Pathological priming causes developmental gene network heterochronicity in autistic subject-derived neurons

Simon T. Schafer1, Apua C. M. Paquola1,2,3, Shani Stern1, David Gosselin4, Manching Ku5,6, Monique Pena1, Thomas J. M. Kuret1, Marvin Liyanage1, Abed Alfatah Mansour1, Baptiste N. Jaeger1,7, Maria C. Marchetto1, Christopher K. Glaes8, Jerome Mertens1,9 and Fred H. Gage1,*

Autism spectrum disorder (ASD) is thought to emerge during early cortical development. However, the exact developmental stages and associated molecular networks that prime disease propensity are elusive. To profile early neurodevelopmental alterations in ASD with macrocephaly, we monitored subject-derived induced pluripotent stem cells (iPSCs) throughout the recapitulation of cortical development. Our analysis revealed ASD-associated changes in the maturational sequence of early neuron development, involving temporal dysregulation of specific gene networks and morphological growth acceleration. The observed changes tracked back to a pathologically primed stage in neural stem cells (NSCs), reflected by altered chromatin accessibility. Concerted over-representation of network factors in control NSCs was sufficient to trigger ASD-like features, and circumventing the NSC stage by direct conversion of ASD iPSCs into induced neurons abolished ASD-associated phenotypes. Our findings identify heterochronic dynamics of a gene network that, while established earlier in development, contributes to subsequent neurodevelopmental aberrations in ASD.

Autism spectrum disorder is a highly heritable neurodevelopmental condition defined by deficits in social interaction and communication accompanied by restrictive, repetitive behaviors. Considerable genetic and phenotypic heterogeneity has complicated the efforts to find common biological substrates of the syndrome. Meanwhile, recent advances in systems biology approaches have caused a paradigm shift in the field of biomedical sciences from a single-gene causation model to pathway perturbation models. Implementing high-confidence risk variants discovered in ASD cohorts into a data-driven framework of gene network analysis revealed a convergence on molecular processes that implicate specific perturbations during human fetal cortical development. One complication is that many of the genes identified are involved at multiple time points and in different neural cell types, such as neural stem cells (NSCs), as well as in maturing neurons during human cortical development. Thus, a major challenge in the field is to determine the critical developmental periods, their associated cellular states and the molecular networks that might provide a basis for the emergence of common pathological phenotypes during ASD development.

Here we have performed a time series analysis to follow human idiopathic ASD subject-specific iPSCs during their early neuronal development to examine when and how the earliest ASD-specific molecular abnormalities and phenotypes arise. Time series transcriptome and cellular phenotype analyses in ASD NSCs and their progeny identified dysregulation of specific transcriptional networks that caused aberrant neuronal maturation of ASD cortical neurons. Temporal reconstruction and network-based screening revealed that coherent neurodevelopmental gene modules were accelerated in ASD and appeared to be enriched with several ASD risk genes. Surprisingly, we found that dysregulation of these modules arose even before the neuronal stage, as module-specific signatures were already present during the NSC stage. Genome-wide comparisons of chromatin accessibility revealed a prevalent pathologically primed state in ASD NSCs that precedes the occurrence of aberrant gene network dynamics at the consecutive neuronal stage. Concerted transgenic expression of subsets of module genes was used to mimic ASD-associated gene network-related changes in healthy controls, and was sufficient to recapitulate the occurrence of aberrant neurodevelopmental growth trajectories at the subsequent neuronal stage. We further demonstrated that skipping the NSC stage by direct iPSC-to-neuron conversion (iPSC-iN) also circumvented the critical period for the ASD signature to establish itself, thereby preventing neuronal ASD phenotypes from manifesting. Our study provides a thorough temporal analysis of the neurodevelopmental regulatory gene networks that become impaired in ASD, and shows that some of these changes are triggered during early developmental periods.

Results
Time series-based transcriptome analysis. All iPSC lines used in this study were obtained from a cohort of eight idiopathic ASD

1Laboratory of Genetics, The Salk Institute for Biological Studies, La Jolla, CA, USA. 2Lieber Institute for Brain Development, Baltimore, MD, USA. 3Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. 4Centre de Recherche du Centre Hospitalier Universitaire de Québec-Université Laval, Département de Médecine Moléculaire, Faculté de Médecine, Université Laval, Québec, Canada. 5Next Generation Sequencing Core, The Salk Institute for Biological Studies, La Jolla, CA, USA. 6Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, Faculty of Medicine, University of Freiburg, Freiburg, Germany. 7Laboratory of Neural Plasticity, Faculties of Medicine and Science, Brain Research Institute, University of Zurich, Zurich, Switzerland. 8Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA. 9Department of Genomics, Stem Cell Biology and Regenerative Medicine, Institute of Molecular Biology & CMBI, University of Innsbruck, Innsbruck, Austria. *e-mail: gage@salk.edu
subjects with early developmental brain overgrowth, and five unaffected individuals as controls. Children with such signs of early brain enlargement have been shown to be part of a group of subjects with a high probability of receiving diagnosis\(^5\)\(^6\). Primary skin fibroblasts were harvested and reprogrammed to iPSCs using retroviral vectors expressing the four Yamanaka factors\(^6\) (Supplementary Fig. 1a and Supplementary Table 1). Our previous study showed that neuronal cultures derived from these subjects bear an impaired neuronal network connectivity at later, more mature, neuronal stages. Other iPSC studies have also reported aberrant features of neuronal connectivity as a phenotypic hallmark for various forms of autism\(^1\)\(^2\)\(^3\). However, whether dynamic neurodevelopmental aberrations exist that precede the appearance of such advanced phenotypic disease states remains unresolved. Integrative bioinformatics approaches suggest that ASD pathologies may arise from combined deficiencies during distinct phases of human cortical development\(^4\). To model these early periods, we differentiated the iPSCs into NSCs and then further into functional excitatory MAP2\(^+\) neurons (Supplementary Fig. 1b–d) using a reported protocol\(^9\). All subject-derived cells adopted a cortical fate (Supplementary Figs. 1e–h and 2a,b), and global transcriptional comparison to the BrainSpan dataset\(^7\) revealed high correspondence to the dorsal telencephalon (cerebral cortex with dorsolateral, medial and orbitofrontal cortical areas) during early human fetal cortical development (Supplementary Fig. 2c–e). Furthermore, continuous electrophysiological recordings showed a consecutive transition through stages of functional maturation, including the firing of multiple evoked action potentials after 22 days of differentiation (Supplementary Fig. 2f).

One major aim of this study was to model the temporal dimension of shifting biological processes by treating the disorder as an evolving, dynamic system. We reasoned that a tight recapitulation of the consecutive developmental stages of early neuronal development in ASD and control cases would allow us to reconstruct time-related maturational processes. We first used fluorescence-activated cell sorting (FACS) to purify a homogenous population of NSCs based on the expression of the cell-surface markers CD184\(^+/\)CD271\(^−\)/CD44\(^+/\)CD24\(^−\)/CD15\(^+\) (ref. \(^8\)). To trace ASD and control neurons over time, we performed a series of retroviral lineage-tracing experiments to isolate the progenies of dividing NSCs using a retroviral vector expressing a membrane-tagged enhanced green fluorescent protein (eGFP) (CAG:LCaN-eGFP). Because differentiating neurons express PSA-NCAM on the cell surface (Supplementary Fig. 3a,b), we established a FACS-based protocol for purification of defined subpopulations of retrovirally labeled eGFP\(^+/\)PSA-NCAM\(^+\) double-positive neurons after 2, 4, 7 and 14 days of differentiation (Fig. 1a), which allowed us to capture several maturational stages during lineage progression. We next performed RNA sequencing (RNA-seq) analysis over the time course of neuronal differentiation and observed dramatic but gradual changes in gene expression (Supplementary Fig. 3c). As expected, neuronal genes such as MAP2, MAPT and DCX, voltage-gated channels, synaptics proteins and neurotransmitter receptors (such as GRIN3a and GRIA3) became gradually upregulated, whereas cell cycle-associated genes were consecutively downregulated (Supplementary Fig. 3c, left panel and Supplementary Table 2). Gene ontology term analysis of differentially upregulated genes revealed highly significant enrichment for biological processes that control neuron generation and function (Supplementary Fig. 3c, right panel). A highly significant enrichment for factors that control cell cycle and proliferation was present in the list of the most strongly decreased genes (Supplementary Fig. 3c, right panel and Supplementary Table 3), indicating efficient purification of maturing neurons from defined time points during neuronal differentiation.

Identification of aberrant gene network dynamics in ASD. We next sought to decipher the dynamics of regulatory gene networks to identify deviations in time-critical aspects during ASD development. To cluster genes with similar expression patterns in an unbiased manner, we applied weighted gene co-expression network analysis (WGCNA)\(^11\)\(^12\) to all 65 time series samples (five control and eight ASD lines at five time points) and identified 44 modules of co-expressed genes (Fig. 1b, Supplementary Table 4 and Methods). Here, modules with high gene numbers mostly corresponded to changes within developmental time (Fig. 1b), whereas cell stage-specific events appeared to be a more subtle transcriptional feature reflected in a number of smaller modules (Fig. 1c and Supplementary Fig. 4a). As developmental progression is based on a sequence of events that change over time, we were particularly interested in those modules that showed enrichment for genes holding dynamic temporal profiles. Defining a time-based module significance measure (Methods) allowed us to identify three modules that showed a highly significant correlation with the progression of neurodevelopmental timing, referred to as time modules (TMs) 1, 2 and 3 (Fig. 1d). Alternative network construction strategies (Supplementary Methods) further demonstrated the robustness in identifying all three TMs, suggesting a high preservation within our gene network (Supplementary Fig. 4d and e).

To identify temporal deviations within the developmental sequence between ASD and control neurons, we compared temporal gene expression dynamics on the level of the identified gene modules. To overcome the lack of sensitivity to distortion in the time axis when using the Euclidean distance metric, we applied dynamic time-warping (DTW) algorithms to compare temporal module dynamics between ASD and control neurons. DTW is a variety of time series alignment algorithms designed to measure similarities between two temporal sequences that may vary in speed\(^13\). Module eigengenes (first principal component of a given module) were used as a representative of the gene expression profiles in the three TMs (Fig. 2a). DTW was then applied to compare the module-based dynamics between ASD and control neurons, revealing distinct deviations between the two time series in TM1 (Fig. 2b, left). The two other TMs showed only subtle initial time-line deviations (Fig. 2b, middle and right panels, Supplementary Fig. 4b,c), indicating that the temporal sequences in TM2 and TM3 were almost similar between ASD and control neurons. The temporal progression of TM1, however, appeared markedly accelerated in ASD neurons, with earlier time points corresponding to later time points in control neurons (Fig. 2c), highlighting the emergence of an intrinsic molecular heterochronicity.

To further evaluate the identified developmental gene networks, we computed the relative enrichment for disease-related risk factors in each module based on a set of 859 genes (Supplementary Table 5) evaluated by the Simons Foundation Autism Research Initiative knowledge base\(^9\). Remarkably, TM1 and TM2 showed specific enrichment for ASD risk genes with both featuring highly significant enrichment scores as compared to other modules (Fig. 2d). As further evidence for this confined enrichment, gene expression changes identified in postmortem ASD cortex studies\(^10\)\(^12\) and during human cortical development\(^1\), as well as genes with de novo variants associated with intellectual disability\(^14\)\(^15\), showed significant enrichment in both TMs (Supplementary Fig. 4f). This confined enrichment in TMs suggests an involvement of autism risk factors in the early phases of cortical neuron development, consistent with other recent studies\(^14\), and relates our findings to relevant ASD-associated changes that are present at later stages.

Characterization of the accelerated ASD TM1. Gene ontology term analysis of the genes assigned with TM1 revealed highly significant enrichment for biological processes that control specific aspects of nervous system development (\(P=2.8\times10^{-28}\)), including neuron differentiation (\(P=2.7\times10^{-16}\)), the development of neuronal projections (\(P=3.4\times10^{-15}\)), cell morphogenesis (\(P=3.1\times10^{-15}\))...
To further characterize TM1, we calculated individual gene significance measures for the top 60 hub genes by correlating their expression profiles with developmental time. This module significance measure was defined as the absolute Pearson correlation between time (days of in vitro differentiation) and module eigengenes; TM1: $n = 1,530$ genes; TM2: $n = 294$ genes; TM3: $n = 401$ genes. Student’s $P$ values for associations with time were adjusted for multiple comparisons using the false discovery rate (FDR) approach.

and synaptic signaling ($P = 2.6 \times 10^{-20}$) (Fig. 3a, Supplementary Table 6). As TM1 increases expression with time, the large neurodevelopmental signature of this module suggests that these genes reflect gradually increasing neuronal differentiation. Highly connected genes (‘hub genes’) such as BRSK1, CACNB1, RAB3A and GPC2, together with several other interconnected hub genes, are involved in neuron generation and synapse organization (Fig. 3b).

To further characterize TM1, we calculated individual gene significance measures for the top 60 hub genes by correlating their expression with developmental time and condition. As expression of these genes increases with developmental time, these measures reach highly positive values at later, more mature, neuronal stages (Fig. 3c). When comparing the hub gene trajectories of ASD and control neurons, a pronounced shift in gene significance values markedly emerged at 4 days after differentiation and continued to be prominent at later time points, although differences were already present at the NSC stage (Fig. 3c), suggesting an ASD-specific priming of gene expression already present in pre-neuronal cells. Thus, given the large neurodevelopmental signature of TM1, the presence of such a marked time shift suggests aberrant neurodevelopmental gene network dynamics that could directly disturb the maturational sequence in developing ASD neurons.

Maturing ASD neurons show aberrant neurodevelopmental growth dynamics. We next explored the extent to which certain morphological aspects of ASD neuron development deviate from the ‘normal’ trajectory, in addition to the marked time shift observed at the transcriptional level. Therefore, a series of structural and morphometric analyses was performed on defined subpopulations of eGFP⁺/PSA-NCAM⁺ neurons using a retrovirus-based lineage-tracing approach (Fig. 4a,b and Supplementary Fig. 5a). Strikingly, and consistent with the shift observed in TM1, systematic reconstruction of individual morphological growth trajectories revealed a significant developmental acceleration in differentiating neurons from ASD subjects as compared to control neurons (Fig. 4c,d and Supplementary Fig. 5b). In line with this finding, ASD neurons consistently featured branching patterns that were significantly more complex than those of control neurons (Fig. 4e and Supplementary Fig. 5c). Thus, given the marked time shift in TM1, these data highlight the presence of aberrant maturational dynamics, pertaining to the growth patterns of neuronal projections in developing ASD neurons. As the development of neuronal projections is a time-critical process, we further asked whether this growth acceleration might be accompanied by differences in the states of neuronal maturity. To query intrinsic maturational properties, we designed a retroviral tracing system expressing eGFP under the control of the human doublecortin promoter (hDcx::eGFP; Fig. 4f and Supplementary Fig. 5d). Flow cytometry-based reporter analysis revealed a premature activity state of the introduced hDcx promoter in differentiating neurons from ASD subjects when compared to the trajectory of control neurons (Fig. 4g,h and Supplementary Fig. 5e), indicating differences in both the initiation and dynamics of the neuronal program. Additional measures of maturational features revealed significant deviations in
Fig. 2 | Gene network analysis and DTW identify ASD-specific heterochronicity in a time-related module. a, Module eigengene dynamics of the three notable time-related modules averaged across developmental time. Top: Heatmap showing average module eigengenes across time. Blue, low expression; red, high expression. Bottom: plot of TM1, TM2, and TM3 showing trajectories of module eigengenes across time. b, Three-way plot of the alignment for TM1, TM2, and TM3 between control (reference) and ASD (query). This plot allows display of alignments and places the query time series (ASD) in the small lower panel and the reference time series (control) vertically on the left; the large inner panel holds the warping curve. Left: earlier timepoints in the query index (ASD time series) of TM1 correspond to later timepoints in the reference index (control time series), indicating an accelerated progression of TM1. c, Left: DTW alignment curves superimposed for all three TMs, showing the magnitude and onset of time shift. Right: density plot showing alignment of TM1 between control (reference) and ASD (query). The average cost per step is displayed as a density distribution with contours superimposed. d, Module-specific enrichment for ASD-related risk factors based on the Simons Foundation Autism Research Initiative gene list (see Supplementary Table 5).
GO terms in TM1

Figure 3 | Characterization of the heterochronic TM1. a, Gene ontology (GO) term analysis of the genes assigned to TM1 reveals specific neuro developmental signature. Point sizes correspond to the number of genes assigned to each category (Supplementary Table 6). b, Top gene–gene connections for TM1 are shown, highlighting their involvement in selected biological categories and ASD risk. Different point sizes correspond to the number of timepoints showing increased expression levels. c, Gene significance values of the top 60 hub genes across developmental time and between conditions; measures are represented as mean ± s.d.; n = 60 genes.

Fig. 3 | Characterization of the heterochronic TM1. a. Gene ontology (GO) term analysis of the genes assigned to TM1 reveals specific neuro developmental signature. Point sizes correspond to the number of genes assigned to each category (Supplementary Table 6). b. Top gene–gene connections for TM1 are shown, highlighting their involvement in selected biological categories and ASD risk. Different point sizes correspond to the number of timepoints showing increased expression levels. c. Gene significance values of the top 60 hub genes across developmental time and between conditions; measures are represented as mean ± s.d.; n = 60 genes.

To further explore the developmental implications of this intrinsic neurodevelopmental heterochronicity, we used two different protocols to generate a three-dimensional (3D) model of cortical development. As monolayer cultures prove limited in revealing spatial effects on early cortical development, we used a subset of our cohort (Supplementary Table 7) to assess the formation of the cortical plate in situ. While no major shift in specification of TBR1+ neurons was present in the monolayer system, cerebral organoids as well as TBR1+ cortical regions of forebrain organoids revealed an increased thickness of the cortical plate in the ASD condition (Supplementary Fig. 6f–j), consistent with the hypothesis of increased neuronal numbers and proliferation. These results further emphasize the advantage of 3D models to dissect spatial developmental aberrations.

To exclude the possibility of cell heterogeneity and to study the aberrant cell-intrinsic trajectories, we labeled dividing radial glia-like cells (RGLs) in the ventricular-like zones of 4-week-old forebrain organoids using a GFP-expressing retrovirus (Fig. 5a). This approach allowed us to follow the progenies of labeled RGLs (Fig. 5b, top panels) as they migrated and integrated into the evolving cortical plate (Fig. 5b, bottom panels). In line with our findings from the NSC monolayer system, assessment of developing neurons revealed no difference in overall cell-fate commitment, as both ASD and control RGLs gave rise mostly to TBR1+ deep-layer neurons (Fig. 5c and Supplementary Fig. 6k,l). However, morphological assessment of early-born cortical neurons showed marked differences between the control and ASD groups (Fig. 5d). Quantitative assessment of their early branching patterns showed that ASD neurons harbored aberrantly complex neurite structures at 14 days post-infection (dpi) (Fig. 5e,f). Together with our findings from the monolayer system and in line with the observed transcriptional dysregulation, these data highlight the existence of intrinsic ASD-associated perturbations in the sequence of cortical neuron development.

TM1-specific priming causes aberrant neurodevelopmental growth dynamics. As the timing of neurogenesis is thought to be largely controlled by cell-intrinsic mechanisms encoded within the cortical lineage, we asked whether the disruption of the neurogenic gene program in ASD neurons could have been primed during an earlier pre-neuronal stage in the time line. Differential gene expression analysis indicated early aberrant gene expression dynamics in the ASD group, with an upregulation of twice as many TM1 genes within the first 4 days of differentiation (Supplementary Table 8), pointing to the possible involvement of priming during preceding developmental stages. To assess the stage-specific co-expression of genes in TM1, we excluded network
Fig. 4 | Aberrant neurodevelopmental growth dynamics in maturing ASD neurons. a, Schematic showing experimental design for lineage tracing-based morphometric assessment. b, Representative confocal images of developing neurons 7 days postretroviral infection (dpi) expressing GFP and DCX, and showing immunoreactivity for axonal filaments (Smi312). Scale bars, 50 μm. Immunohistochemistry was performed in all 13 subject lines and two cell culture replicates per timepoint, with similar results (see also Supplementary Fig. 5a). c, Reconstructions of retrovirus (RV)-labeled neurons derived from ASD and unaffected control individuals over time. Scale bar, 50 μm. Experiments were performed in all 13 subject lines and two cell culture replicates per timepoint. Results are shown in d and e. d, ASD neurons showed accelerated growth properties and were significantly longer at 4, 7, and 14 dpi (two-way ANOVA with Sidak correction; 4 dpi: ***P = 0.0002; 7 dpi: **P = 0.006; 14 dpi: ****P < 0.0001). Values show means ± s.d.; ASD (n = 8; 320 technical tracing replicates in total); control (n = 5; 244 technical tracing replicates in total); n refers to biologically independent subject lines. e, Sholl analysis of neurite length from ASD and control neurons at 14 dpi. Total sholl neurite length complexity was higher in the ASD group as compared to controls (two-way ANOVA with Sidak correction, **P = 0.0076; 16 dpi: ****P < 0.0001). Values show means ± s.e.m.; ASD (n = 8; 320 technical tracing replicates in total); control (n = 5; 244 technical tracing replicates in total); n refers to biologically independent subject lines. f, Relative hDCX promoter activities in differentiating neurons. ASD neurons showed premature activation of the inserted hDCX promoter, with significantly elevated levels at the NSC stage and at 16 dpi (two-way ANOVA with Sidak correction, *P = 0.0076; 16 dpi: **P = 0.0059). The data are normalized to the maximum activity over time within each line. Values represent means ± s.e.m.; ASD (n = 8), control (n = 5); n refers to biologically independent subject lines. See also Supplementary Fig. 5.

genes weakly interconnected with TM1 (ME-based connectivity (kME) < 0.7) and calculated the significance of gene-wise correlations to the preceding NSC stage using a signed gene significance measure (Supplementary Fig. 7a). As expected, a high number of TM1-specific genes featured highly significant negative correlations with the NSC stage in the control group (FDR-adjusted P < 0.05; Supplementary Fig. 7b), indicating that these genes were downregulated in comparison to all other stages. Remarkably, almost 80% of these downregulated genes (189 of 237 negatively correlated TM1 genes) lacked significant negative gene-significance values in ASD NSCs (Supplementary Fig. 7c and Supplementary Table 9). This difference between groups was also reflected in a marked shift in gene significance values for these 189 TM1 genes (Supplementary Fig. 7c), suggesting a subtle primed state of the neuronal network program in ASD NSCs. Among the top negatively correlated TM1 genes, the F-box protein 2 (FBXO2) gene is involved in regulating the postmitotic neuronal state and the initiation of synapse formation27,28, whereas other genes encoding for the small guanosine 5′-triphosphate-binding protein RASD2 or the transmembrane protein ISLR2 have either been implicated in schizophrenia29 or have important functions in neurite extension and regulating forebrain connectivity30,31 (Supplementary Fig. 7d–h). We tested whether an early, dynamically aberrant over-representation of TM1-related genes is sufficient to induce ASD-specific morphological and maturational changes. To assess their functional implication, we performed controlled overexpression of selected TM1 genes using FBXO2 alone as well as in combination with additional network genes (Supplementary Fig. 7i–j). As expected, priming the
expression of \textit{FBXO2} at the NSC stage caused a dynamically aberrant initiation of the subsequent neurodevelopmental program, as measured by neurite outgrowth after 4 days of differentiation (Supplementary Fig. 7k,l). This effect was even more pronounced when combining \textit{FBXO2} with two additional TM1-specific network genes (Supplementary Fig. 7l,m), pointing to synergistic gene network-wide effects. These findings highlight the possibility that ASD-specific neurodevelopmental dysregulation is controlled by...
genes identified in TM1 and could arise even before the neuronal stage through an aberrant priming of network-related genes.

Circumventing the pathologically primed NSC stage restores ASD neuron development. To evaluate whether these ASD-associated changes are indeed primed during early periods of neural development, we sought to optimize a technology that could be used to force iPSCs to circumvent early NSC-like stages. We thus generated neurons directly from iPSCs by optimizing a Neurogenin2 (Ngn2)-based induced neuron protocol. IPSCs were lentivirally transduced to express rtTA and a 2A peptide-linked transcript coding for Ngn2 and eGFP (N2AG), resulting in transgenic but silent and expandable N2AG iPSCs (Fig. 6a and Supplementary Fig. 8a).

Following iPSC-iN conversion, iPSCs underwent marked morphological changes by establishing neuronal projections as early as within 4 days (Fig. 6b and Supplementary Fig. 8b). As seen with neurons derived from NSCs through directed differentiation, forced expression of Ngn2 also resulted in the generation of functional neurons with high correspondence to the dorsal telencephalon during human fetal cortical development (Fig. 6c, Supplementary Fig. 8c–i and Supplementary Table 10). However, stringent assessment of early trans-differentiation events revealed that iPSC-iN conversion did not require proliferation (Fig. 6d) and bypassed the emergence of NSC-specific signatures (Supplementary Fig. 9a–e), indicating that this iPSC-iN method indeed generated neurons that had never been in an NSC-like state.

We reasoned that applying this technology in the context of ASD could circumvent the critical NSC-associated priming period required for establishment of the ASD signature. A series of in-depth analyses focusing on maturation and structural development throughout early neuronal growth periods were performed to re-evaluate the differences between control and ASD iPSC-iNs during early neuronal differentiation. Both ASD and control iPSC-iNs underwent marked morphological changes over the time course of differentiation (Fig. 6e). However, quantitative assessment of neuronal differentiation revealed no difference between iPSC-iNs from ASD subjects and those generated from unaffected control individuals, as measured by neuronal commitment and neurite branching patterns (Fig. 7a,b and Supplementary Fig. 9f,g). Further, in sharp contrast to the ASD-specific differences in the growth of neuronal projections observed in NSC-derived neurons (Fig. 4), in-depth morphometric analysis of iPSC-iNs derived from the same iPSC clones did not show this aberrant trajectory (Fig. 7c). To compare the developmental gene network dynamics during iPSC-iN maturation to those identified in NSC-derived neurons, we performed time-series RNA-seq analysis at different stages during iPSC-iN
Fig. 7 | Bypassing the NSC state restores early neurodevelopmental aberrations. a. A significant enrichment of PSA-NCAM+ cells was present in ASD NSC-derived neurons at 4dpi as compared to their respective controls (**P = 0.0016, Mann–Whitney U test), whereas no difference was observed in iPSC-iNs (P = 0.805, Mann–Whitney U test, not significant). Box plots show median (center line), mean (+), and interquartile range, with whiskers representing the minimum and maximum of data points; ASD: n = 8 biologically independent subject lines (42 technical replicates in total); control: n = 5 biologically independent subject lines (22 technical replicates in total). b. Sholl analysis of subtle branching patterns in subject-derived iPSC-iNs at 14 days post-conversion. Values represent mean ± s.e.m.; ASD: n = 8 biologically independent subject lines; control: n = 5 biologically independent subject lines. c. Left: Total neurite growth assessment of converting iPSC-iNs. Values represent mean ± s.d.; ASD: n = 8 biologically independent subject lines (482 tracing replicates total); control: n = 5 biologically independent subject lines (264 tracing replicates total). Right: ASD NSC-derived neurons had significantly longer neurites at 14 dpi as compared to their respective controls (control: 421.8 ± 14.14 μm, ASD: 552.3 ± 24.86 μm, **P = 0.0062, Mann–Whitney U test), whereas no difference was observed in iPSC-iPSC-iNs at 14 days post-conversion (control: 442.3 ± 23.93 μm, ASD: 439.07 ± 14.51 μm, P = 0.9433, Mann–Whitney U test, not significant). Values represent mean ± s.e.m.; ASD: n = 8 biologically independent subject lines; control: n = 5 biologically independent subject lines. d. WGCNA cluster dendrogram of all 52 iPSC-iN time series samples groups genes into distinct modules (top row). The middle row shows strong differential expression relationships for developmental time (days of iPSC-iN conversion). TM1 genes (bottom row; Fig. 2) show high module preservation in the iPSC-iN gene network (also see Supplementary Fig. 10c–e). e. Density plot showing the temporal module eigengenes alignment for the TM1-equivalent iPSC-iN module (MEblue) between control (reference) and ASD (query) during iPSC-iN maturation. The average cost per step is displayed as a density distribution with contours superimposed. f. GS values of the top 60 TM1 hub genes for ASD and control individuals in NSC-derived neurons (left) or when bypassing the NSC stage (iPSC-iN technology, right) after 14 days. Violin plots show median (center line), interquartile range (rectangle), 95% confidence interval, and the kernel probability density at different values; n = 60 genes. Also see Supplementary Figs. 8, 9, and 10.

conversion (Supplementary Fig. 10a). We applied WGCNA to identify modules of co-expressed genes, similar to the approach used for NSC-derived neurons (Fig. 7d, Supplementary Fig. 10b and Supplementary Table 11). Comparing the networks of NSC-derived neurons (Fig. 1) to those of iPSC-iNs revealed a significant conservation of TM1 (Supplementary Fig. 10c–e). In sharp contrast to the NSC differentiation data, DTW revealed no trajectory difference in the progression of the TM1-equivalent gene network between control and ASD in iPSC-iNs (Fig. 7e and Supplementary Fig. 10f,g). Furthermore, the prominent ASD-associated transcriptional shift among the top 60 TM1 hub genes, which was prevalent in NSC-derived neurons, appeared to be abolished in iPSC-iNs (Fig. 7f).

Thus, circumventing early NSC-like stages by using the iPSC-iN technology rescued ASD-specific phenotypes and restored aberrant transcriptional signatures in ASD neurons, suggesting a pathological priming period during early neurodevelopmental periods.

Aberrant gene network dynamics at early neuronal stages are associated with changes in chromatin accessibility at preceding NSC stages. Given the contribution of early NSC stages in the subsequent ASD-associated phenotype, we sought to further characterize possible mechanisms underlying this pathologically primed state. We reasoned that the assay for transposase-accessible chromatin using sequencing (ATAC-seq) in NSCs would allow us to obtain a depiction of preceding changes in the chromatin landscape. We thus isolated NSCs by FACS using the same surface markers as described above before subjecting them to ATAC-seq. All size-selected ATAC-seq libraries showed the characteristic fragment length distribution (Supplementary Fig. 11a), with the majority of reads falling within the nucleosome-free region (Supplementary Fig. 11b). We assessed differential accessibility between the two conditions and discovered 833 peaks with greater accessibility in ASD NSCs and 760 peaks with decreased accessibility in the ASD group as compared to control NSCs (Fig. 8a,b, FDR-adjusted P < 0.05). Hierarchical clustering based on the correlation of accessibility of differential accessibility sites separated the samples into two groups based on their condition (Fig. 8c). Similarly, the identified differential accessibility peaks could be used to separate the sample groups along the first and second principal components of variation, explaining 75 and 16% of variation, respectively (Supplementary Fig. 11d).
Fig. 8 | Aberrant gene network dynamics at early neuronal stages are associated with changes in chromatin accessibility at preceding NSC stages.

a. Coverage maps of normalized ATAC-seq signals from ASD and control NSCs showing a differentially accessible (DA) peak (highlighted in red) near the DNM8P gene on chromosome 10. Group-wise sample coverages are displayed, and annotated chromatin states are based on the ENCODE 25-state model (see Supplementary Methods).

b. Binding affinity heatmap showing normalized accessibilities for DA peaks.

c. Correlation heatmap showing hierarchical clustering based on DA peaks (Pearson correlations of peak scores); ASD: n = 8 biologically independent subject lines; control: n = 5 biologically independent subject lines.

d. Log2-fold enrichment of DA peaks in epigenetically annotated regions of the genome (ENCODE 25-state model) shows significant enrichment in promoter and enhancer regions (GAT randomization test; \( P < 6.717 \times 10^{-4} \)).

e. Enrichment of DA peaks in gene-distal regions of TMs.

f. Metagene profiles of normalized ATAC-seq signals at promoters around ±1kb from the transcription start-site (TSS) in NSCs from ASD subjects (red) or controls (black). Left: metagene profiles of promoter regions from TM1 hub genes (kME > 0.8) show higher accessibility in ASD NSCs (\( P = 6.717 \times 10^{-4} \); Hotelling’s t test; n = 273 elements). Right: metagene profiles of promoter regions from randomly accessible background genes (\( P = 0.7761 \); Hotelling’s t test, n.s.; n = 273 elements). Numbers of plotted elements (genes) are similar.
Differential accessibility peaks are significantly enriched in previously annotated chromatin states, including promoters and enhancers, but depleted in regions containing quiescent and actively transcribed regions of the genome (Fig. 8d). Previous studies have shown that regulatory modules as far as 50 kb and more from the transcriptional start site have important contributions to the prediction of a gene’s transcriptional activity. Examining 50 kb windows around genes assigned to any of the three TMs showed that differential accessibility peaks with increased accessibility in ASD NSCs fall significantly more often in TM1 gene-distal regions than would be expected by chance, whereas gene-distal regions of TM2 and TM3 did not show any enrichment for differential accessibility peaks (Fig. 8e and Supplementary Fig. 11e,f). Remarkably, metagene analysis of gene promoters from TM1 hub genes (kME > 0.8) showed a significantly increased amount of nucleosomal depletion in ASD NSCs compared to control NSCs (273 elements, P < 6.717 × 10⁻⁴); this depletion was not present in accessible promoters of randomly selected background genes (Fig. 8f and Supplementary Fig. 11g). These findings highlight the involvement of altered epigenetic chromatin remodeling processes and suggest that, at least in part, increased accessibility in promoter and gene-distal regions may contribute to the aberrant enhancement of transcription that occurs at the subsequent neuronal stage. In line with the finding that the manifestation of the aberrant TM1-specific signature establishes itself during early NSC-like stages, these data provide further evidence for the existence of early pathologically primed states that may trigger subsequent aberrations during ASD development.

**Discussion**

Our work demonstrates how ASD-associated neurodevelopmental aberrations that disturb the maturational sequence of cortical neuron development are triggered by a pathological priming of gene regulatory networks that evolve during early neural development. These findings contribute to the emerging picture of convergent molecular pathologies in ASD and identify NSC stages as critical developmental periods that lay the groundwork for disease propensity.

Despite etiological heterogeneity, recent large-scale gene expression studies on ASD postmortem brain regions have shown consistent transcriptome changes that evolve during the first decade of brain development. These ASD-associated changes are thought to represent the consequences of an ongoing process that is triggered largely by genetic and epigenetic factors early in development. However, little is known about whether this plethora of heritable genetic components might converge on a coherent gene network during specific periods of fetal cortical development. Our current study is based on a carefully selected cohort of macrocephalic ASD subjects representing a highly replicated and significant phenotype within the spectrum. Previous ASD postmortem studies have pointed to pathologies in neuronal and micro-columnar organization, spine growth and dendrite morphology, suggesting aberrant processes during early brain development. Considering that live neurons from subjects are inaccessible and most postmortem tissues are limited to later stages of disease pathogenesis, iPSC-based model systems provide an unprecedented tool to study the progression of early developmental events. Here, we have developed a time series approach that allowed us to follow human idiopathic ASD individual-specific iPSCs during their early cortico-neuronal development. Transcriptome-wide screening revealed an ASD-associated temporal dysregulation within a specific neurodevelopmental gene module that caused marked changes in the maturational sequence of cortical neuron development, including morphological growth acceleration and heterochronous initiation of the neuronal program. It is likely that a disruption of the neotenic gene expression program during cortical neuron development has wide-ranging consequences for a timely, coordinated establishment of interconnected neuronal networks. Experiments in rodents have shown that altering the trajectory of early postnatal cortical development can lead to structural and behavioral features of autism. In line with this finding, human iPSC-based studies have confirmed that neuronal connectivity is affected in various forms of autism. However, while cortical neurons are thought to be particularly susceptible for certain ASD risk genes, it is currently not well understood whether common ASD-associated changes that evolve at the neuronal stage are the consequence of an ongoing process that is primed early in development.

Here we show that the observed neurodevelopmental heterochrony in ASD neurons originates from functional neurodevelopmental gene networks that become pathologically primed during the pre-neuronal NSC stage. By mimicking this ASD-specific signature in control NSCs using controlled over-representation of a subset of network genes, we show that pathological dysregulation in NSCs is sufficient to induce ASD-associated morphological and maturational changes at later neuronal stages. While we tested only a selected subset of network-related genes, it would be interesting to investigate how different combinations of network genes may contribute to ASD-associated changes in a context- and subject-specific manner.

Trans-differentiation of somatic cells such as fibroblasts into induced neurons has been used as an alternative technology to directed differentiation. More recently this technology has been expanded to enable rapid generation of functional neurons by transgenic overexpression of proneural transcription factors in iPSCs. However, it is well established that forced expression of induced neuron pioneer factors during fibroblast-based induced neuron conversion does not require proliferation. A recent study further demonstrated that the direct programming path diverges into a novel transitional state, bypassing intermediate progenitors but converging onto the same final state. In line with our findings, these data indicate that iPSC-2Ns indeed generate neurons that have never been in a NSC-like state, thus making them a promising tool to evaluate the implication of early neurodevelopmental states in priming disease progression. Using this technology, we demonstrate that bypassing NSC-like stages by direct conversion of ASD iPSCs into postmitotic neurons prevents manifestation of the observed neuronal ASD-associated phenotypes. Our analysis suggests that some ASD-associated changes are likely to be the consequence of pathological events that are triggered during NSC stages early in development. A genome-wide analysis of the chromatin state further identified a contribution of preceding epigenetic changes to priming disease propensity. Future studies are needed to assess the implication of specific epigenetic aspects on the onset and severity of autism-related disease trajectories. Finally, given the growing evidence of specific neurodevelopmental alterations in ASD, these findings and the methodological framework will be valuable for guiding future mechanistic studies aimed at identifying a convergence of causal genetic variations that contribute to ASD risk.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at [https://doi.org/10.1038/s41593-018-0295-x](https://doi.org/10.1038/s41593-018-0295-x).

Received: 4 August 2018; Accepted: 13 November 2018; Published online: 7 January 2019

**References**

1. de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* 22, 345–361 (2016).

2. Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* 16, 441–458 (2015).
3. Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008–1021 (2013).

4. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, 997–1007 (2013).

5. Courchesne, E., Carper, R. & Akshoomof, N. Evidence of brain overgrowth in the first year of life in autism. J. Am. Med. Assoc. 290, 337–344 (2003).

6. Courchesne, E. et al. Mapping early brain development in autism. Neuron 56, 399–413 (2007).

7. Emerson, R. W. et al. Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. Sci. Transl. Med. 9, eaag2882 (2017).

8. Hazlett, H. C. et al. Early brain development in infants at high risk for autism spectrum disorder. Nature 542, 348–351 (2017).

9. Marchetto, M. C. et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. Mol. Psychiatry 22, 820–835 (2017).

10. Mariani, J. et al. FOXP1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. Cell 162, 375–390 (2015).

11. Shcheglovitov, A. et al. SHANK3 and IGF1 restore synaptic deficits in neurons from 22q13 deletion syndrome patients. Nature 503, 267–271 (2013).

12. Stein, J. L. et al. A quantitative framework to evaluate modeling of cortical development by neural stem cells. Neuron 83, 69–86 (2014).

13. Allen Institute for Brain Science. Atlas of the developing human brain. BrainSpan http://www.brainspan.org/ (2016).

14. Yuan, S. H. et al. Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. PLoS One 6, e17540 (2011).

15. Zhang, B. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).

16. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559 (2008).

17. Sakoe, H. Dynamic programming algorithm optimization for spoken transcription factor occupancy and chromatin state. J. Neurosci. 24, 1238–1254 (2016).

18. Wilson, J. A. et al. Mouse embryonic stem cells can differentiate via multiple defined factors. Nature 493, 423–427 (2013).

19. Bhandari, A. et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 521, 216–221 (2014).

20. Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. Nature 501, 373–379 (2013).

21. Qian, X. et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. Nature 521, 216–221 (2014).

22. Molyneaux, B. J., Arlotta, P., Menezes, J. R. L. & Macklis, J. D. Neuronal synaptic connectivity. J. Neurosci. 35, 262–262 (1998).

23. Willsey, A. J. et al. Coexpression networks implicate human midfetal forebrain anterior commissure. J. Comp. Neurol. 352, 223–234 (1970).

24. Courchesne, E. & Kemper, T. L. Neuroanatomic observations of the brain postnatal cortical development can lead to structural and behavioural features of autism. BMC Neurosci. 11, 102 (2010).

25. Casanova, M. F. et al. Abnormalities of cortical minicolumnar organization in the prefrontal lobes of autistic patients. Clin. Neurosci. Res. 6, 127–133 (2006).

26. Wang, B. & Horvath, S. A general framework for weighted gene coexpression network analysis. Stat. Appl. Genet. Mol. Biol. 4, e17 (2005).

27. Vierbuchen, T. et al. Direct conversion of fibroblasts into neuronal cells. Nature 468, 1035–1041 (2010).

28. Fishman, V. S. et al. Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells. Cell Cycle 14, 1188–1196 (2015).

29. Briggs, J. A. et al. Mouse embryonic stem cell lines that can differentiate via multiple pathways to the same state. Elife 6, e26945 (2017).

30. Heeger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics 29, 2046–2048 (2013).

31. Fishman, V. S. et al. Cell divisions are not essential for the direct conversion of fibroblasts into neuronal cells. Cell Cycle 14, 1188–1196 (2015).

32. Fein, E. et al. High-resolution mapping of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294 (2013).

33. Chow, M. L. et al. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature age. PLoS Genet. 8, e1002592 (2012).

34. Gupta, S. et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat. Commun. 5, 5748 (2014).

35. Sun, W. et al. Histone acetylome-wide association study of autism spectrum disorder. Cell 167, 1385–1397 (2016).

36. Carver, C. et al. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. Nat. Neurosci. 20, 1217–1224 (2017).

37. HaZlett, H. C. et al. Early brain overgrowth in autism associated with an increase in cortical surface area before age 2 years. Arch. Gen. Psychiatry 68, 467–476 (2011).

38. Schumann, C. M. et al. Longitudinal magnetic resonance imaging study of cortical development through early childhood in autism. J. Neurosci. 30, 4419–4427 (2010).

39. Bauman, M. L. & Kemper, T. L. Neuroanatomical observations of the brain in autism: a review and future directions. Int. J. Dev. Neurosci. 23, 183–187 (2005).

40. Belmonte, M. K. et al. Autism and abnormal development of brain connectivity. J. Neuroscience 24, 9228–9231 (2004).

41. Casanova, M. F., Buchhoeveden, D. P., Switala, A. E. & Roy, E. Minicolumnar abnormalities in autism. Neurology 58, 428–432 (2002).

42. Casanova, M. F. et al. Abnormalities of cortical minicolumnar organization in the prefrontal lobes of autistic patients. Clin. Neurosci. Res. 6, 127–133 (2006).

43. Weir, R. K., Bauman, M. D., Jacobs, B. & Schumann, C. M. Protruded dendritic growth in the typically developing human amygdala and increased spine density in young ASD brains. J. Comp. Neurol. 526, 262–274 (2018).

44. Chomiak, T., Karrlik, Y., Block, E. & Hu, B. Altering the trajectory of early postnatal cortical development can lead to structural and behavioural features of autism. BMC Neurosci. 11, 102 (2010).

45. Chen, S. et al. Auditory-cued sensorimotor task reveals disengagement deficits in rats exposed to the autism-associated teratogen valproic acid. Neurosci. 268, 212–220 (2014).

46. Heeger, A., Webber, C., Goodson, M., Ponting, C. P. & Lunter, G. GAT: a simulation framework for testing the association of genomic intervals. Bioinformatics 29, 2046–2048 (2013).

Acknowledgements We thank the subjects who participated in this study. We also thank M.L. Gage and M. Pena for editorial comments; E. Courchesne for his contribution to fibroblast collection; L. Moore, A. Mendes, B. Miller, K. Rehder, A. Mar and K. Niedrig for technical assistance; C. Difenderfer and the Salk STEM core facility for technical support; C. Fitzgerald and C. O’Connor for help with flow cytometry; and T. Toda for helpful discussions. This work was supported by the Flow Cytometry Core Facility and the Salk Institute with funding from NIH-NICHD CC555 P03 014195, the Chapman Foundation and the Helmsley Charitable Trust. We also acknowledge support from The James S. McDonnell Foundation, G. Harold & Leila Y. Mathers Charitable Foundation, JPB Foundation, the March of Dimes Foundation, NIH (grant nos MH095741 and MH090258 to E.H.G. and R03 MH151426-01A1 to M.C.M.), The Engman Foundation, Annette C. Merle-Smith, The Paul G. Allen Family Foundation and The Leona M. and Harry B. Helmsley Charitable Trust (grant no. 2017 PG-MED001). S.T.S. was funded by a fellowship from the German Research Foundation and was recently awarded the NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation.

Author contributions S.T.S. designed, performed, analyzed and contributed to all experiments. A.P. aligned the RNA-seq dataset, performed statistical analysis and helped with data interpretation. S.S. conducted and analyzed the electrophysiological recording experiments. M.K. performed RNA-seq experiments and helped with interpretation of results. D.G. performed ATAC-seq experiments. D.G. and C.K.G. analyzed data and helped with interpretation of results. M.P. conducted flow cytometry, imaging experiments, analyzed data and assisted with experiments and organoid experiments. M.I. and T.K. conducted structural and morphological analysis and performed cloning experiments. A.M. helped with organoid experiments and analyzed data. B.I. contributed to flow cytometry analysis. M.C.M. provided iPSC lines and helped with data interpretation. J.M. provided constructs, assisted with establishment of iPSC-IN protocol and helped with data interpretation. E.H.G. supervised the experimental design and analysis, interpreted results and provided funding. S.T.S. and E.H.G. wrote the manuscript and conceptualized the study.
Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0295-x.

Reprints and permissions information is available at www.nature.com/reprints.
Correspondence and requests for materials should be addressed to F.H.G.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods

**iPSC reprogramming and maintenance.** Primary human dermal fibroblasts were obtained from subjects recruited through the University of California, San Diego, Autism Center of Excellence and reprogramming was performed using the classical retroviral approach. This study protocol was approved by Salk Institute’s IRB Committee (FWA 00005316) and the Embryonic Stem Cell Research and Oversight Committee. The Salk Institute is committed to protecting the rights and welfare of human research participants and ensures compliance with all applicable ethical and legal requirements. After a complete description of the study, written informed consent was provided by the parents of the human participants. iPSCs were generated from eight male subjects diagnosed with ASD according to the DSM-IV (average age 13.25 ± 5.67 years) and five male control participants (average age 11.4 ± 5.22 years), none of whom had a history of psychological, genetic or other disorders (Supplementary Table 1). Control subjects were selected randomly from lists of typically developing individuals who had been administered a modified autism diagnostic assessment (MADI) scan when they were toddlers. ASD subjects were selected from lists of those who had been identified and diagnosed with ASD and MRI scanned when they were toddlers. From among these potential ASD subjects, those with larger than normal average total brain volume as compared to typically developing toddlers were selected. The iPSC lines obtained from this cohort have previously been described and were reprogrammed in the same facility (Salk Institute for Biological Studies, Laboratory of Genetics) and under the same conditions. Briefly, fibroblasts were transfected with retroviruses containing SOX2, OCT4, KLF4 and MYC to induce overexpression of these genes and were transferred to a co-culture system with murine embryonic fibroblasts. iPSC colonies were identified after around 2 weeks in this culture system, plated onto Matrigel-coated plates (BD Biosciences) and maintained with mTeSR1 media (Stem Cell Technologies). All cell lines were regularly tested for mycoplasma contamination. Experiments were performed with one or two independent iPSC clones (for details see Supplementary Table 7).

**Neural differentiation.** We differentiated iPSCs into NSCs by following a default contamination. Experiments were performed with one or two independent iPSC lines (Stem Cell Technologies). All cell lines were regularly tested for mycoplasma and were transferred to a co-culture system with murine embryonic fibroblasts. Under the same conditions. Briefly, fibroblasts were transfected with retroviruses containing SOX2, OCT4, KLF4 and MYC to induce overexpression of these genes and were transferred to a co-culture system with murine embryonic fibroblasts. iPSC colonies were identified after around 2 weeks in this culture system, plated onto Matrigel-coated plates (BD Biosciences) and maintained with mTeSR1 media (Stem Cell Technologies). All cell lines were regularly tested for mycoplasma contamination. Experiments were performed with one or two independent iPSC clones (for details see Supplementary Table 7).

**Retroviral lineage tracing and neuronal differentiation.** Replication-incompetent retrovirus was produced as described in Supplementary Methods. For retroviral lineage tracing, human NSCs were plated in a density of 5 × 10^4 cells per cm² on coverslips or dishes coated with poly-ornithine (10 μg ml⁻¹, Life Technologies) and laminin (5 μg ml⁻¹, Life Technologies)-coated surfaces. At day 16, emergent rosette structures were manually picked, and, gene expression profiles across all 65 samples. This measure was used to compute the module significances as the average gene significance measure for all genes in a given module (Fig. 1f).

**DTW.** Expression dynamics of time-related modules were compared using a DTW-based approach to calculate the distance between ASD and control time series gene expression data. In brief, we used the expression profiles of the ME calculated by WGCNA and applied smoothing to enhance time resolution. The DTW package was then used to create a distance measure by locally compressing or stretching (warping) the query ME profile to best match the reference ME profile. The more warping that is required, the greater the difference between the two time series. This measure was used to compute the module significances as the average gene significance measure for all genes in a given module (Fig. 1f).

**Gene significance measures and network visualization.** Gene significance measures were used to produce a network overview of the co-expression network. These measures were defined as correlations between gene expression values and trait-specific cellular stage (time point). The advantages of using correlations to quantify trait relationships are that this measure can take on either positive or negative values (−1, 1), and that a P value is readily computed. Thus, gene significance indicates whether the gene has a positive or negative relationship with the trait-specific time point. In combination with kME, gene significance was used to assess stage-specific expression or suppression of intramodular genes. Central hub genes of TM1 (genes with the highest kME) were depicted using Cytoscape (version 3.3.0). In these plots, each point is a gene (or node) and gene–gene connections (or edges) are shown as light yellow or light red lines. Edge width values (gene–gene connection weight) are displayed with increasing thickness. The positioning of the nodes is done by first using an automatic layout and then by manually adjusting the positions so that gene names are visible. In the plot showing the hub genes of TM1, ASD risk genes are displayed in red and selected term memberships as a colored bar graph next to the node. Different point sizes correspond to the number of time points showing (≥0.25 log-fold change) for the genes displayed. Significantly suppressed hub genes at the NSC stage are depicted in a second plot. Here, a gradually increasing point size of nodes corresponds to gene-significance values at the control NSC stage.

**Gene ontology analysis.** Functional annotation was performed with the Database for Annotation, Visualization and Integrated Discovery Bioinformatics resource and The Gene Ontology Consortium. For differentially expressed genes and...
co-expression modules, the background was set to the total list of genes expressed in our dataset. In Database for Annotation, Visualization and Integrated Discovery Bioinformatics-based analysis, the reported P values are derived from the expression analysis systematic explorer score probability, a modified Fisher’s exact test that is more stringent than the standard Fisher’s exact test in examining gene lists66. Where indicated, significance of over-representation was adjusted to control the false discovery rate by means of the Benjamini–Hochberg procedure. The statistical significance threshold for all gene ontology analyses was P < 0.01.

Immunocytochemistry. Cells were washed in PBS, fixed with 4% paraformaldehyde for 20 min at room temperature and blocked with 5% horse serum in PBS containing 0.1% Triton X-100. Primary antibodies were incubated overnight at 4 °C, washed three times with PBS, incubated with secondary antibodies for 2 h at room temperature, stained with DAPI solution and mounted in DABCO mounting solution (Sigma Aldrich). For confocal imaging, cells were grown on ibidi µ-slides. Antibodies and representative concentrations are listed in Supplementary Methods.

Morphometric analyses. Tiled image stacks were processed using ZEN Imaging software (Carl Zeiss Microscopy) and exported as TIFF files. Neuroacula version 11 (MBF Bioscience) was used to reconstruct individual retrovirally labeled newborn neurons over time. In total, 10–20 traces per line (eight ASD and five control lines with one or two clones per line, n = 564 cells) were analyzed at three different time points (4, 7 and 14 dpi) and plotted as mean for every individual. Identification of individual neurons was based on co-expression of the markers PSA-NCAM and DCX. Reconstructions were based on the extent of membrane-bound GFP expression labeled through RV-CAG:LckN-eGFP. IPSC-iNs were traced at the time points indicated post-induction (4, 7 and 14 days post-doxycycline). In total, 10–30 traces per line (eight ASD and five control lines with one or two clones per line, n = 746 cells; Supplementary Table 7) were analyzed and plotted as mean for every individual. For Sholl analysis, 5 or 10 µm increments were used to define the gradually increasing radius of concentric circles centered at the centroid of the cell soma. Subsequent analysis of global and subtle branching patterns was performed using Neuroacula Explorer (MBF Bioscience) and Microsoft Excel.

Generation of cerebral organoids. Cerebral organoids were generated and processed for analysis as described previously24 with minor modifications. Briefly, iPSCs were washed with Dulbecco’s phosphate-buffered saline (Invitrogen) and dissociated with EDTA followed by Accutase to generate single cells. A total of 1 × 10⁴ cells were then seeded into each well of an ultra-low-attachment 96-well plate (Nunc) to form single embryoid bodies in medium containing DMEM/F12, 20% KSR, 3% fetal bovine serum, 2 mM GlutaMAX, 1% non-essential amino acids, 30 mM β-mercaptoethanol and 4 mg ml⁻¹ β-mercaptoethanol (Sigma). Embryoid body colonies were grown for another 3–4 days. Medium was then changed to neural induction medium, consisting of DMEM/F12, 1× N2 supplement, 1% non-essential amino acids, 2 mM GlutaMAX and 1 mg ml⁻¹ heparin (Sigma). On days 10–12, embryoid bodies were subsequently cultured in cerebral differentiation medium comprising 50% DMEM/F12, 50% Neurobasal, 0.5× N2 supplement, 1× B27 (without RA) supplement, 2 mM GlutaMAX, 2.8 mg ml⁻¹ human insulin (Sigma), 0.5% non-essential amino acids and 55 mM β-mercaptoethanol. Droplets were cultured in stationary condition in 6 cm suspension culture dishes (Corning) for 4 days, transferred to an orbital shaker (65–80 rpm) and cultured in cerebral differentiation medium + RA media for up to five weeks.

Generation of forebrain organoids and retrotinal labeling. Forebrain organoids were generated as described previously24 with minor modifications. Human iPSC colonies were detached before reaching confluence with collagenase Type IV (Gibco) and transferred to an Ultra-Low attachment 10 cm plate (Corning Costar), containing 10 ml hPSC medium, consisting of DMEM:F12 (Invitrogen), 20% Knockout Serum Replacer (Gibco), 1× Non-essential Amino Acids (Invitrogen), 1× 2-mercaptoethanol (Gibco), 1× GlutaMAX (Invitrogen), 10 mg ml⁻¹ FGF-2 (Peprotech) and KROCK inhibitor Y27632 (10 µM). Twenty-four hours later, the medium was replaced with induction medium containing hPSC media without FGF-2, 2 µM dorsomorphin (Cayman Chemical) and 2 µM A-083 (Stemgent). At day 3, the media were replaced with neural induction medium comprising DMEM:F12 (Invitrogen), 1× N2 Supplement (Invitrogen), 1× Non-essential Amino Acids (Invitrogen), 1× GlutaMAX (Invitrogen), 10 µg ml⁻¹ Heparin (Sigma), 1× Penicillin/Streptomycin (Gibco), 10 µg ml⁻¹ CHIR99021 (Cayman) and 1 µM SB-431542 (Stemgent). Seven days after induction, organoids were embedded in 20 µl Matrigel (Trevigen) droplets and continued to grow for an additional week in 6-cm Ultra-Low attachment plates (Corning Costar). From day 14 onwards, organoids were cultured in differentiation medium comprising DMEM:F12 (Invitrogen), 1× N2 and B27 Supplements (Invitrogen), 1× Non-essential Amino Acids (Invitrogen), 1× GlutaMAX (Invitrogen), 1× 2-Mercaptoethanol (Gibco), 1× Penicillin/Streptomycin (Gibco) and 2.5 pg ml⁻¹ Insulin (Sigma), and transferred to an orbital shaker (65–80 rpm). At day 20, residual Matrigel was removed and medium changes were performed every 2–3 days using the average medium composition described above. Cultured for up to 10 weeks.

Forebrain organoids were stained with TRITC at 6 weeks to define CP-like regions and to measure the thickness of the evolving plate. Four independent measurements per ventricular structure of sub-segmented fields (specified in Supplementary Fig. 6; g) were averaged, and a total of three forebrain organoids with two 3′VZ-like regions per organoid were quantified for every subject line using three ASD and three control individuals (n = 17 organoids; see Supplementary Table 7 for details). Retrovirally labeled newborn neurons were analyzed in forebrain organoids at 14 dpi in regions that showed clearly defined VZ-like and CP-like structures. Three-dimensional reconstructions were based on 4× magnification single-plane GFP images in newborn neurons located within cortical plate-like regions and analyzed with the software detailed above. Immunohistochemical analysis (Supplementary Fig. 7) was performed on forebrain organoids with three control lines with three organoid replicates per individual, n = 264 cells; Supplementary Table 7) were analyzed and plotted as mean for every individual.

Assay for ATAC-seq. A total of 70,000 FACS-purified NSCs were lysed in 50 µl lysis buffer (10 mM Tris- HCl pH 7.5, 10 mM NaCl, 3 mM MgCl₂, 0.1% IGEPA, CA-630, in water) on ice and nuclei were pelleted by centrifugation at 500 g for 10 min. Nuclei were then resuspended in 50 µl transposase reaction mix (1× Tagment DNA buffer (Illumina 15027866), 2.5 µl Tagment DNA enzyme I (Illumina 15027865), in water) and incubated at 37 °C for 30 min. DNA was then purified with Zymo Chip DNA concentrator columns (Zymo Research D5205) and eluted with 10 µl of elution buffer. DNA was then amplified with PCR mix (1.25 µM Nextera primer 1, 1.25 µM Nextera index primer 2-bar code, 0.6x SYBR Green I (Life Technologies, S7563), 1× NEBNext High-Fidelity and 2× PCR MasterMix (NEB0541)) for 8–12 cycles, size selected for fragments (160–500 bp) by gel extraction (10% TBE gels, Life Technologies EC62752BOX) and 125 bp paired-end sequenced on an Illumina HiSeq 2500 platform.

ATAC-seq pre-processing and peak calling. FASTQ sequencing files from ATAC-seq experiments were mapped to the University of California, Santa Cruz genome (hg38) using Bowtie2 with default parameters25. For all data files, optical and PCR duplicates were removed using PicardTools. De-duplicated BAM files from the same sample were merged and PicardTools was again used to remove duplicate reads on the uniquely mapped properly paired reads. Reads falling into the mitochondrion, unmapped contigs and blacklisted regions (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/) were removed, resulting in more than 55 M high-quality reads on average. All samples showed the characteristic fragment length distribution, with more than 70% of reads falling within the nucleosome-free region and a mono-nucleosomal peak representing reads that contain a single nucleosome (180–247 bp in length). For peak calling of open chromatin, reads were filtered using the depeaks²⁶ command ‘alignment.Sieve.py’ with the following parameters: ‘-atACShift --minMappingQuality 30 --maxFragmentLength 100′, obtaining only properly paired reads within nucleosome-free regions (<100 bp). MACS2 (version 2.1.2.10160309) was then used to peak calls on filtered BAM files using the following command: ‘macs2 callpeak -g hs -q 0.05 –nomodel –keep-dup all –broad -f BAMPE’.

ATAC-seq data analysis. For differential accessibility analysis, we used DiffBind v2.4.8 and first performed a weighted overlap analysis (db.count). The number of reads within each peak of the union peak set was counted and normalized for sequencing depth using the total number of non-mitochondrial reads present in the peak for each sample. The edger v3.18.1 package was then used to contrast the control and ASD group using the DiffBind function ‘db.analyse’ with the settings for edger (method = DBA_EDGER). The differential accessibility heatmap (Fig. 8c) displays the normalized peak accessibility scores for each sample. These scores were calculated using the DiffBind default setting ‘DBA_SCORE_TMM_READS_FULL’, which uses the trimmed mean of M values normalization built into the edger package using read counts and full library size. Coverage plots (Fig. 8a and Supplementary Fig. 11f) display normalized read counts—the number of reads within a peak divided by the sequencing depth—and were generated with Gviz using hg38 NCBI RefSeq gene model annotations downloaded from the University of California, Santa Cruz genome browser.

To generate metagene profiles of promoter regions, HOMER was first used to convert aligned and filtered reads into ‘tag directories’ for further analysis27. For the
analysis of promoter accessibility, read counts for each sample were normalized to 10 million; the mean numbers of reads located ±1,500 bp of gene transcription start sites were then computed across the regions of the selected genes using HOMER's 'annotatePeaks' command, with the 'bin' option and a bin size of 10 bp. Statistics were calculated in a ±250 bp window surrounding the center of the transcription start sites.

Statistics. All information related to statistical tests is documented in the corresponding figure legends and/or in the main text and in the Supplementary Methods. No statistical methods were used to predetermine sample size, but our sample sizes are similar to those reported in previous publications that showed significance. No method of randomization was applied, but all biological assays and sequencing samples were randomly assigned to different experimental groups and batch order to control for covariates. For data collection and data analysis (counting, tracing, imaging and FACS analysis), investigators were blinded with regard to the group category. All biological assays were performed with iPSCs derived from eight ASD subjects and five healthy controls (n = 8 ASD subject lines, n = 5 control cell lines; all male), with the exception of the organoid validation experiments, for which we randomly selected a subset of three ASD subjects and three controls (n = 3 ASD subject lines, n = 3 control cell lines; Fig. 5 and Supplementary Fig. 6f–l). No other collected samples were excluded from the analysis. Statistical comparisons for biological assays were performed in Prism 6.0 (GraphPad Software) using the Mann–Whitney U-test (one independent variable, single comparison), one-way analysis of variance using Dunnett’s multiple comparisons test (to compare means of three or more sets of unpaired measurements) and two-way analysis of variance using the Sidak multiple comparison test (two independent variables: dpe and neurite length per neuron). Threshold for significance (α) was set at 0.05. Data distribution was assumed to be normal, but this was not formally tested. Means, median, s.d., s.e.m., percentages and confidence intervals of all other quantitative data were calculated using Prism 6.0 and Microsoft Excel.

Differential gene expression was calculated using Wald test statistics as implemented in the DESeq2 package6. Log-transformed RPKM values were used for transcriptome correlation analysis (Supplementary Figs. 2 and 8) using the Pearson correlation coefficient to measure the linear dependence of our samples against the BrainSpan dataset (see Supplementary Methods). Gene set enrichment analysis was performed using a hypergeometric test, and for gene ontology term enrichment a modified Fisher’s exact test was used7. For gene significance calculations and other WGCNA-related statistics, a two-tailed unpaired Student’s t-test was used and, where indicated, P values were adjusted using the FDR approach. For ATAC-seq, Hotelling’s t-square test was used to determine the difference in transcription start sites of the genes between ASD and control samples (Fig. 8f), considering n as the number of genes analyzed. Significance calculations for the enrichment of differential accessibility peaks within genomic regions were performed using the Genomic Association Test (GAT) randomization test8.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Code availability. The analysis routines and R packages used for data analysis are publicly available (http://bioconductor.org/), and the parameters and settings used are described in the Methods and Supplementary Methods. The following R packages were used for data analysis: DESeq2 v1.20.0, WGCNA v1.31, dtw v1.20-1, DiffBind v2.4.8, edger v3.18.1 and Gviz v1.20.0. Other packages used in this study include: STAR v2.5.2a, Caudapdt v1.6 with Python 3.4.1, FastQC v0.10.1, Bowtie2 v2.2.8, PicardTools MarkDuplicates v2.18.3-SNAPShot, samtools v1.6, deepTools2 v2.5.4, macs2 v2.1.1.20.160309, Homer, GAT v1.3.5, Cytoscape v3.3.0 and Prism v6.0 (Graphpad Software).

Data availability
RNA-seq datasets are available at EMBL-EBI ArrayExpress, with the accession code E-MTAB-6018. These raw data are associated with the following figures: Figs. 1, 2, 3, and 7d–f, and Supplementary Figs. 3c, 4a, 7a–d, 8f–g, and 10. Additional data generated or analyzed during this study are included in this published article and its Supplementary information tables. Supplementary tables are available for the following figures: Figs. 1, 2, 3, and 7d–f, and Supplementary Figs. 3c, 4a, 7a–d, 8f–g, and 10, as well as for additional differential expression analysis (Supplementary Tables 8 and 10). Supplementary Table 1 (Supplementary Data and Notes) provides additional information on all subject-derived iPSC lines used in this study. ATAC-seq datasets are available from the corresponding author upon request.

References
54. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872 (2007).
55. Marchetto, M. C. N. et al. A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell 143, 527–539 (2010).
56. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30, 2114–2120 (2014).
57. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
59. Yip, A. M. & Horvath, S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics 8, 22 (2007).
60. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics 24, 719–720 (2008).
61. Aach, J. & Church, G. M. Aligning gene expression time series with time warping algorithms. Bioinformatics 17, 495–508 (2001).
62. Giorgino, T. Computing and visualizing dynamic time warping alignments in R: the dtw package. J. Stat. Softw. 31, 1–24 (2009).
63. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
64. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57 (2009).
65. Ashburner, M. et al. The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. Nat. Genet. 25, 25–29 (2000).
66. Hosack, D. A., Dennis, G. Jr., Sherman, B. T., Lane, H. C. & Lempicki, R. A. Identifying biological themes within lists of genes with EASE. Genome Biol. 4, R70 (2003).
67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).
68. Ramirez, F. et al. deepTools2: a next generation web server for deep-seq data analysis. Nucleic Acids Res. 44, W160–W165 (2016).
69. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature 481, 389–393 (2012).
70. Halme, F. & Ivanek, R. Visualizing genomic data using Gviz and Bioconductor. Methods Mol. Biol. 1418, 335–351 (2016).
71. Heinz, S. et al. Simple combinations of lineages-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ The statistical test(s) used AND whether they are one- or two-sided
☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ A description of all covariates tested
☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
☒ Give P values as exact values whenever suitable.
☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated
☒ Clearly defined error bars
☒ State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
For data collection, the following software was used: Zeiss ZEN (black edition; Carl Zeiss Microscopy), Zeiss ZEN (blue edition; Carl Zeiss Microscopy), Neurolucida v11 (MBF Bioscience), StereoInvestigator v11 (MBF Bioscience) and the BD FACSDiva acquisition software (Becton-Dickinson).

Data analysis
The following R packages were used for data analysis: DESeq2 v1.20.0, WGCNA v1.51, DiffBind v2.4.8, edgeR v3.18.1, Gviz v1.20.0. Other packages and software used in this study included: STAR v2.5.2a, Cudadapt v1.6 with Python 3.4.1, FastQC v0.10.1, Bowtie2 v2.2.8, PicardTools MarkDuplicates v2.18.3-SNAPSHOT, samtools v1.6, deeptools2 v2.5.4, macs2 v2.1.1.20160309, Homer, GAT v1.3.5. and GraphPad Prism v6.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA-seq datasets are available at EMBL-EBI ArrayExpress with the accession code: E-MTAB-6018. This raw data is associated with the following figures: 1, 2a-c, 4i-k, S3c, S4, S7a-e, S8f-g and S10. Additional data generated or analyzed during this study is included in this published article and its supplementary information files. Supplementary tables are available for the following figures: 1, 2a-c, 4i-k, S3c, S4a, S7a-d, S8f-g and S10 as well as for additional differential expression analysis (Supplementary table 8). Supplementary table 1 (Supplementary Data and Notes) is providing additional information about all patient-derived iPSC lines used in this study. ATAC-seq datasets are available from the corresponding author upon request.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No statistical methods were used to predetermine sample size but our sample sizes are similar to those reported in previous publications and were estimated based on previous experience with a similar setup that showed significance (Marchetto et al., 2016).

Data exclusions

All biological assays were performed with iPSCs derived from eight ASD patients and five healthy controls (nASD=8, ncontrol=5; all male), with the exception of the organoid validation experiments, for which we randomly selected a subset of three ASD patients and three controls (nASD=3, ncontrol=3; Fig.3 and Extended Data Fig.6f-l). No other collected samples were excluded from the analysis.

Replication

Biological assays were performed with iPSCs derived from eight ASD patients and five healthy controls (nASD=8, ncontrol=5; all male), with the exception of the organoid validation experiments, for which we randomly selected a subset of three ASD patients and three controls (nASD=3, ncontrol=3; Fig.3 and Extended Data Fig.6f-l). In all experiments, two lines from each patient were prepared, and one or two lines were eventually used for the experiment. Replications were defined as such in the associated figure legends. All attempts of replication were successful.

Randomization

Cell culture experiments were not completely randomized, however all biological assays and sequencing samples were randomly assigned to different experimental groups and batches to control for covariates.

Blinding

For data collection, every cell line had a unique code that could not reveal the identity of the subject but could tell which lines belonged to the same subject, so that the investigators were blinded with regard to the group category. For data analysis (counting, tracing, imaging and FACS analysis), technicians and other investigators were blinded for group categories.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | Unique biological materials |
| ✗   | Antibodies |
| ✗   | Eukaryotic cell lines |
| ✗   | Palaeontology |
| ✗   | Animals and other organisms |
| ✗   | Human research participants |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ✗   | ChIP-seq |
| ✗   | Flow cytometry |
| ✗   | MRI-based neuroimaging |
Antibodies used
All antibodies used in this study were commercially available antibodies. FACS and flow cytometry were performed with the following antibodies: APC-conjugated PSA-NCAM (Miltenyi Biotec, cat.no. 130-093-273, 1:50) and APC-conjugated IgM Isotype control (Miltenyi Biotec, cat.no. 130-093-176, 1:50); all antibodies supplied in the BD Stemflow™ Human Neural Cell Sorting Kit (BD Biosciences, cat.no. 562271, 1:20). Validations and references at http://www.milenyi.com/US/en/products/macs-flow-cytometry/antibodies-primary-antibodies-anti-psa-ncam-antibodies-human-mouse-rat-2-2b-1-11.html#macs-for-B0-tests. All antibodies supplied in the BD Stemflow™ Human Neural Cell Sorting Kit (BD Biosciences, cat.no. 562271, 1:20). Validations and references at http://www.bd.com/us/applications/research/immunofluorescence/stem-cell-research/stem-cell-kits-and-60-skis/human-neural-cell-sorting-kit/p/562271. FBXO2 (abcam, cat.no. ab133171, 1:100). Validations and references at https://www.abcam.com/fbxo2-antibody-epr73282-ab133171-references.html. 647-conjugated SOX2 (BD Biosciences, cat.no. 562139, 1:25) and 647-conjugated IgG1, k Isotype Control (BD Biosciences, cat.no. 562139, 1:25). Validations and references at https://www.bd.com/us/applications/research/intracellular-flow/intracellular-antibodies-and-isotype-controls/anti-human-antibodies-alex-fluor-488-anti-sox2-30-678-p/562139. 488-conjugated SSEA-4 (BioLegend, cat.no. 330411, 1:20) and 488-conjugated IgG3, k Isotype Control (BioLegend, cat.no. 401323, 1:25). Validations and references at https://www.biolegend.com/en/us/products/plex-fluor-488-anti-human-ssea-4-antibody-4825. 647-conjugated TRA1-81 (BioLegend, cat.no. 330705, 1:20) and IgG1, k Isotype Control (BioLegend, cat.no. 401618, 1:200). Validations and references at https://www.biolegend.com/fr/ch/products/plex-fluor-488-anti-human-tra-1-81-antibody-4827.

Validation
FACS and flow cytometry antibodies: APC-conjugated PSA-NCAM (Miltenyi Biotec, cat.no. 130-093-273, 1:50) and APC-conjugated IgM Isotype control (Miltenyi Biotec, cat.no. 130-093-176, 1:50). Validations and references at https://www.milenyi.com/US/en/products/macs-flow-cytometry/antibodies-primary-antibodies-anti-psa-ncam-antibodies-human-mouse-rat-2-2b-1-11.html#macs-for-B0-tests. All antibodies supplied in the BD Stemflow™ Human Neural Cell Sorting Kit (BD Biosciences, cat.no. 562271, 1:20). Validations and references at http://www.bd.com/us/applications/research/immunofluorescence/stem-cell-research/stem-cell-kits-and-60-skis/human-neural-cell-sorting-kit/p/562271. FBXO2 (abcam, cat.no. ab133171, 1:100). Validations and references at https://www.abcam.com/fbxo2-antibody-epr73282-ab133171-references.html. 647-conjugated SOX2 (BD Biosciences, cat.no. 562139, 1:25) and 647-conjugated IgG1, k Isotype Control (BD Biosciences, cat.no. 562139, 1:25). Validations and references at https://www.bd.com/us/applications/research/intracellular-flow/intracellular-antibodies-and-isotype-controls/anti-human-antibodies-alex-fluor-488-anti-sox2-30-678-p/562139. 488-conjugated SSEA-4 (BioLegend, cat.no. 330411, 1:20) and 488-conjugated IgG3, k Isotype Control (BioLegend, cat.no. 401323, 1:25). Validations and references at https://www.biolegend.com/en/us/products/plex-fluor-488-anti-human-ssea-4-antibody-4825. 647-conjugated TRA1-81 (BioLegend, cat.no. 330705, 1:20) and IgG1, k Isotype Control (BioLegend, cat.no. 401618, 1:200). Validations and references at https://www.biolegend.com/fr/ch/products/plex-fluor-488-anti-human-tra-1-81-antibody-4827. The following antibodies were used for immunofluorescence: MAP2AB (Abcam, cat.no. ab5392, RRID: AB_2138153, 1:500), Validations and references at https://www.abcam.com/map2-antibody-ab5392.html. PSA-NCAM (EMD Millipore, cat.no. MAB5324, RRID: AB_95211, 1:500), Validations and references at http://www.emdmillipore.com/US/en/product/Anti-Polylysic-Acid-NCAM-Asymbio-clone-2-2B-MM_NF-MAB5324. Synapsin I (EMD Millipore, cat.no. MAB5324, 1:500), Validations and references at https://www.emdmillipore.com/US/en/product/Anti-SynapsinI-Rabbit-pAb;EMD_BIO-574778. Mouse anti-TUJ1/TUBB3 (BioLegend, cat.no. 801201, RRID: AB_2313773, 1:1000). Validations and references at https://www.biolegend.com/en-us/products/plex-fluor-488-anti-human-ssea-4-antibody-4825. 647-conjugated TRA1-81 (BioLegend, cat.no. 330705, 1:20) and IgG1, k Isotype Control (BioLegend, cat.no. 401618, 1:200). Validations and references at https://www.biolegend.com/fr/ch/products/plex-fluor-488-anti-human-tra-1-81-antibody-4827.

Eukaryotic cell lines
Policy information about cell lines
Cell line source(s) A detailed description of all reprogrammed iPSCs is listed in Supplementary Table 1. Additional information was published in Marchetto et al., Mol. Psychiatry (2016). The iPSC lines obtained from this cohort were reprogrammed in the same facility (Salk Institute for Biological Studies, Laboratory of Genetics) and under the same conditions. Briefly, fibroblasts were transduced with retroviruses containing SOX2, OCT4, KLF4 and MYC to induce overexpression of these genes and were transferred to a co-culture system with murine embryonic fibroblasts. iPSC colonies were identified after around two weeks.
in this culture system, plated onto Matrigel-coated plates (BD Biosciences) and maintained in mTeSR1 media (Stem Cell Technologies). Experiments were performed with 8 independent ASD patient lines (referred to as: Able, Avid, Aqua, Arch, Ahoy, Aero, Acai, Apex) and 5 independent control lines (Clue, Cent, Cove, Chap and Cent) with one or two clones per line. The above mentioned identifiers, accompanied by their respective clone ID, have also been used to allow sample identification in the deposited raw sequencing data.

Authentication

The characterization and authentication of these lines was recently performed in Marchetto et al., Mol. Psychiatry (2016). Here, analyses of patient blood cells, fibroblasts and iPSCs were performed using CNV analysis, whole-exome sequencing and G-banding analysis to authenticate and characterize patient-derived cell lines. For subsequent experiments, iPSC lines were grown separately and authentication was performed through the use of standardized names and clone-specific unique identifiers.

Mycoplasma contamination

Cell lines were regularly tested for mycoplasma contamination. All cell lines used for this study were tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics

IPSCs were generated from eight male subjects diagnosed with ASD according to the DSM-IV (average age 13.25 ± 5.67 years) and five male control participants (average age 11.4 ± 5.22 years), which had no history of psychological, genetic or other disorders (also shown in Supplementary Table 1). ASD subjects were selected from lists of ASD subjects with larger than normal average total brain volume (at 3.56 ± 0.67 years) as measured by MRI and compared with typically developing toddlers.

Recruitment

Subjects were recruited through the University of California San Diego, Autism Center of Excellence from a pool of volunteers formerly included in previous brain imaging studies. Control subjects were selected randomly from lists of typically developing individuals who had had the magnetic resonance imaging (MRI) scan when they were toddlers. ASD subjects were selected from lists of ASD subjects with larger than normal average total brain volume as compared with typically developing toddlers. The ASD subjects demonstrated a behavioral presentation consistent with autism as defined by the criteria set forth in the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV; APA). Assessment of the history and presence of the disorder was achieved via standardized behavioral, cognitive and functional assessments, including the appropriate Wechsler Intelligence Scale, the Autism Diagnostic Observation Schedule, the Autism Diagnostic Interview, Revised and Vineland Adaptive Behavior Scales. Participants in the control group had no history of psychological, genetic or other disorders. The recruitment criteria thus was specific to a subset of autistic patients, which show early signs of macrocephaly. We have highlighted this in the manuscript and abstract to specify the sampling of a highly stratified subset of macro-cephalic ASD patients.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

NSCs and maturing neuron cultures as well as iPSCs and iPSC-iNs were detached using accutase (Innovative Cell Technologies) and then washed and stained with fluorophore-coupled antibodies for 45 min at 4ºC in phosphate buffered saline (PBS). Cells were washed, suspended in PBS and filtered using a 40-μm cell strainer.

Instrument

Flow cytometry analysis was performed with a BD LSR II (Becton-Dickinson) and the BD InFlux Cytometer (Becton-Dickinson) was used for FACS-based purification of neuronal cell types.

Software

Flow cytometry data was analyzed using the FlowJo 10.4.2 software.

Cell population abundance

For flow cytometry analysis (measuring MFIs for reporter assays), a similar number of cells were analyzed for each replicate as indicated in the Supplementary Information (50,000 cells). Gates and histograms are shown for the populations analyzed and percentages for cell populations were indicated.

Gating strategy

Gating strategies for flow cytometry analysis are shown in the figures, including the respective isotype controls to determine the gates. Briefly, for FACS sorting we gated on APC-conjugated PSA-NCAM+ and GFP+ populations, which we determined in every experiment by using the same non RV-infected cell line stained with the appropriate APC-conjugated isotype control antibody. Gating for the flow cytometry analysis of reporter activity (hDCX and TCF/LEF) was performed similar by gating on APC-conjugated PSA-NCAM+ and GFP+ populations.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.