al., 2013; Parikshak; et al., 2016; Voineagu; et al., 2011). We first undertook RNA-sequencing and differential gene expression in autism and control iDN, followed by weighted gene co-expression network analysis (WGCNA) – an unsupervised analysis method that clusters genes based on their expression profiles (Langfelder; Horvath, 2008), followed by enrichment and transcription factor binding analyses. Additionally, we undertook exome sequencing to identify odds of *de novo* variants in participants being associated with gene expression networks. As we are proposing scaling up of transcriptome studies in autism, we chose an iPSC source cell collection method that is relatively non-invasive and painless for donors, i.e., keratinocytes from plucked hair follicles. Hair follicles is also easier to extract compared to skin fibroblasts or blood mononuclear cells. Using this method it has been possible to collect more than 300 hair samples from autistic individuals as part of the European Autism Interventions - A Multicentre Study for Developing New Medications (EU-AIMS) consortium. However, methodological challenges associated

with reprogramming iPSCs have meant efficiency is low, while costs high (Streckfuss-Bomeke; et al., 2013).

Nevertheless, using WGCNA and associated downstream analyses, we achieved the following: 1) We validated autism gene networks previously identified in post mortem brain tissue by comparing these with gene networks in iDN; 2) We demonstrated the reliable nature of RNA-sequencing and unsupervised clustering analysis methods used in the detection of autism molecular phenotypes. We conclude that neurons and brain tissue from hair follicles of autistic individuals have similar molecular phenotypes as are seen in autistic post mortem brains. We propose that iDN minimizes dependence on post mortem brain tissue and can be scaled up in transcriptome studies in autism.

Results

Differential gene expression analysis reveals descriptive gene expression differences between autism and control day 35 iDN

We undertook RNA-sequencing followed by bioinformatics analysis of early cortical lineage neurons differentiated using established protocol (Shi; Kirwan; Livesey, 2012) (Fig 2(A)). We focused on day 35 from start of neural induction as at this stage there is a mix of early born deep layer neurons and late born upper layer neurons (Shi; Kirwan; Livesey, 2012), making it highly suitable as a snapshot for studying cortical neuron development. We first investigated descriptive gene expression differences by calculating FPKM for each gene (Fragments Per Kilobase of transcript per Million). Multiple linear regression used to study differential gene expression revealed significant differences in gene expression in thousands of genes between

autism and healthy iDN. Fig 2(B) shows the top 50 differentially expressed genes in autism and control participants.

Genome-wide coexpression networks reveal biological processes essential to both neural development and immune activation in autism

We applied signed weighted coexpression network analysis (WGCNA) and identified 11 coexpression modules significantly correlated to autism (labelled by colour, e.g., salmon, Fig 2C), and ranked according to their module eigengene values (ME, the first principal component of the module). These modules represent genes that share highly similar expression patterns (Fig 2D, Fig S1). We then evaluated the shared consensus function of each module by enrichment for Gene Ontology (GO). Of the 11 modules, 5 modules were positively correlated in autism iPSC neurons, while 6 modules were negatively correlated to autism iPSC neurons. The top 3 positively correlated modules – 'steelblue' (Cellular Metabolic Processes), 'lightgreen' (Neural Development) and 'white' (Immune Activation), and the top 3 negatively correlated modules – 'grey60' (Epigenetic Regulation), 'salmon' (Gene Regulation) and 'sienna3' (Chromosome Organisation), were also functionally most significant with resepct to autism (Fig 2D). The consensus functions of each module are non-exclusive, and each module will henceforth be referred to by their corresponding R package designated module colour names.

MEs for 'steelblue', 'lightgreen' and 'white' increase in autism iPSC neurons (Fig 2D). The 'steelblue' module is enriched for metabolic functions associated with post-mitotic cells (Fig 2G) suggesting terminal differentiation. The 'lightgreen' module is enriched for GO terms including regulation of cell-cell adhesion, cognition, calcium mediated signalling and regulation of dendrite maturation associated with late phase neuron development. The most interconnected genes of this module (based on correlation to ME, also known as 'hub genes')

include GABRA4 – a subunit of the inhibitory GABA-A receptor (Roberts; et al., 2005) and CX3CL1 – negative modulator of excitatory glutamatergic neurotransmission (Ragozzino; et al., 2006; Sheridan; et al., 2014) (Fig 2H). The 'white' module is enriched for cytokine binding, regulation of DNA damage response, positive regulation of apoptosis and negative regulation of neuronal death, and its hubs include ADAM9 – implicated in inflammatory events related to neutrophil activation (Amendola; et al., 2015) and GPNMB – expressed in microglia and regulator of immune/inflammatory responses in the brain (Huang; Ma; Yokoyama, 2012) (Fig 2I).

Compared to the positively correlated modules, the negatively correlated modules (Fig 2D) reflect global gene regulatory functions associated with developing neurons, deficient in autism iDN. The 'salmon' module is enriched for RNA methyltransferase activity, epigenetic regulation of gene expression and s-adenosylmethionine-dependant methyltransferase activity (Fig 2I). The 'sienna3' module is enriched for nucleic acid binding, regulation of RNA metabolic process and regulation of gene expression (Fig 2J), while the 'grey60' module is enriched for regulation of histone H3-K4 methylation, DNA binding and chromosome organisation (Fig 2K). HTR7 ('salmon' module), ROBO1 ('sienna3' module) and SLITRK5 ('salmon' module) are known autism risk genes in the negatively correlated modules significantly downregulated in this study, demonstrating a causal link between dysregulated expression of risk genes and autism, explored in detail in the next section.

To take a deeper look at potential gene regulatory function deficiencies, we analysed the interaction of transcription factors with DNA recognition motifs upstream of the start site of differentially expressed genes. Our aim was to study enrichment for transcription factor binding sites that characterize signals regulating transcription (Frith; et al., 2004). We obtained DNA motifs for TF binding from TRANSFAC (Matys; et al., 2003), scanned 1000bp upstream

sequences of top 200 genes from each module and calculated motif enrichment using Clover (Frith; et al., 2004) and MEME (Bailey; Elkan, 1994) algorithms. We identified several TFs (Fig 3A-F) amongst which STAT3 is prominent in the positively correlated modules (Fig 3G). STAT3 motifs were found upstream of 66/163 genes of the 'steelblue' module. It is involved in neuroprotection against inflammatory insults, differentiation and proliferation (Leibinger; et al., 2013; Park; Nozell; Benveniste, 2012). STAT3 is also known to interact with KLF4, which is enriched here for binding targets in the 'white' module, and anti-correlated modules 'sienna3' and 'grey60' (Fig 3F). KLF4 interacts with the JAK-STAT pathway to reduce cell proliferation, and induce apoptosis (Qin; Zhang, 2012). Furthermore, ARID3A – a part of the ARID family of transcriptional regulators is functionally similar to KLF4 (Lin; et al., 2014), and both ARID3A and KLF4 have downstream targets in all three negatively correlated modules 'salmon', 'sienna3' and 'grey60', and the positively correlated 'white' module (Fig 3H).

Thus, gene co-expression modules revealed functional variations between autism and control iDN in this study. This approach also helped us correlate mechanistic variation in iDN with those observed in post mortem brain studies and iPSC-derived minibrain studies, as discussed below.

Gene modules associated with autism brains are highly conserved in iPSC-derived early cortical neurons

Our next aim was to search for autism-associated genes sharing common autism-associated physiological functions. We used a set of 155 autism associated candidate genes from a previous study (Parikshak; et al., 2013) using the Simons Foundation Autism Research Initiative (SFARI) database. The SFARI list of autism associated genes is a database of genes associated with autism, 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 (Basu; Kollu; Banerjee-Basu, 2009) and an up-to-date reference for all known risk genes can be found at: https://gene.sfari.org/autdb/HG_Home.do. We mapped the genes in our gene networks with the SFARI autism risk genes, and found that the SFARI autism risk genes are enriched in the negatively correlated 'salmon' module (p=0.002; odds ratio [OR] = 1.5) (Fig 4A).

First, we mapped 5 developmental gene modules dysregulated in the autism post mortem brain (APMB), from Parikshak et al., (2013) (dev_asdM2, dev_asdM3, dev_asdM13, dev_asdM16, dev_asdM17) (Parikshak; et al., 2013), shown in Fig 4A. Of these 5 sets, dev_asdM2 and dev_asdM3 represent DNA-binding and transcriptional regulation and are downregulated in autism, while dev_asdM14, dev_asdM16 and dev_asdM17 include GO terms that represent nervous system development and synaptic maturation, and are upregulated in autism. The dev_asdM2 set is enriched in the top downregulated genes ('Top –ve DE', $p = 2\times10^{-4}$; OR = 1.8), the 'grey60' module (p = 0.004; OR = 1.6) and the 'sienna3' module ($p = 10^{-5}$; OR = 2). The dev_asdM3 set is enriched in the top downregulated genes ('Top –ve DE', p = 0.008; OR = 1.5), the 'grey60' module ($p = 3\times10^{-14}$; OR = 2.5) and the 'sienna3' module ($p = 4\times10^{-4}$; OR = 1.7). The dev_asdM13 set is enriched in the top upregulated genes ('Top +ve DE', $p = 10^{-6}$; OR = 2.1), the 'lightgreen' module ($p = 3\times10^{-9}$; OR = 3.1) and the 'white' module ($p = 10^{-6}$; OR = 2.3). The dev_asdM16 set is enriched in the 'lightgreen' module ($p = 10^{-4}$; OR = 2.6), and, the dev_asdM17 set is 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 known to be highly correlated with autism in the temporal and frontal cortex, APMB_asdM12 (a neuronal function module) and APMB_asdM16 (an immune module) from Voineagu et al., (2011) (Voineagu; et al., 2011) (Fig 4a). Although the

APMB_asdM12 module has been reported to be downregulated in autism post mortem brain, it shows nominal enrichment in the upregulated 'white' module ($p = 0.04$; OR = 1.8). The APMB_asdM16, on the other hand, is an upregulated module in autism post mortem brains, and is also 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) (Fig 4a). Gene sets associated with attenuated cortical patterning or ACP (ACP_asdM5, ACP_asdM13, ACP_asdM14) (Parikshak; et al., 2016) were also mapped (Fig 4a). This suggested greater prediction of ACP in autism iDN. In Fig 4A we observed significant enrichment of upregulated autism gene sets in positively correlated modules while enrichment of downregulated autism gene sets in negatively correlated modules. This clearly suggests recapitulation of autism post mortem brain phenotypes in iPSC-derived neural cells.

Autism iDN gene modules enriched for genes expressed in cells from mature cerebral cortex

We then plotted enrichment of neuronal cell types in adult human brains (Zhang; et al., 2016) in autism iDN to explore if pathway enrichments discussed above corresponded to cell type differences between autism and control iDN in this study. We were expecting a rise in non-neuronal cell types to reflect enriched immune functions in autism iDN. Not surprisingly, astrocyte-associated markers were detected in the 'white' module ($p < 0.0005$; OR = 2.1), while microglial lineage markers also in the 'white' module ($p < 0.0005$; OR = 2.4). Additionally, mature neuron markers were detected in the 'lightgreen' module ($p < 0.0005$; OR = 1.9), and endothelial markers in the 'steelblue' module ($p < 0.05$; OR = 2.1) and 'white' module ($p < 0.0005$; OR = 2.4) (Fig 4B). Enrichment of mature neuron and glial markers exclusively in the positively correlated modules is indicative of terminally differentiated cell lineages with higher prevalence of immune cells in autism iDN compared to controls. Furthermore, the appearance of these cell types is atypical at day 35 of cortical neuronal differentiation.

Gene networks in autism iPSC-derived early cortical neurons and developing autism iPSC minibrains show moderate preservation

We then mapped gene networks identified in a previous autism iPSC study (Mariani; et al., 2015) with gene networks in this study. We calculated module enrichment of gene modules in the previous autism iPSC study with equivalent colour-assigned modules in this study (Fig 4C). A module with a $Z_{summary} > 10$ corresponds to strong module preservation, and $2 < Z_{summary} < 10$ is equivalent to moderate module preservation. The 'white', 'sienna3' and 'grey60' modules are moderately well preserved in this study ($2 < Z_{summary} < 10$; $p < 0.05$), while the 'lightgreen' and 'salmon' module moderately preserved ($2 < Z_{summary} < 10$; $p < 0.05$) (Fig 4C). Module preservation was only moderate as minibrains (or 'cerebral organoids') were used previously instead of monolayer neuron cultures in this study.

Nominal genetic architecture alterations and de novo variants enrichment in exome

Molecular genetics studies have shown that autism-associated gene mutations are distributed across the whole population with no single major responsible genetic locus, and there is accumulation of thousands of low risk alleles in a subset of autistic individuals (Bourgeron, 2015; O'Roak; et al., 2012). Transcriptome studies have revealed common dysregulated biological processes in autism in independent populations of autistic individuals, using three different methods – post mortem brain studies (Parikshak; et al., 2013; Voineagu; et al., 2011), minibrain studies (Mariani; et al., 2015), and in this study using monolayer neuronal cultures (Fig 4a). To find out whether genetic mutations corresponded to dysregulation of gene expression in our samples, we visualised genomic variants from top 50 differentially expressed genes ($p < 0.005$) found in the individual participants (Fig 5A). Only some of the variants seem to correspond to dysregulated expression, e.g., mutations in WDR49 in CTRM2 and DDIT4L

in CTRM3 corresponds to reduced gene expression in controls, while mutations in ARMC3 and MAP3K19 in ASDM1 and 004ASM corresponds to reduced gene dosage in autism, and a mutation in TRIM14 appears to correspond to reduced gene expression in 004ASM. However these associations are only causal and there maybe additional layers of transcriptional control affecting gene dosage.

Studies have shown rare and unique de novo variants to be major contributors to the genetic signatures of autism (Betancur, 2011; Bourgeron, 2015; O'Roak; et al., 2011; O'Roak; et al., 2012; Yuen; et al., 2015) although it is unclear whether they are associated with converging biological processes. To investigate the correlation between biological processes and genetic events in individual participants, we checked enrichment of gene expression modules in rare and unique exome variants from each participant. We did not find any strong correlation of exome variants with gene expression modules. Rare variants in 010ASM are enriched in top downregulated genes (Top –ve DE) ($p = 0.04$; OR = 1.7), while those in CTRM1 are enriched in the 'salmon' module ($p = 0.03$; OR = 2) (Fig 5B). Unique variants in 010ASM are enriched in the 'grey60' module ($p = 0.03$; OR = 1.4), while those in CTRM3 are enriched in the top downregulated genes (Top –ve DE) ($p = 0.004$; OR = 2.1) and the 'sienna3' module ($p = 0.02$; OR = 1.9) (Fig 5C). This could further suggest additional layers of transcriptional regulation controlling gene expression. We also observed only minor variations in the genetic architecture in the exomes between autism and control participants (Fig S2). However, these variations were not significant and given the small sample size of this study no distinction could be made between the autism and control participants based on their exomes alone.

Discussion

This study aimed to explore the potential of iPSC derived neurons and brain tissue as a proxy for post mortem brain in undertaking autism transcriptome studies. As autism is known to affect

the cerebral cortex, we generated two-dimensional cortical neuron cultures to test efficacy of this method in autism transcriptome analyses, in order to propose: 1) minimizing dependence on post mortem brain tissue and, 2) scale up transcriptome studies in autism. Through this study we have succeeded in delivering evidence of autism associated convergent biological processes and pathophysiological mechanisms in iDN from our cohort. This is also the first study to compare gene expression in autism iPSC neurons with post mortem brain tissue. Using similar network correlation analyses as used in post mortem brain studies and an iPSC-derived organoid study it was possible to identify validated autism-associated mechanisms in iDN generated in this study. The enrichment of post mortem brain cortical networks – dev_asdM13, dev_asdM16 and dev_asdM17 in the iDN neural development ('lightgreen') module is evidence of autism iPSC neurons recapitulating neuronal maturation pathways associated with autism, while enrichment of post mortem brain immune and inflammatory network (APMB_asdM16) in the immune activation ('white') module is evidence of the autism-associated non-neuronal pathways being recapitulated. The dysregulated gene networks also point to a post-mitotic, non-proliferative state. STAT3 was identified as a significant TF and is suggestive of iDN-specific inflammatory responses in autism. STAT3 is also known to interact directly with KLF4 to inhibit cellular regeneration (Qin; Zhang, 2012; Qin; Zou; Zhang, 2013), which has also been identified as a potential transcription factor associated with the dysregulated pathways in this study.

We also found dysregulation in other autism-associated molecular phenotypes recently discovered in autism post mortem brain. For example, we found some evidence of attenuation of cortical patterning (ACP) of differentially expressed genes in autism iDN (Parikshak; et al., 2016; Voineagu; et al., 2011) (Fig 4A), with two of the autistic individuals clustering together based on gene expression (Fig S3). Interestingly, the two individuals clustering together, ASDM1 and 010ASM, also share similar clinical diagnosis (Table S1). There is also a

suggestion of GABA-glutamate signalling imbalance – a prominent phenotype in autism (Coghlan; et al., 2012; Mariani; et al., 2015). GABRA4 and CX3CL1 are hub genes in the positively correlated 'lightgreen' module (Fig 2H). GABRA4 increases GABA neurotransmission while CX3CL1 downregulates glutamate transmission respectively. There might also be disruption of L-glutamate-gated ion channel function, as GRIA4 is enriched in a negatively correlated 'grey60' module (Fig 2L). Excitatory-inhibitory imbalance is a well-known cellular phenotype of autism, and identification of genes related to this phenomenon in autism-correlated networks further demonstrates ability of iDN to recapitulate autism brain phenotypes *in vitro*.

The primary limitation of this study is its low sample size, and that is related to the low reprogramming efficiency afforded by current transcriptional factor-induced reprogramming methods. Collecting hair follicle, however, is the most ethical method for generating source cell for iPSC generation. Thus although lacking statistical power, this study was able to inform us that iPSC derived brain tissue can act as a proxy for post mortem brain for undertaking transcriptome studies, and by using the combination of methods described in here (Fig S4) dependence on the scarce post mortem resource can be reduced while also enabling scaling up.

To summarize, the conclusions in Fig 6 are the result of stringent and exhaustive multiple correlations taking an integrative and unbiased approach, to show how iDN from individuals with idiopathic autism having heterogeneous genetic backgrounds can give rise to an amalgamated pathophysiology that have been previously associated with the adult autism brain through studies using post mortem brain tissue. Through these analyses we propose greater application of the iPSC system in autism transcriptome studies, along with reduced dependence on autism post mortem brain tissue. With technological improvements and collaboration in this

field, bigger cohorts can be studied in the future to study the origin of biological divergence in autism.

Study participants and methods

Induced pluripotent stem cells

IPSCs (2 clones from each individual) were produced from 3 autistic individuals and 3 controls (Fig 1a), using keratinocytes from plucked hair follicles (Aasen; Izpisua Belmonte, 2010) (see extended experimental procedures). Idiopathic autistic individuals were chosen. For clinical diagnosis details, see Table S1.

Neuronal differentiation

We differentiated iPSC lines into cortical neurons using a well-established method (Shi; Kirwan; Livesey, 2012). IPSCs were differentiated till early neuron stage – day 35 (see extended experimental procedures).

Immunocytochemistry

Neuronal differentiation in the iPSC lines were characterised using immunocytochemistry. IPSC cells were differentiated till day 8, day 21 and day 35 and tagged with antibodies of appropriate markers associated with each developmental stage (see extended experimental procedures). Nuclei were stained using DAPI, and imaging was performed using a 40× objective on a confocal microscope (Leica) (Fig 1B).

RNA-sequencing

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 (Dobin; et al., 2013). Quality control was performed using Picard tools (Broad Institute) and QoRTs (Hartley; Mullikin, 2015). Gene expression levels were quantified using an union exon model with HTSeq (Anders; Pyl; Huber, 2015).

Differential gene expression (DGE)

DGE analysis was performed using R statistical packages 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/. 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) (Langfelder; Horvath, 2008) was used to construct coexpression networks as previously shown (Parikshak; et al., 2013). 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 risk gene set was compiled using the online SFARI gene database, AutDB, using "Gene Score" as shown previously (Parikshak; et al., 2013). 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 (Parikshak; et al., 2013). Modules APMB_asdM12 and APMB_asdM16 were obtained from an autism post mortem gene expression study (Voineagu; et al., 2011). We obtained another three autism-associated modules: ACP_asdM5, ACP_asdM13 and ACP_asdM14 from an independent gene expression study profiling dysregulated cortical patterning genes in autism post mortem brain (Parikshak; et al., 2016). All three studies applied WGCNA to identify modules of dysregulated genes in autism.

Gene set overrepresentation analysis

Enrichment analyses were performed either with logistic regression (all enrichments analyses in Fig 5a, 6b, 6c, S3b) or Fisher's exact test (cell type enrichment, Fig 5b). All GO enrichment analysis to characterize gene modules was performed using GO Elite (Zambon; et al., 2012) with 10,000 permutations. Molecular function and biological process terms were used for display purposes.

Transcription Factor Binding Site Enrichment

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

Exome sequencing

200ng of genomic DNA was purified, library prepared and exome enriched using Illumina Nextera Rapid Capture Exome (Illumina). Exome DNA was sequenced on an Illumina HiSeq 2500 machine. Reads were aligned to human genome GRCh37.75 (UCSC version hg19), using the Burrows Wheeler Aligner (BWA) (Li; Durbin, 2009). The genome analysis toolkit (GATK) (Broad Institute) was used to identify single nucleotide polymorphisms (SNPs) and insertion-deletion mutations (indels) by the local de-novo assembly of haplotypes. The genetic variants were evaluated for ratio of transition mutations to transversion mutations (Ti/Tv), heterozygous/homozygous (het:hom) ratio, and insertion/deletion ratio. SNPs and indels were annotated using ANNOVAR (Yang; Wang, 2015) and variant effector predictor (VEP) (McLaren; et al., 2010). Important annotations such as minor allele frequency from the 1000G project, and exonic function were used.

For full methods, see extended experimental procedures.

Acknowledgments

We gratefully acknowledge the participants in this study. This study was supported by grants from the European Autism Interventions (EU-AIMS), 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 awarded to D.P.S., and 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 (J.P. and S.B.C), 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. 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 Annie Kathuria, 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.

Figure legends

Fig 1. Characterisation of iPSC from individuals with and without autism. (a) Schematic of iPSC generation process from keratinocytes, followed by cortical neuron differentiation. Cortical neuron differentiation protocol was characterised by imaging at early neuroprogenitors (day 8), late neuroprogenitors (day 21) and neurons (day 35). (b) Immunofluorescence staining to show morphological changes during developmental progression of autism and control iDN. Confirmation of Ki67^{+} and Nestin^{+} early progenitor (day 8) (scale bar: 10µm), Pax6^{+} late progenitor (day 21) (scale bar: 10µm), TBR1^{+} and MAP2^{+} neurons (day 35) (scale bar: 10µm).

Fig 2. Transcriptome-wide gene co-expression network analysis in autism and control neurons. (a) Schematic of RNA-seq and exome seq experiments and analyses. (b) Sample clustering based on gene expression patterns. Top 50 differentially expressed genes shown here. (c) Signed association of mRNA module eigengenes with autism. Modules with positive values indicate increased expression in autism iPSC-neurons, while modules with negative values indicate decreased expression in autism iPSC-neurons. Red dotted lines indicate Benjamini-Hochberg corrected $p<0.05$. R package designated module colour names are used. (d) Correlation network dendogram with consensus function of significant gene modules and their relationship with autism (red denotes a positive relationship). (e) Module eigengene values of positively correlated gene modules. (f) Module eigengene values of negatively correlated gene modules. (g) Coexpression network plot and GO term enrichment of the ‘steelblue’ module. Top 25 genes are indicated. (h) Coexpression network plot and GO term enrichment of the ‘lightgreen’ module. Top 25 genes are indicated. (i) Coexpression network plot and GO term enrichment of the ‘white’ module. Top 25 genes are indicated. (j) Coexpression network plot and GO term enrichment of the ‘salmon’ module. Top 25 genes are indicated. (k) Coexpression

network plot and GO term enrichment of the ‘sienna3’ module. Top 25 genes are indicated. (l) Coexpression network plot and GO term enrichment of the ‘grey60’ module. Top 25 genes are indicated.

Fig 3. Transcription factor binding site analysis. (a) TFs predicted to bind the 1000bp upstream region of genes in ‘steelblue’ module. (b) TFs predicted to bind the 1000bp upstream region of genes in ‘lightgreen’ module. (c) TFs predicted to bind the 1000bp upstream region of genes in ‘white’ module. (d) TFs predicted to bind the 1000bp upstream region of genes in ‘salmon’ module. (e) TFs predicted to bind the 1000bp upstream region of genes in ‘sienna3’ module. (f) TFs predicted to bind the 1000bp upstream region of genes in ‘grey60’ module. (g) STAT3 binding site motif and GO term enrichment. STAT3, involved in neuroprotection against inflammatory responses is a TF significantly upregulated in autism iPSC-neurons. (h) KLF4 and ARID3A are the most commonly found TFs in this study, having putative binding with gene upstream regions in positively as well as negatively correlated modules.

Fig 4. Selected autism-associated gene modules and cell type enrichment analyses. (a) Enrichment of selected autism post mortem gene modules and SFARI autism risk gene list in iDN gene modules. Odds ratio after logistic regression have been shown. Only OR>1.5 has been shown (p-value in parenthesis). (b) Cell type enrichment in iDN gene modules (*$p<0.05$, ***$p<0.0005$). (c) Module preservation of autism minibrain gene modules in iDN gene modules from this study.

Fig 5. Relationship between differential gene expression and exome variants. (a) Sample-wise clustered top 50 differentially expressed genes, and associated coMut plot of genetic variants (mutations in genes marked in purple seem to have causal effect on gene expression). (b) Enrichment of participant rare variants in iDN gene modules. (c) Enrichment of participant unique variants in iDN gene modules.

Fig 6. Study synopsis showing molecular functions common in autism post mortem brain transcriptome studies and iDN transcriptome in this study.

Supplementary Info

1. Extended experimental procedures
2. Fig S1
3. Fig S2
4. Fig S3
5. Fig S4
6. Fig S5
7. Supplementary figure legends
8. Supplementary table S1

Keywords:

Autism, iPSC, neurons, minibrain, post mortem brain, transcriptome, functional genomics, molecular pathways

Extended experimental procedures

Neuronal differentiation

Keratinocytes were collected from volunteers with and without an autism diagnosis (Ethics approved, 13/LO/1218) as part of a larger European study (EU-AIMS). These were reprogrammed into iPSCs using previously described methods (Aasen; Izpisua Belmonte, 2010; Takahashi; et al., 2007). IPS cells were cultured in E8 medium (Life Technologies) with E8 supplement (Life Technologies). Induction of neurons of cortical lineage was established using a modified dual SMADi protocol(Shi; Kirwan; Livesey, 2012). 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 supress embryonic development thereby inducing neural differentiation pathways, while SB431542 mediates loss of pluripotency.

Immunocytochemistry

Cultures were fixed in 4% formaldehyde followed by ice-cold 100% methanol and processed for immunofluorescence staining and confocal microscopy. 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).

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(Dobin; et al., 2013). Aligned reads were sorted using samtools(Li; et al., 2009), while biases were removed using Picard tools (Broad Institute). Quality control was performed using Picard tools (Broad Institute) and QoRTs (Hartley; Mullikin, 2015). Gene expression levels were quantified using an union exon model with HTSeq (Anders; Pyl; Huber, 2015), 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 (Hansen; Irizarry; Wu, 2012) in R.

mRNA weighted co-expression network analysis

Co-expression network analysis was performed using the R library, WGCNA(Langfelder; Horvath, 2008). 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 (Parikshak; et al., 2013; Parikshak; et al., 2016). 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) (Fig S5). 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 (Csardi; Nepusz, 2006).

Module preservation analysis

Module preservation analysis was performed to validate co-expression in a previous autism-iPSC study (Mariani; et al., 2015). 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 (Langfelder; et al., 2011).

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 (Parikshak; et al., 2013; Parikshak; et al., 2016; Voineagu; et al., 2011) 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 (Zhang; et al., 2016).

Module genes were characterised using GO Elite (version 1.2.5) (Zambon; et al., 2012) 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.

Transcription factor binding site enrichment

Transcription factor binding site (TFBS) enrichment analysis was performed by scanning the canonical promoter region (1000bp upstream of the transcription start site) for the genes in each co-expression module. For each transcription factor (TF), the top 200 connected genes in each module were assessed, using the following steps: 1) putative motifs bound by the TF were obtained from TRANSFAC (Matys; et al., 2003). 2) upstream sequences of these 200 genes were scanned with the Clover algorithm (Frith; et al., 2004) to calculate motif enrichment; and 3) enrichment above background was calculated using the MEME algorithm (Bailey; Elkan, 1994).

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) (Li; Durbin, 2009). 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 (Yang; Wang, 2015) and variant effector predictor (VEP) (McLaren; et al., 2010). Important annotations include minor allele frequency from the 1000G project, SIFT score (Ng; Henikoff, 2003), PolyPhen 2 score (Ramensky; Bork; Sunyaev, 2002), base change, and exonic function.

Supplementary figure legends:

Fig S1: More 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.

Fig S2: Comparison of overall genetic architecture between autism and control participants. (a) Transition vs Transversion (TiTv) ratio. (b) Single nucleotide polymorphisms (SNPs), also indicating average novel mutations and known mutations in each group. (c) Insertion/deletion type mutations (indels), also indicating average novel mutations and known mutations in each group. (d) Heterozygous vs homozygous mutations. (e) Homozygous rare variants. (f) Heterozygous rare variants.

Fig S3: Clustering of individual participant samples based on differential expression patterns of top 50 genes associated with attenuated cortical patterning.

Fig S4: RNA-sequencing sample preparation and bioinformatics pipeline.

Fig S5: Soft thresholding used for exploring co-expressed genes conserved across our cohort.

References

Aasen, T.; Izpisua Belmonte, J. C. (2010): Isolation and cultivation of human keratinocytes from skin or plucked hair for the generation of induced pluripotent stem cells. Nat Protoc 5, 2, 371-382. http://dx.doi.org/10.1038/nprot.2009.241.

Amendola, R. S.; Martin, A. C.; Selistre-de-Araujo, H. S.; et al. (2015): ADAM9 disintegrin domain activates human neutrophils through an autocrine circuit involving integrins and CXCR2. J Leukoc Biol. http://dx.doi.org/10.1189/jlb.3A0914-455R.

Anders, S.; Pyl, P. T.; Huber, W. (2015): HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 2, 166-169. http://dx.doi.org/10.1093/bioinformatics/btu638.

APA (2013): Diagnostic and Statistical Manual of Mental Disorders (DSM-5®).(American Psychiatric Pub).

Bailey, T. L.; Elkan, C. (1994): Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2, 28-36, http://www.ncbi.nlm.nih.gov/pubmed/7584402.

Basu, S. N.; Kollu, R.; Banerjee-Basu, S. (2009): AutDB: a gene reference resource for autism research. Nucleic Acids Res 37, Database issue, D832-836. http://dx.doi.org/10.1093/nar/gkn835.

Berg, J. M.; Geschwind, D. H. (2012): Autism genetics: searching for specificity and convergence. Genome Biol 13, 7, 247. http://dx.doi.org/10.1186/gb4034.

Betancur, C. (2011): Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res 1380, 42-77. http://dx.doi.org/10.1016/j.brainres.2010.11.078.

Bourgeron, T. (2015): From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci 16, 9, 551-563. http://dx.doi.org/10.1038/nrn3992.

Coghlan, S.; Horder, J.; Inkster, B.; et al. (2012): GABA system dysfunction in autism and related disorders: from synapse to symptoms. Neurosci Biobehav Rev 36, 9, 2044-2055. http://dx.doi.org/10.1016/j.neubiorev.2012.07.005.

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

Dobin, A.; Davis, C. A.; Schlesinger, F.; et al. (2013): STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 1, 15-21. http://dx.doi.org/10.1093/bioinformatics/bts635.

Frith, M. C.; Fu, Y.; Yu, L.; et al. (2004): Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res 32, 4, 1372-1381. http://dx.doi.org/10.1093/nar/gkh299.

Hansen, K. D.; Irizarry, R. A.; Wu, Z. (2012): Removing technical variability in RNA-seq data using conditional quantile normalization. Biostatistics 13, 2, 204-216. http://dx.doi.org/10.1093/biostatistics/kxr054.

Hartley, S. W.; Mullikin, J. C. (2015): QoRTs: a comprehensive toolset for quality control and data processing of RNA-Seq experiments. BMC Bioinformatics 16, 224. http://dx.doi.org/10.1186/s12859-015-0670-5.

Huang, J. J.; Ma, W. J.; Yokoyama, S. (2012): Expression and immunolocalization of Gpnmb, a glioma-associated glycoprotein, in normal and inflamed central nervous systems of adult rats. Brain Behav 2, 2, 85-96. http://dx.doi.org/10.1002/brb3.39.

Kretzschmar, H. (2009): Brain banking: opportunities, challenges and meaning for the future. Nat Rev Neurosci 10, 1, 70-78. http://dx.doi.org/10.1038/nrn2535.

Lancaster, M. A.; Renner, M.; Martin, C. A.; et al. (2013): Cerebral organoids model human brain development and microcephaly. Nature 501, 7467, 373-379. http://dx.doi.org/10.1038/nature12517.
Langfelder, P.; Horvath, S. (2008): WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. http://dx.doi.org/10.1186/1471-2105-9-559.
Langfelder, P.; Luo, R.; Oldham, M. C.; et al. (2011): Is my network module preserved and reproducible? PLoS Comput Biol 7, 1, e1001057. http://dx.doi.org/10.1371/journal.pcbi.1001057.
Leibinger, M.; Andreadaki, A.; Diekmann, H.; et al. (2013): Neuronal STAT3 activation is essential for CNTF- and inflammatory stimulation-induced CNS axon regeneration. Cell Death Dis 4, e805. http://dx.doi.org/10.1038/cddis.2013.310.
Lewis, D. A. (2002): The human brain revisited: opportunities and challenges in postmortem studies of psychiatric disorders. Neuropsychopharmacology 26, 2, 143-154. http://dx.doi.org/10.1016/S0893-133X(01)00393-1.
Li, H.; Durbin, R. (2009): Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 14, 1754-1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
Li, H.; Handsaker, B.; Wysoker, A.; et al. (2009): The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 16, 2078-2079. http://dx.doi.org/10.1093/bioinformatics/btp352.
Lin, C.; Song, W.; Bi, X.; et al. (2014): Recent advances in the ARID family: focusing on roles in human cancer. Onco Targets Ther 7, 315-324. http://dx.doi.org/10.2147/OTT.S57023.
Marchetto, M. C.; Carromeu, C.; Acab, A.; et al. (2010): A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell 143, 4, 527-539. http://dx.doi.org/10.1016/j.cell.2010.10.016.
Mariani, J.; Coppola, G.; Zhang, P.; et al. (2015): FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. Cell 162, 2, 375-390. http://dx.doi.org/10.1016/j.cell.2015.06.034.
Matys, V.; Fricke, E.; Geffers, R.; et al. (2003): TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31, 1, 374-378, http://www.ncbi.nlm.nih.gov/pubmed/12520026.
McLaren, W.; Pritchard, B.; Rios, D.; et al. (2010): Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26, 16, 2069-2070. http://dx.doi.org/10.1093/bioinformatics/btq330.
Ng, P. C.; Henikoff, S. (2003): SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 31, 13, 3812-3814, http://www.ncbi.nlm.nih.gov/pubmed/12824425.
O'Roak, B. J.; Deriziotis, P.; Lee, C.; et al. (2011): Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43, 6, 585-589. http://dx.doi.org/10.1038/ng.835.
O'Roak, B. J.; Vives, L.; Girirajan, S.; et al. (2012): Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 7397, 246-250. http://dx.doi.org/10.1038/nature10989.
Parikshak, N. N.; Luo, R.; Zhang, A.; et al. (2013): Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 5, 1008-1021. http://dx.doi.org/10.1016/j.cell.2013.10.031.
Parikshak, N. N.; Swarup, V.; Belgard, T. G.; et al. (2016): Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature. http://dx.doi.org/10.1038/nature20612.

Park, K. W.; Nozell, S. E.; Benveniste, E. N. (2012): Protective role of STAT3 in NMDA and glutamate-induced neuronal death: negative regulatory effect of SOCS3. PLoS One 7, 11, e50874. http://dx.doi.org/10.1371/journal.pone.0050874.

Pasca, S. P.; Portmann, T.; Voineagu, I.; et al. (2011): Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. Nat Med 17, 12, 1657-1662. http://dx.doi.org/10.1038/nm.2576.

Qin, S.; Zhang, C. L. (2012): Role of Kruppel-like factor 4 in neurogenesis and radial neuronal migration in the developing cerebral cortex. Mol Cell Biol 32, 21, 4297-4305. http://dx.doi.org/10.1128/MCB.00838-12.

Qin, S.; Zou, Y.; Zhang, C. L. (2013): Cross-talk between KLF4 and STAT3 regulates axon regeneration. Nat Commun 4, 2633. http://dx.doi.org/10.1038/ncomms3633.

Ragozzino, D.; Di Angelantonio, S.; Trettel, F.; et al. (2006): Chemokine fractalkine/CX3CL1 negatively modulates active glutamatergic synapses in rat hippocampal neurons. J Neurosci 26, 41, 10488-10498. http://dx.doi.org/10.1523/JNEUROSCI.3192-06.2006.

Ramensky, V.; Bork, P.; Sunyaev, S. (2002): Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30, 17, 3894-3900, http://www.ncbi.nlm.nih.gov/pubmed/12202775.

Roberts, D. S.; Raol, Y. H.; Bandyopadhyay, S.; et al. (2005): Egr3 stimulation of GABRA4 promoter activity as a mechanism for seizure-induced up-regulation of GABA(A) receptor alpha4 subunit expression. Proc Natl Acad Sci U S A 102, 33, 11894-11899. http://dx.doi.org/10.1073/pnas.0501434102.

Sheridan, G. K.; Wdowicz, A.; Pickering, M.; et al. (2014): CX3CL1 is up-regulated in the rat hippocampus during memory-associated synaptic plasticity. Front Cell Neurosci 8, 233. http://dx.doi.org/10.3389/fncel.2014.00233.

Shi, Y.; Kirwan, P.; Livesey, F. J. (2012): Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. Nat Protoc 7, 10, 1836-1846. http://dx.doi.org/10.1038/nprot.2012.116.

Streckfuss-Bomeke, K.; Wolf, F.; Azizian, A.; et al. (2013): Comparative study of human-induced pluripotent stem cells derived from bone marrow cells, hair keratinocytes, and skin fibroblasts. Eur Heart J 34, 33, 2618-2629. http://dx.doi.org/10.1093/eurheartj/ehs203.

Takahashi, K.; Tanabe, K.; Ohnuki, M.; et al. (2007): Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 5, 861-872. http://dx.doi.org/10.1016/j.cell.2007.11.019.

Voineagu, I.; Wang, X.; Johnston, P.; et al. (2011): Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474, 7351, 380-384. http://dx.doi.org/10.1038/nature10110.

Woolfenden, S.; Sarkozy, V.; Ridley, G.; et al. (2012): A systematic review of two outcomes in autism spectrum disorder - epilepsy and mortality. Dev Med Child Neurol 54, 4, 306-312. http://dx.doi.org/10.1111/j.1469-8749.2012.04223.x.

Yang, H.; Wang, K. (2015): Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 10, 10, 1556-1566. http://dx.doi.org/10.1038/nprot.2015.105.

Yuen, R. K.; Thiruvahindrapuram, B.; Merico, D.; et al. (2015): Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med 21, 2, 185-191. http://dx.doi.org/10.1038/nm.3792.

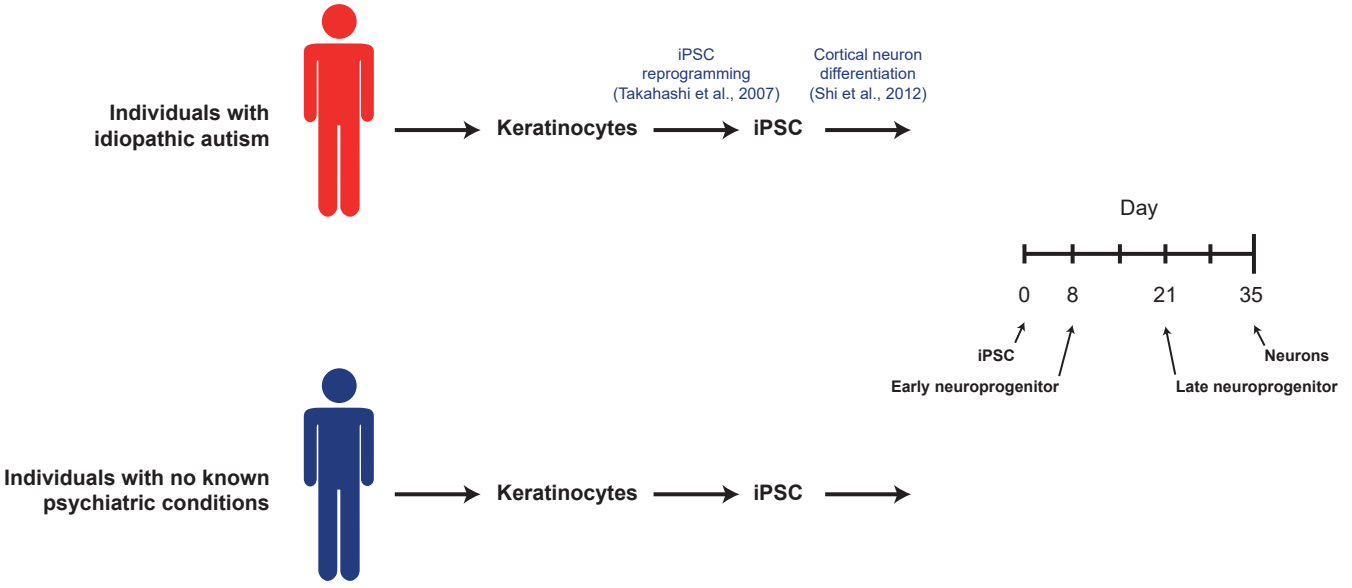
Zambon, A. C.; Gaj, S.; Ho, I.; et al. (2012): GO-Elite: a flexible solution for pathway and ontology over-representation. Bioinformatics 28, 16, 2209-2210. http://dx.doi.org/10.1093/bioinformatics/bts366.

Zhang, Y.; Sloan, S. A.; Clarke, L. E.; et al. (2016): Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. Neuron 89, 1, 37-53. http://dx.doi.org/10.1016/j.neuron.2015.11.013.

781

Figure 1

A


Individuals with idiopathic autism
Keratinocytes
iPSC reprogramming (Takahashi et al., 2007)
iPSC
Cortical neuron differentiation (Shi et al., 2012)
Day
0
8
21
35
iPSC
Early neuroprogenitor
Late neuroprogenitor
Neurons
Individuals with no known psychiatric conditions
Keratinocytes
iPSC

B

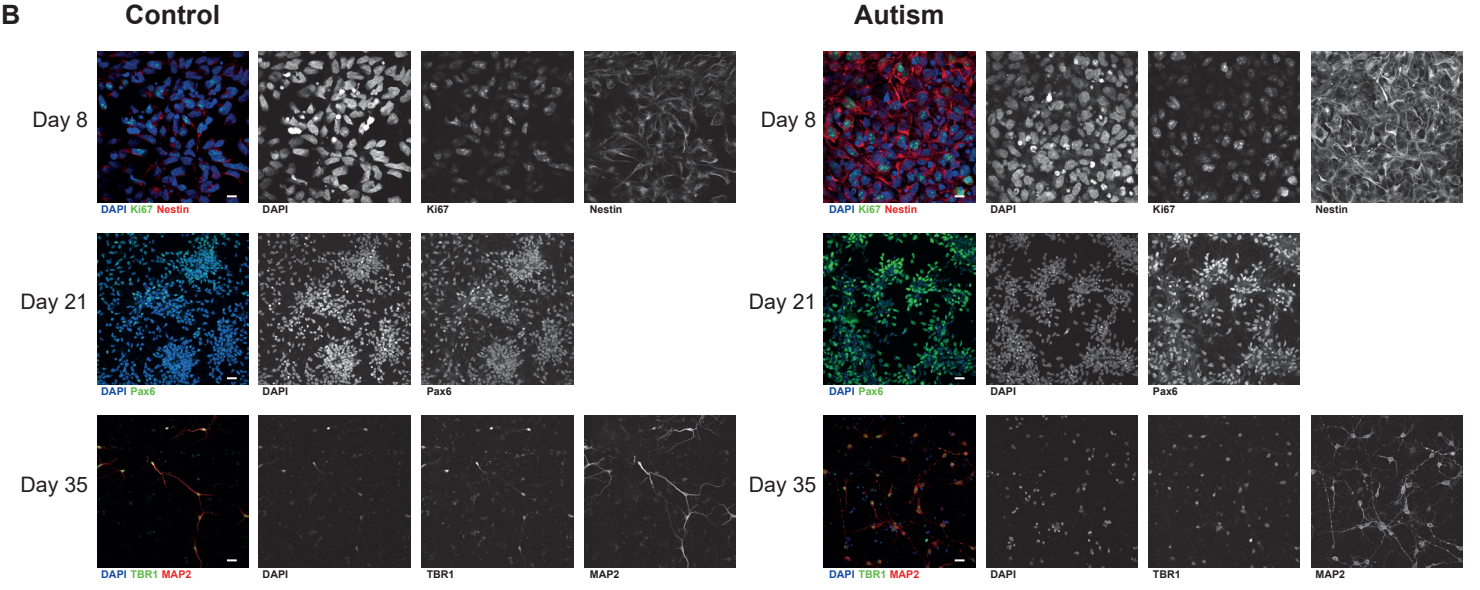

Control
Autism
Day 8
Day 21
Day 35
DAPI Ki67 Nestin
DAPI
Ki67
Nestin
DAPI Pax6
Pax6
DAPI TBR1 MAP2
TBR1
MAP2

Figure 2

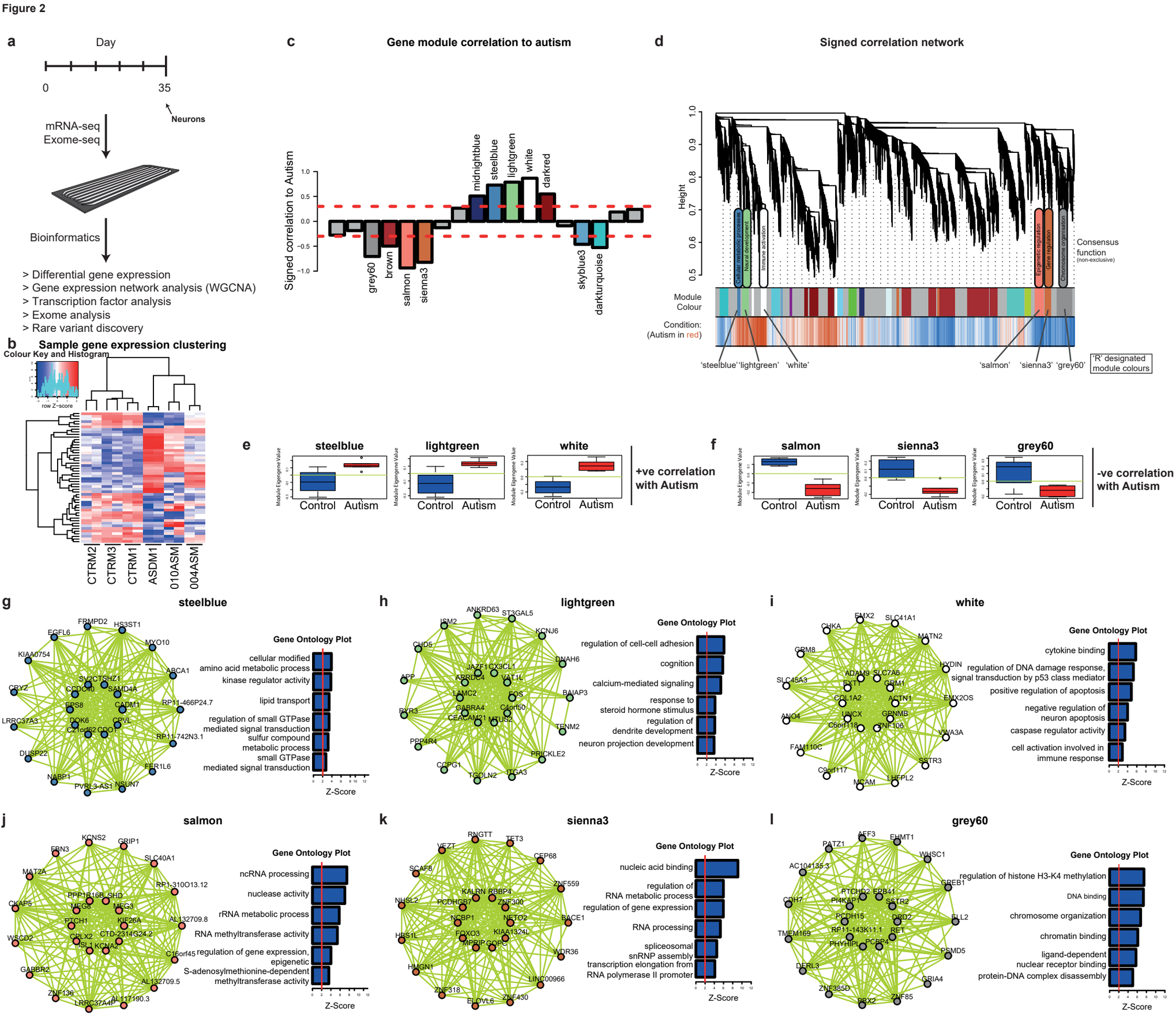

a
Day
0
35
Neurons
mRNA-seq
Exome-seq
Bioinformatics
> Differential gene expression
> Gene expression network analysis (WGCNA)
> Transcription factor analysis
> Exome analysis
> Rare variant discovery
b
Sample gene expression clustering
Colour Key and Histogram
row Z-score
CTRM2
CTRM3
CTRM1
ASDM1
010ASM
004ASM
c
Gene module correlation to autism
Signed correlation to Autism
grey60
brown
salmon
sienna3
midnightblue
steelblue
lightgreen
white
darkred
skyblue3
darkturquoise
d
Signed correlation network
Height
Cellular metabolic processes
Neural development
Immune activation
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
steelblue
lightgreen
white
Control
Autism
+ve correlation with Autism
f
salmon
sienna3
grey60
Control
Autism
-ve correlation with Autism
g
steelblue
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
h
lightgreen
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
i
white
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
j
salmon
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
k
sienna3
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
l
grey60
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

Figure 3

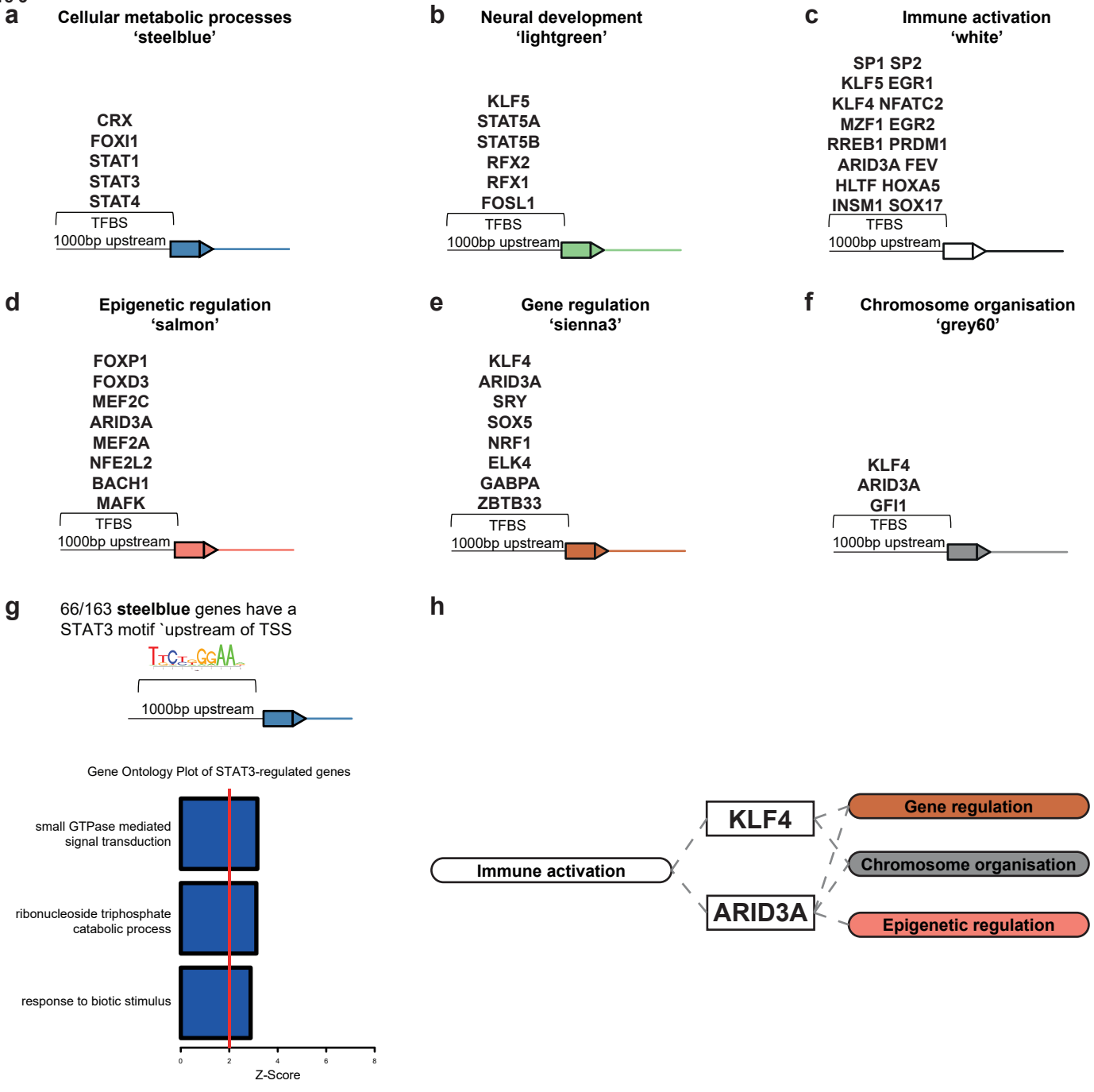

a
Cellular metabolic processes
'steelblue'
CRX
FOXI1
STAT1
STAT3
STAT4
TFBS
1000bp upstream
b
Neural development
'lightgreen'
KLF5
STAT5A
STAT5B
RFX2
RFX1
FOSL1
TFBS
1000bp upstream
c
Immune activation
'white'
SP1 SP2
KLF5 EGR1
KLF4 NFATC2
MZF1 EGR2
RREB1 PRDM1
ARID3A FEV
HLTF HOXA5
INSM1 SOX17
TFBS
1000bp upstream
d
Epigenetic regulation
'salmon'
FOXP1
FOXD3
MEF2C
ARID3A
MEF2A
NFE2L2
BACH1
MAFK
TFBS
1000bp upstream
e
Gene regulation
'sienna3'
KLF4
ARID3A
SRY
SOX5
NRF1
ELK4
GABPA
ZBTB33
TFBS
1000bp upstream
f
Chromosome organisation
'grey60'
KLF4
ARID3A
GFI1
TFBS
1000bp upstream
g
66/163 steelblue genes have a
STAT3 motif `upstream of TSS
1000bp upstream
Gene Ontology Plot of STAT3-regulated genes
small GTPase mediated
signal transduction
ribonucleoside triphosphate
catabolic process
response to biotic stimulus
0
2
4
6
8
Z-Score
h
Immune activation
KLF4
ARID3A
Gene regulation
Chromosome organisation
Epigenetic regulation

Figure 4

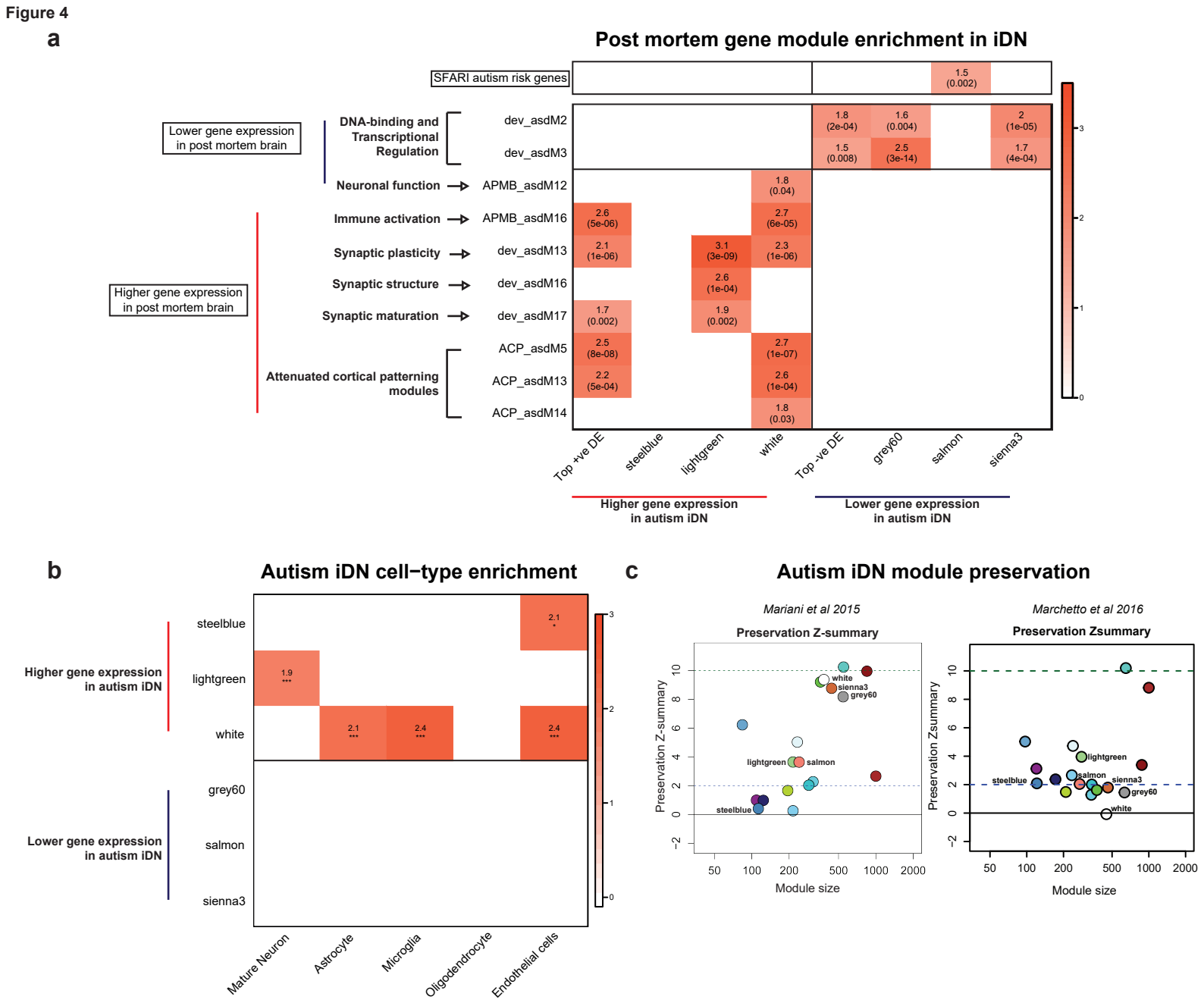

a
Post mortem gene module enrichment in iDN
SFARI autism risk genes
Lower gene expression in post mortem brain
Higher gene expression in post mortem brain
DNA-binding and Transcriptional Regulation
Neuronal function
Immune activation
Synaptic plasticity
Synaptic structure
Synaptic maturation
Attenuated cortical patterning modules
dev_asdM2
dev_asdM3
APMB_asdM12
APMB_asdM16
dev_asdM13
dev_asdM16
dev_asdM17
ACP_asdM5
ACP_asdM13
ACP_asdM14
Top +ve DE
steelblue
lightgreen
white
Top -ve DE
grey60
salmon
sienna3
Higher gene expression in autism iDN
Lower gene expression in autism iDN
b
Autism iDN cell−type enrichment
Higher gene expression in autism iDN
Lower gene expression in autism iDN
steelblue
lightgreen
white
grey60
salmon
sienna3
Mature Neuron
Astrocyte
Microglia
Oligodendrocyte
Endothelial cells
c
Autism iDN module preservation
Mariani et al 2015
Preservation Z-summary
Marchetto et al 2016
Preservation Zsummary
Module size

Figure 5

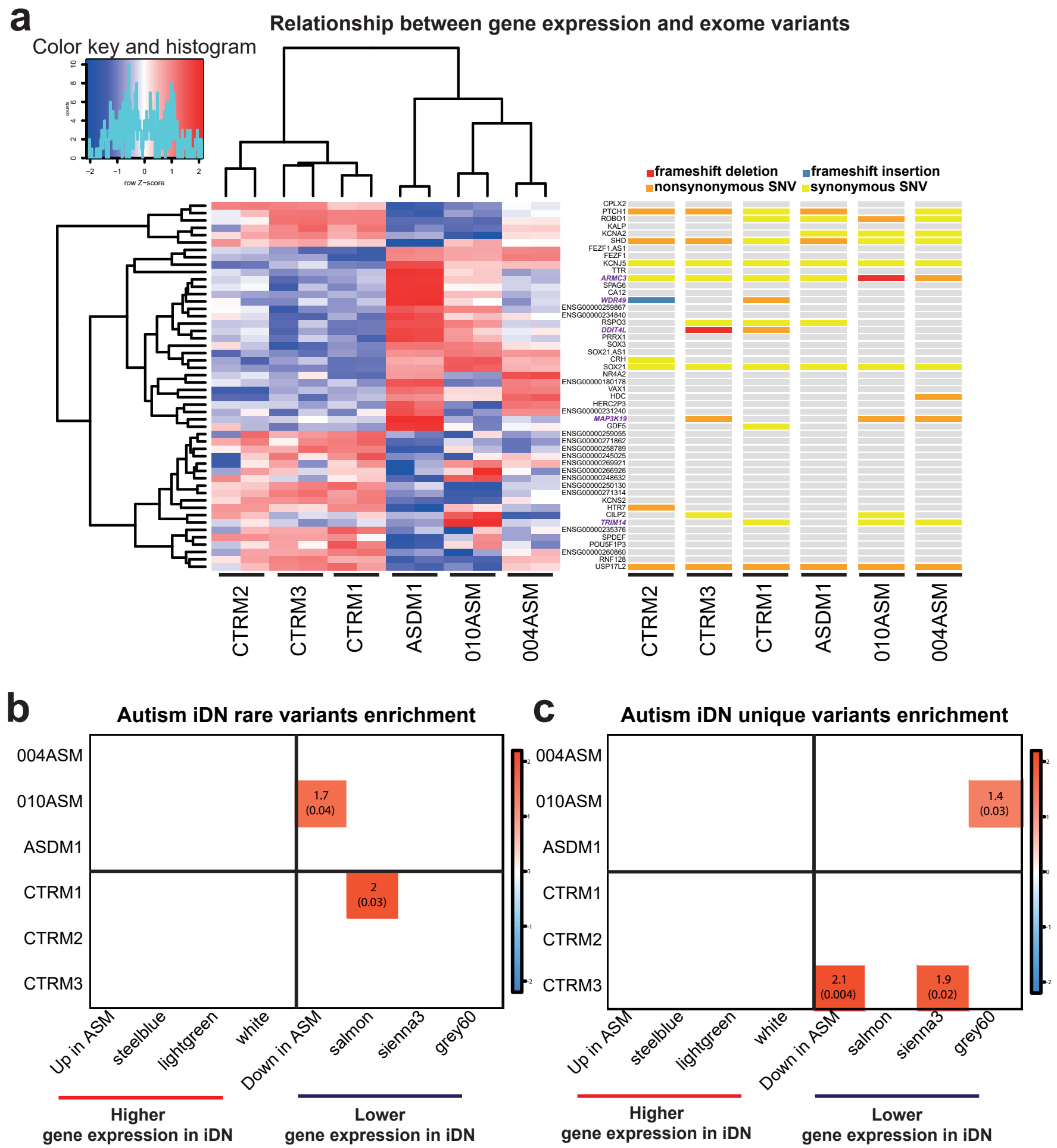
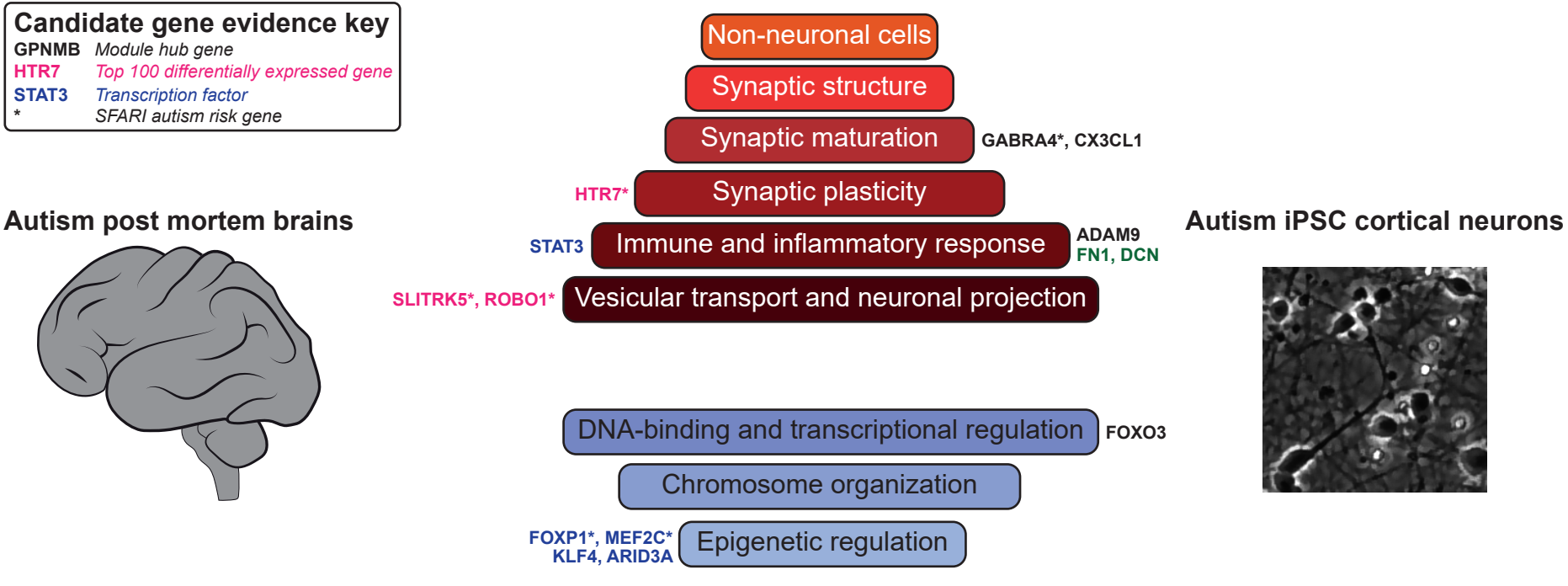

a
Relationship between gene expression and exome variants
Color key and histogram
counts
row Z-score
frameshift deletion
frameshift insertion
nonsynonymous SNV
synonymous SNV
CPLX2
PTCH1
ROBO1
KALP
KCNA2
SHD
FEZF1.AS1
FEZF1
KCNJ5
TTR
ARMC3
SPAG6
CA12
WDR49
ENSG00000259867
ENSG00000234840
RSPO3
DDIT4L
PRRX1
SOX3
SOX21.AS1
CRH
SOX21
NR4A2
ENSG00000180178
VAX1
HDC
HERC2P3
ENSG00000231240
MAP3K19
GDF5
ENSG00000259055
ENSG00000271862
ENSG00000258789
ENSG00000245025
ENSG00000269921
ENSG00000266926
ENSG00000248632
ENSG00000250130
ENSG00000271314
KCNS2
HTR7
CILP2
TRIM14
ENSG00000235376
SPDEF
POU5F1P3
ENSG00000260860
RNF128
USP17L2
CTRM2
CTRM3
CTRM1
ASDM1
010ASM
004ASM
b
Autism iDN rare variants enrichment
004ASM
010ASM
ASDM1
CTRM1
CTRM2
CTRM3
1.7
(0.04)
2
(0.03)
Up in ASM
steelblue
lightgreen
white
Down in ASM
salmon
sienna3
grey60
Higher
gene expression in iDN
Lower
gene expression in iDN
c
Autism iDN unique variants enrichment
1.4
(0.03)
2.1
(0.004)
1.9
(0.02)

Figure 6


Candidate gene evidence key
GPNMB Module hub gene
HTR7 Top 100 differentially expressed gene
STAT3 Transcription factor
* SFARI autism risk gene
Autism post mortem brains
Non-neuronal cells
Synaptic structure
Synaptic maturation
GABRA4*, CX3CL1
HTR7*
Synaptic plasticity
STAT3
Immune and inflammatory response
ADAM9
FN1, DCN
SLITRK5*, ROBO1*
Vesicular transport and neuronal projection
DNA-binding and transcriptional regulation
FOXO3
Chromosome organization
FOXP1*, MEF2C*
KLF4, ARID3A
Epigenetic regulation
Autism iPSC cortical neurons