

<b>Supplementary Methods</b>	<b>2</b>
<i>Brain sample metadata</i>	2
<i>Library preparation</i>	3
<i>Read alignment and quality control</i>	4
<i>Definition of datasets for analysis</i>	5
<i>Quantification of gene expression</i>	5
<i>Identification and utilization of covariates for differential gene expression and differential alternative splicing</i>	5
<i>Analysis of lncRNAs during evolution and development</i>	6
<i>Quantification of transcript splicing and differential splicing analysis</i>	7
<i>Cortical patterning analysis</i>	7
<i>Transcription factor binding site analysis in the cortical patterning gene set</i>	8
<i>Experimental validation of SOX5 targets in human neurons</i>	8
<i>Module preservation analysis</i>	9
<i>Enrichment analysis of gene sets from exome studies</i>	9
<b>Supplementary Tables</b>	<b>10</b>
<b>Supplementary References</b>	<b>11</b>

## Supplementary Methods

### *Brain sample metadata*

Brain samples were obtained from 49 neurotypical controls and 48 individuals with ASD (39 idiopathic with no identified cause of autism and 9 with confirmed duplications in the 15q region, Extended Data Fig. 1a).

Individuals defined as autistic for this study had either a confirmed autism score based on the Autism Diagnostic Interview – Revised<sup>1</sup> (ADI-R, 30/48 individuals), duplication 15q syndrome with confirmed ASD (9/48), or a diagnosis of autism supported by other factors such as clinical history (9/48). Available metadata from brain banks included age, sex, medical history, and sample quality information. Variable levels of detail were available regarding medical case history and previous sample quality information. The medical case history from the brain bank was used to identify use of psychiatric medications, history of seizures, existence of co-morbidities, length of postmortem interval, results of neuropsychiatric testing, and cause of death where possible. Notably, medication status, seizures, ADI-R scores, and IQ test results were available only for individuals with ASD or dup15q syndrome, with 48/48, 48/48, 27/48, and 11/48 individuals having available information for these factors, respectively. History of medication and seizures were categorized as having a reported history as supported by medical information (Yes, 25/48) or not having mention of psychiatric medications or seizures (No, 23/48) despite other medical records being available.

Additional postmortem tissue metrics available in a subset of individuals included previously recorded RNA integrity number (RIN), pH, and brain weight. Notably, where it could be assessed, 88% of cortical and 91% of cerebellar RNA extractions were within 2 RIN values of previously documented RIN values, and 44% of cortical and 48% of cerebellar RNA extractions had better RIN values than what was previously documented. A comparison of pH between 14 case and 18 control samples revealed little correlation between RIN and pH ( $R^2 = 0.001$ ,  $P = 0.86$ ) or between diagnosis and brain weight ( $R^2 = 0.007$ ,  $P = 0.50$ ). Together, these data demonstrate that 1) elapsed time at the brain bank had minimal effect of RNA quality despite some brains being stored for many years, 2) pH has a minimal effect on RNA quality, and 3) postmortem brain weight was not different between ASD and controls.

Cause of death was categorized based on agonal state as previously described in a study assessing relative effects of multiple factors on postmortem gene expression (S = short, I = intermediate, P = prolonged)<sup>2</sup>. The main difference between cases and controls in cause of death was that cases have causes of death related to seizures, while controls do not. Additionally, cause of death information was much more complete for cases compared to controls. Otherwise, the proportion of individuals with different agonal state classifications was similar between groups. Finally, some samples came from individuals with a diagnosis of ASD whose primary cause of death was cancer. We performed an analysis of cortical regions where we removed all samples from individuals with cancer and re-analyzed DGE, and found similar results ( $R^2 = 0.95$ ). We also compared the samples from cancer against all other ASD samples, and found no DGE signal, demonstrating that a cancer diagnosis and associated therapies did not significantly affect our findings. Complete and quantitative data types were used as described below in covariate analyses while variables with missing values were assessed to evaluate their effect on observed relationships where appropriate. Relevant phenotypic and sample quality information used in the analysis is provided in Supplementary Table 1.

Up to three brain regions from each individual were assessed in this study: dorsolateral or medial prefrontal cortex (frontal cortex, FC, from Brodmann area 9, BA9), superior temporal gyrus (temporal cortex, TC, from Brodmann areas 41, 42, or 22, BA41/BA42/BA22), and cerebellar vermis. Not all regions were evaluated for each individual due to limitations in sample availability and tissue or RNA quality. Dissections, RNA extraction, and sequencing were randomized over age, sex, brain region, and diagnostic status, and this

information was blinded until gene and splicing quantification for the first two batches of data. The third batch contained a bias toward younger controls, and we accounted for this by performing initial analyses (Figures 1 and 2) with samples from the first two batches and by confirming a minimal contribution from batch effects in later analyses that included the third batch. Frozen brain regions acquired from the brain banks were dissected on dry ice in a dehydrated custom-built dissection chamber to reduce degradation effects from sample thawing or humidity. Approximately 50-100mg of tissue across the cortical region of interest was isolated from each sample using the miRNeasy kit with no modifications (Qiagen). For each RNA sample, RNA quality was initially quantified using the RNA Integrity Number (RIN)<sup>3</sup> on an Agilent Bioanalyzer.

### *Library preparation*

We performed pilot experiments and evaluated existing RNA-seq data from postmortem brain to select parameters for library preparation and RNA-seq. We compared both poly(A) tail selection for mRNAs (referred to as polyA+) and depletion of cytoplasmic and mitochondrial rRNAs (referred to as rRNA-depletion) for RNA-seq using existing data and a pilot experiment, and found that the quality of polyA+ RNA-seq drops off with RIN < 9, and considerably after RIN < 8, largely due to transcript degradation, resulting in a strong 3' bias that can introduce large variability in gene expression measurement across the transcript and across samples.

RIN values are rarely above 8 in frozen postmortem brain samples, so we opted for rRNA depletion with the RiboZero Gold kit (Illumina). This approach captures a large fraction of transcripts which are not polyadenylated<sup>4,5</sup>, and about 40% of reads are expected to align to intronic regions<sup>6</sup>. Due to the start date of library preparation and the need to keep a consistent protocol, we did not use strand-specific library preparation. We took several precautions to ensure intronic and antisense transcription did not confound our interpretation of protein coding genes and long noncoding RNAs (lncRNAs), which are discussed in more detail below. Additionally, we also explored sequencing read length, and evaluated 50bp and 100bp single-end and 50bp and 100bp paired-end (1x50bp, 1x100bp, 2x50bp, 2x200bp, respectively) RNA-seq through artificially trimming 2x100bp data in pilot samples. We found highly similar mapping and transcriptome quantification results using 2x50bp or 2x100bp RNA-seq, and opted to use 2x50bp RNA-seq to maximize the number of fragments given a constant average sequencing depth.

The general approach to our RNA sequencing and analysis is described in the Methods. In greater detail, ribosomal RNA was depleted from 1 µg total RNA with the Ribo-Zero Gold kit (Epicentre). Remaining RNA was then size selected with AMPure XP beads (Beckman Coulter) and resuspended in 8.5 µL of Illumina resuspension buffer and an additional 8.5 µL of 2x EPF buffer. Subsequent steps followed the Illumina TruSeq protocol (starting at page 84 of the sample prep v2 guide, no changes) to generate standard fragment sizes (100-300bp, mean 150bp). After this protocol was followed, libraries were quantified with the Quant-iT PicoGreen assay (Life Technologies) and validated on an Agilent 2200 TapeStation system. Libraries were pooled to multiplex 24 samples per lane using Illumina TruSeq barcodes, and each lane was sequenced six times on a HiSeq2000 or 2500 instrument using high output mode with standard chemistry and protocols for 50bp paired end reads. Notably, sample pooling was designed to purposefully minimize the difference across lanes in ASD diagnosis, age, sex, brain region, and RIN (by minimizing the adjusted R<sup>2</sup> for each of these factors across lanes). Libraries were prepared by one core facility and then sequenced at that facility and another at UCLA (batches 1 and 2) or at just one core facility (batch 3). After confirming that read quality was similar across both core facilities and lanes, reads from six lanes for each sample were pooled after demultiplexing (with CASAVA v1.8) for downstream analysis.

### Read alignment and quality control

The paired-end raw reads were mapped to the human reference genome assembly GRCh37.73<sup>7</sup> using Tophat2<sup>8</sup> as follows:

```
tophat -o outputfolder -g 10 -p 8 -r 99 --no-novel-juncs -G GRCh37.73.gtf
GRCh37bowtieindex pairedread1.fastq pairedread2.fastq
```

Aligned reads were sorted and alignments mapped to different chromosomes were removed from the BAM file using samtools<sup>9</sup>. We also called genotypes from RNA-seq data using a modification of an existing pipeline shown to detect SNPs optimally from RNA-seq data<sup>10</sup>. Genotypes reflecting sites that are heterozygous or homozygous for the minor allele relative to the reference genome were called using by removing duplications (*rmdup*) from .bam files and constructing .vcf files from piled up reads<sup>9</sup>:

```
samtools mpileup -I -S -gu -f GRCh37reference sorted_reads_rmdup.bam | bcftools
view -bvcg - > var.raw.bcf
bcftools view var.raw.bcf | vcutils.pl varFilter -D100 > var.flt.vcf
```

Genotypes were then coded as NA (homozygous for the major allele or not enough depth to detect), 1 (detected heterozygous), or 2 (homozygous for the minor allele). Pairwise spearman correlations were assessed between samples, and any sample from an individual that did not match the genotype of another sample from the same individual was assessed for contamination or sample mix-up and removed if the sample mix-up was not resolvable.

Further quality control (QC) analysis after read alignment was performed using PicardTools v1.100 (commands *ReorderSam*, *CollectAlignmnetSummaryMetrics*, *CollectRnaSeqMetrics*, *CollectGcBiasMetrics*) and samtools (duplication metrics and other statistics from the *flagstat* command). Sequencing metrics were used to remove samples with poor sequence quality based on the following sequencing metrics: %Total Reads, %High-quality Aligned Reads, %mRNA Bases, %Intergenic Bases, Median 5' to 3' Bias, GC Dropout, and AT dropout. To detect outliers, a quality z-score was calculated for each metric, and samples with low quality ( $Z > 2$  for %Intergenic Bases, GC Dropout, or AT Dropout and  $Z < -2$  for %Total Reads, %High-quality Aligned Reads, %mRNA Bases, or Median 5' to 3' Bias) in this matrix were identified as outlier values, and any sample with greater than one outlier value was removed due to sequencing quality concerns.

The QC analysis was performed with 310 initial samples: 28 samples were not used in this study (due to small sample size from different brain banks, co-morbid diseases, or complete confounding between brain bank and disorder), and the QC described above removed 31 samples (10%). Of these, about one-third had very low RIN ( $< 4$ ), one-fourth had a high 5' to 3' bias ( $> 0.7$ ), and one-third had a high proportion of reads aligned to intergenic regions (intergenic read proportion  $> 20\%$ ). The remaining 251 samples from 97 individuals along with RNA-seq QC metrics are shown in Supplementary Table 1 with summary statistics in Extended Data Fig. 1.

### Definition of datasets for analysis

Extended Data Figures corresponding to the individual analyses (Extended Data Figs. 1h, 7b, 8a, 10a) delineate the different subsets of the data used for different analyses in this study, and each table is summarized from Supplementary Table 1a. Briefly, initial exploratory data analysis revealed a strong dependence of global transcriptomic changes in a nonlinear manner with younger individuals (age  $< 10$ ), consistent with what has been observed in other studies that have evaluated the cortical or cerebellar transcriptome<sup>11,12</sup>. Therefore, in CTX we selected an age-matched subset, which after one round of gene expression based outlier removal using sample-sample correlations (at a signed Z score  $> 2$ )<sup>13</sup> yielded 43 ASD samples and 63 CTL samples from 26 ASD individuals and 33 CTL individuals (median age 30 for ASD, 28 for CTL). We then used the samples

from 9 individuals with dup15q syndrome in a separate analysis with an independent set of 10 CTL individuals. We combined these sets plus additional samples that were unused due to age matching for the co-expression network analysis in CTX and CB and performed one round of outlier analysis as described above with all samples included prior to defining the final sample sets. The specific samples used in each analysis are described in Supplementary Table 1.

### *Quantification of gene expression*

Gene expression levels were quantified for samples passing QC using multiple methods:

HTSeq (v.0.6.1) with a union exon model:

```
python -m HTSeq.scripts.count --stranded=no --mode=union --type=exon --quiet
inputfile.sam --GTF GRCh37.73.gtf >> outputfolder
```

HTSeq with a whole gene (union exon + introns) model:

```
python -m HTSeq.scripts.count --stranded=no --mode=union --type=gene --quiet
inputfile.sam --GTF GRCh37.73.gtf >> outputfolder
```

Cufflinks v2.1.1:

```
cufflinks -o outputfolder --num-threads 8 --GTF GRCh37.73.gtf --frag-bias-correct
GRCh37reference --multi-read-correct --compatible-hits-norm inputfile.bam
```

HTSeq simply counts paired read fragments on the given gene models and does not use reads with multiple alignments or that overlap gene models, while Cufflinks uses these multi-mapped reads, quantifies transcripts first, and then provides a gene level estimate. At a global level, all three approaches are highly correlated (Extended Data Figure 1e-g). The least concordance was between the “whole gene” model compared to the union exon or Cufflinks approaches. This is because the whole gene model includes introns, unprocessed RNAs, and possibly mis-spliced transcripts. For analyses, we used the HTSeq union exon quantifications, and applied the other methods in filtering criteria to improve data quality. Genes were kept if they passed each of the following criteria within all CTX and CB samples separately:

- Expressed in 80% of samples with HTSeq union exon quantification of 10 counts or more (to remove genes supported by only a few reads)
- Expressed in 80% of samples with HTSeq whole gene quantification 10 counts or more (removes genes supported solely by intronic reads of other genes)
- Expressed in 80% of samples with Cufflinks (lower bound FPKM estimate > 0)

The resulting read counts, reflecting fragments mapped to each union exon model, were converted to  $\log_2$  transformed and GC content, gene length, and library size normalized fragments per kilobase million mapped reads (FPKM) values using the cqn package in R (default options, but setting `cqn=FALSE` which turns off quantile normalization)<sup>14</sup>. Where used, these values are referred to as  $\log_2(\text{Normalized FPKM})$ .

### *Identification and utilization of covariates for differential gene expression and differential alternative splicing*

Normalized FPKM data were assessed for effects from biological covariates (condition, age, sex, brain region), technical variables related to sample processing (RIN, Brain Bank, Sequencing Batch), and technical variables related to sequencing quality metrics (sequencing metrics plus the proportion of exonic reads, as defined by total quantified reads in the union exon model divided by total quantified reads in the whole gene model for each sample). For technical variation, there was no strong relationship between any factor and diagnosis in the idiopathic ASD set, largely due to the fact that we randomized all samples at multiple points (dissection, RNA extraction, and multiplexing) over all biological covariates and balanced age. However, there was a strong relationship between the first principal component of gene expression and RNA and sequencing quality metrics, including the RIN, sequencing 5’-3’ bias, and the proportion of exonic reads. Given the large number of

sequencing quality features, we performed principal components analysis (PCA) on these data and found that the first two PCs on the unstandardized features explain nearly 99% of the variance. Consequently, we opted to use these 2 sequencing surrogate variables (seqSVs) as covariates. The final linear mixed effects model was implemented using the R package lme as follows:

```
lme(expression ~ diagnosis + age + sex + region + sequencing.batch +
      brain.bank.batch + RIN + seqSV1 + seqSV2, rand = ~1|individualID)
```

Where all categorical variables were coded in a binary manner, seqSV1 and seqSV2 are the 1<sup>st</sup> and 2<sup>nd</sup> PCs of the sequencing statistics, and individualID is a random effect accounting for the fact that multiple samples come from the same individual. Principal components analysis was performed with mean centering, but without variance normalization for seqSV and gene set PCA. Notably, we used a linear model without random effects for cerebellum, where we removed a handful of technical replicates from the analysis. Additionally, for dup15q DGE, we did not utilize sequencing batch or brain bank batch as covariates, as there were confounding effects between diagnosis, age, brain bank batch, and sequencing batch for this subset of the data.

To ensure robustness, we also explored additional approaches to modeling DGE: we added additional seqSVs, applied a permutation analysis to compute P values using the above model (randomizing case/control status of all samples belonging to a given individual together), used the above model with limma voom<sup>15</sup>, applied surrogate variable analysis<sup>16</sup> to compute 17 additional surrogate variables (SVs) and added them as covariates, and we used the above model with a negative binomial distribution<sup>17</sup>. The P value rankings of genes are concordant across most of these methods (Extended Data Fig. 1j), with the greatest difference being the surrogate variable analysis method. Most importantly, we confirmed that the DGE results were concordant with RT-PCR validations for a subset of transcripts (Extended Data Fig. 2c), demonstrating that our model accurately detects DGE in postmortem tissue.

To demonstrate independent replication of the ASD DGE signal without using any covariates, we used a machine learning approach where a model was trained to differentiate ASD and CTL on RNA-seq samples from Voineagu et al., 2011 data (the samples from the y axis of Extended Data Fig. 1i). We utilized the glmnet package in R to apply a regularized regression model (alpha = 0.5, an elastic net model), and trained a regularized regression model with 10-fold cross validation on these data and classified ASD and CTL status in independent samples (x axis in Extended Data Fig. 1i). We evaluated the ability to discriminate case and control status by the difference in classification score and the area under the receiver-operator characteristic curve (AUROC, Extended Data Fig. 2a).

When plotting gene expression levels, splicing event PSI values, evaluating cortical patterning, and constructing co-expression networks, we adjusted the data for technical covariates. We regressed out variability attributed to technical covariates by applying a linear model on the covariate matched samples (Extended Data Fig. 1h) with the covariates described above, computing the bootstrapped coefficients (resampling 1000 times for each gene), then removing the variability attributed to technical factors:

```
adjusted_value = original_value - sequencing.batch*beta.sequencing.batch -
      brain.bank.batch*beta.brain.bank.batch - RIN*beta.RIN - seqSV1*beta.seqSV1 -
      seqSV1*beta.seqSV2
```

Where beta refers to the computed coefficient for each technical variable. Bootstrapping the coefficient ensured the estimated coefficient was not driven by a subset of samples.

### *Analysis of lncRNAs during evolution and development*

We further evaluated lncRNA expression in independent gene expression data from multiple species<sup>18</sup> and human prefrontal cortical development<sup>19</sup>. For the evolutionary data, we compared our detected transcripts with transcript ages mapped to human by intersecting lncRNA annotations with bedtools<sup>20</sup> bedIntersect (using

Supplementary Dataset “AllInfo\_LncRNAs\_MainDataset\_NonStrandSpecific\_Human.txt”). 331 unique entries intersected with our Gencode v18 annotated lncRNAs that were expressed in cortex of gene biotype antisense, lincRNA, or processed\_transcript. 239/331 (72%) were categorized as having a transcript age of “GreatApe”, “AfricanApe”, or “Primate”, suggesting they are not transcribed in non-primates and are therefore primate-specific. 19/331 lncRNA transcripts were DGE at  $FDR < 0.05$ , and out of these 15/19 (79%) are primate-specific in terms of sequence age. Thus, a large fraction of the lncRNAs dysregulated in ASD CTX is primate-specific, and primate-specific lncRNAs are not more likely to be DGE than other lncRNAs. One caveat to this analysis is that the gene expression platforms were markedly different in the studies (the evolutionary study used polyA+ RNA-seq) and therefore many lncRNAs found in our data were not detected. For developmental data, we utilized polyA+ transcriptome RPKM quantifications from prefrontal cortex spanning prenatal life to late age<sup>19</sup>, categorized as “Fetal”, “Infant”, “Child”, “Teen”, “Adult”, and “50+” (6 samples in each group). We classified transcripts as developmentally regulated if they passed  $FDR < 0.05$  for multiple test corrected ANOVA P values across the six groups. 28 of the 60 show enriched expression in brain relative to other tissues. Additional information, including interactions with miRNA binding complexes and FMRP can be found in Supplementary Table 2.

#### *Quantification of transcript splicing and differential splicing analysis*

In addition to applying Multivariate Analysis of Transcript Splicing (MATS, v3.08)<sup>21</sup> as described in the Methods, we also computed PSI values with OLego and Quantas<sup>22</sup>. Using PSI values from these two different methods, the linear mixed effects model yielded similar results in differential splicing analysis (Extended Data Fig. 4d). We also confirmed that our model accurately identifies differences in alternative splicing by confirming a high correlation between the findings from our model and splicing changes in a representative subset of samples with RT-PCR (Extended Data Fig. 4g).

For additional splicing analyses, we processed splicing data from independent studies that were evaluated with different alternative splicing programs and pipelines<sup>23-27</sup>. For mouse data (*Rbfox1* KO), we first used liftOver to convert mm9 coordinates to hg19, and then used bedIntersect to intersect the alternatively spliced exons from MATS with those from other datasets. For human datasets (*SRRM4* overexpression, *PTBPI* knockdown), we used the latter step to convert exon coordinates from each dataset to our MATS events. When performing over-representation analyses with these datasets, we restricted the background set of events to those that were successfully converted after the liftOver and bedIntersect steps, and used a two-sided Fisher’s exact test.

#### *Cortical patterning analysis*

The aim of the cortical patterning analysis was to identify any differences in genes that distinguish cortical regions in ASD compared to CTL. We first identified which individuals had both a FC and TC sample for evaluation, and identified 39 pairs corresponding to unique individuals (16 ASD, 23 CTL). We ensured that the number of individuals was similar in both cohorts, as we use a P value threshold for identifying the set of cortically patterned genes, and having a greater sample size in one group would artificially inflate the number of patterned genes. Given that there were 7 additional CTLs than ASD, we attempted to match age across both ASD and CTL, as regional patterning varies dramatically by age<sup>11</sup>. These criteria resulted in 16 pairs of ASD and 16 pairs of control samples (Supplementary Table 1). Differences between brain regions were evaluated with a paired Wilcoxon rank-sum test, which helps account for the variability across individuals. To demonstrate our findings are not dependent on previously evaluated samples<sup>28</sup>, we also confirmed attenuation in

patterning without using samples from Voineagu et al., 2011 by applying an unpaired Wilcoxon rank-sum test and comparing differences between FC and TC (Extended Data Fig 6b, sample details in Supplementary Table 2).

To utilize our full sample set and demonstrate attenuation of patterning with an independent method, we applied a machine learning approach where a model was trained to differentiate FC and TC on BrainSpan data. We utilized the *glmnet* package in R to apply a regularized regression model ( $\alpha = 0.5$ , an elastic net model), and trained a regularized regression model with 10-fold cross validation on BrainSpan FC (dorsolateral prefrontal cortex, DFC, medial frontal cortex, MFC) and TC (primary auditory cortex, A1C, and superior temporal cortex, STC) to discriminate FC and TC (Extended Data Fig. 6e) using all genes with a weak correlation to brain region (Extended Data Fig. 6f) and only genes demonstrating the greatest attenuation of patterning (Extended Data Fig. 6g). We then applied this model to our FC and TC samples in CTL and ASD, and evaluated the ability to discriminate brain region with each gene set by the difference in classification score and the area under the receiver-operator characteristic curve (AUROC, Extended Data Fig. 6h). This machine learning analysis enabled us to demonstrate that the attenuation of cortical patterning is seen in the full dataset without having to match or adjust any covariates.

#### *Transcription factor binding site analysis in the cortical patterning gene set*

Transcription factor binding site enrichment analysis was performed by scanning 1000bp upstream of the transcription start site for the genes in the attenuated cortical patterning set. Then, for each TF in the JASPAR database (<http://jaspar.genereg.net/>), we applied the following steps: 1) putative motifs bound by the TF were obtained from TRANSFAC database. 2) upstream sequences of genes in the attenuated cortical patterning set were scanned with the Clover algorithm<sup>29</sup> to calculate motif enrichment; and iii) enrichment above background was calculated using the MEME algorithm<sup>30</sup>. We compared enrichment for each TF motif in three different background data sets to ensure robustness: 1000 bp sequences upstream of all human genes, human CpG islands, and the sequence of human chromosome 20. We calculated P values by drawing 1,000 sequences of the same length and testing them for enrichment using MEME, and computed the P value based on the observed motif enrichment rank versus the randomized sets. We considered TF binding site predictions with significant motif enrichment across multiple backgrounds as high-confidence predictions. An online tool that serially executes the above steps for TF binding site enrichment for a given gene set is publicly available (<https://tfenrichment.semel.ucla.edu/>) and the P values for each background against the ACP set are reported in Extended Data Fig. 6k for each high-confidence TF. We chose to perform validation experiments for *SOX5* in particular due to the motif enrichment results in conjunction with our observation that *SOX5* exhibited attenuated patterning between FC and TC.

#### *Experimental validation of SOX5 targets in human neurons*

*SOX5* was overexpressed in primary human neuronal cultures using a lentiviral system. Briefly full-length human *SOX5* gene was cloned in pLVU/GFP vector (gift from Lars Ittner [Addgene plasmid #24177]) using the gateway recombination technique. Lentivirus was produced in HEK293T cells using a second generation packaging vector system (psPAX2, a gift from Didier Trono [Addgene plasmid #12260] and pCMV-VSV-G, a gift from Bob Weinberg [Addgene plasmid #8454]) as described by Stewart et al., 2003<sup>31</sup>. Primary human neurons were infected at plating at a multiplicity of infection (MOI) of 10 with either the *SOX5* overexpressing construct or a control pLVU-GFP backbone vector. 14 days after infection, RNA from the samples were isolated using miRNeasy micro kit (Qiagen, Carlsbad) and 50bp paired-end libraries were prepared using



SMARTer Stranded Total RNA sample prep kit (Clontech) with rRNA depletion. Libraries were then multiplexed and sequenced with HiSeq 2500 instrument (Illumina).

### *Module preservation analysis*

We also performed module preservation analyses<sup>32</sup> to demonstrate that modules found in the network analyses including ASD and CTL individuals could be found in network analysis of ASD individuals only and CTL individuals only (Supplementary Table 4). We also used module preservation analysis to show that CTX modules are found in CB, and vice versa (Supplementary Table 4).

### *Enrichment analysis of gene sets from exome studies*

Gene set enrichment analyses for different parts of the study were performed as described in the Methods. We note that we use logistic regression when controlling for covariates such as gene length and overall brain gene expression level. Methodologically, most studies performing gene set enrichment analysis use Fisher's exact test (or the hypergeometric test, which is equivalent to a one-sided Fisher's exact test) to evaluate whether a gene set is enriched over background, providing a P value and enrichment value for gene set enrichment. An assumption with such a test is that the background set is similar to the gene set considered for enrichment for all factors other than pathway membership. In the specific case of identifying genes affected by *de novo* variants in genes by whole exome sequencing, gene length alone is highly correlated to the mutation rate on that gene<sup>33,34</sup>. To correct for this gene length bias, Iossifov et al., 2014 utilized a stratified permutation analysis. This is equivalent to using gene length as a covariate. We therefore adopted a modified enrichment analysis using logistic regression. Logistic regression involves specifying predictors and a binary outcome and assessing how well the predictors explain the outcome. Beta values can be interpreted as odds-ratios (odds ratio =  $\exp(\text{beta value})$ ) and P values are equivalent to enrichment P values when the predictor is binary membership for one gene set membership (e.g. if a gene is in a given pathway or not) and the outcome is another gene set membership (e.g. if a gene is in a given module or not). The major benefit of this approach is that we can add gene length or overall gene expression in brain as a predictor, and remove these effects from the final gene set enrichment estimate.

## Supplementary Tables

**Supplementary Table 1 | Metadata for samples used in this study.** **a**, Brain sample metadata containing biological and technical information related to analyses. **b**, Data dictionary describing contents of the metadata spreadsheet. **c**, Subsets of samples used in the cortical patterning analysis (Figure 2, Extended Data Fig. 6).

**Supplementary Table 2 | Differential gene expression changes in cortex and cerebellum, cortical patterning results, and co-expression network module assignments.** **a**, Results for DGE between ASD and CTL in ASD and dup15q, as well as cortical patterning analysis results and co-expression network results for CTX. **b**, DGE results between ASD and CTL in CB and co-expression network results for CB. **c**, Validation of selected genes with quantitative real time PCR. **d**, Information on the differentially expressed ENSEMBL-annotated lncRNAs between ASD and CTL CTX (FDR < 0.05), including HGNC symbol (when available); log<sub>2</sub>(fold change) between ASD and CTL; summary notes of interest; sequence age, minimum transcript age, and maximum transcript age from Necsulea et al., 2014<sup>18</sup>; mean expression at fetal, infant, child, teen, adult and 50+ developmental epochs from Jaffe et al., 2015<sup>19</sup> with ANOVA P value (genes significant after a Bonferroni correction are highlighted); nuclear versus cytosolic expression in fetal and adult brain from Jaffe et al., 2015<sup>19</sup> with ANOVA P value (genes significant after a Bonferroni correction are highlighted) mean expression in stem cells at various stages from Xie et al., 2013<sup>35</sup> with ANOVA P value (genes significant after a Bonferroni correction are highlighted); mean expression in 16 diverse tissues from Farrell et al., 2014<sup>36</sup> with ANOVA P value (genes significant after a Bonferroni correction are highlighted); locus conservation information in mouse and zebrafish, and relevant literature from Incipedia.org<sup>37,38</sup>; lncRNA-protein interactions as well as conservation scores for promoter, start, end, exon, and intron across placental mammals, primates, vertebrates from Incrator.ewha.ac.kr<sup>39</sup>; miRNA-lncRNA interactions from <http://starbase.sysu.edu.cn><sup>40</sup>; disease associations from <http://www.cuilab.cn/lncnadisease><sup>41</sup>; chromosome, start, stop, strand, band, and biotype from Ensembl; log<sub>2</sub>(fold changes) and P values, and FDR adjusted P values for ASD vs CTL and dup15q vs CTL; the module to which the lncRNA was assigned in co-expression network analysis, and the module membership (kME) values for all modules in the network analysis.

**Supplementary Table 3 | Differential splicing changes in cortex and cerebellum.** **a**, Differential splicing results for cortex. **b**, Differential splicing results for cerebellum. **c**, Validation of selected splicing events with RT-PCR. **d**, Explanation of splicing event annotations. SE, skipped exon; A3SS, alternative 3' splicing site; A5SS, alternative 5' splicing site; MXE, mutually spliced exons.

**Supplementary Table 4 | Module preservation analyses and Gene Ontology term enrichment analyses.** **a**, Module preservation of CTX co-expression network modules in the ASD subset and CTL subset of the CTX data. **b**, Gene Ontology terms for CTX network analysis. **c**, Gene Ontology terms for CB network analysis. **d**, Module preservation of CTX co-expression network modules in CB. **e**, Module preservation of CB co-expression network modules in CTX.

## Supplementary References

1. Lord, C., Rutter, M. & Le Couteur, A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659–685 (1994).
2. Monoranu, C. M. *et al.* pH measurement as quality control on human post mortem brain tissue: a study of the BrainNet Europe consortium. *Neuropathol Appl Neurobiol* **35**, 329–337 (2009).
3. Schroeder, A. *et al.* The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol* **7**, 3 (2006).
4. Su, Z. *et al.* A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol* **32**, 903–914 (2014).
5. Li, S. *et al.* Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study. *Nat Biotechnol* **32**, 915–925 (2014).
6. Ameer, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18**, 1435–1440 (2011).
7. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol* **7**, S4–9 (2006).
8. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* **31**, 46–53 (2012).
9. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
10. Quinn, E. M. *et al.* Development of Strategies for SNP Detection in RNA-Seq Data: Application to Lymphoblastoid Cell Lines and Evaluation Using 1000 Genomes Data. *PLoS ONE* **8**, e58815 (2013).
11. Kang, H. J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
12. Colantuoni, C. *et al.* Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011).
13. Oldham, M. C., Langfelder, P. & Horvath, S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol* **6**, 63 (2012).
14. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
15. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).
16. Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLoS Genet.* **3**, e161 (2007).
17. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 31 (2014).
18. Necșulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014).
19. Jaffe, A. E. *et al.* Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat Neurosci.* **18**, 154–161 (2015).
20. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**, 11.12.1–34 (2014).
21. Shen, S. *et al.* MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. *Nucleic Acids Res* **40**, e61–e61 (2012).
22. Wu, J., Anczuków, O., Krainer, A. R., Zhang, M. Q. & Zhang, C. OLego: fast and sensitive mapping of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res* **41**, 5149–5163 (2013).
23. Raj, B. *et al.* A Global Regulatory Mechanism for Activating an Exon Network Required for Neurogenesis. *Mol Cell* **56**, 90–103 (2014).

24. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014).
25. Zhang, Y. *et al.* An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* **34**, 11929–11947 (2014).
26. Maze, I. *et al.* Critical Role of Histone Turnover in Neuronal Transcription and Plasticity. *Neuron* **87**, 77–94 (2015).
27. Gueroussov, S. *et al.* An alternative splicing event amplifies evolutionary differences between vertebrates. *Science* **349**, 868–873 (2015).
28. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
29. Frith, M. C. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* **32**, 1372–1381 (2004).
30. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28–36 (1994).
31. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).
32. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is My Network Module Preserved and Reproducible? *PLoS Comput Biol* **7**, e1001057 (2011).
33. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat Genet* **46**, 944–950 (2014).
34. Michaelson, J. J. *et al.* Whole-Genome Sequencing in Autism Identifies Hot Spots for De Novo Germline Mutation. *Cell* **151**, 1431–1442 (2012).
35. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
36. Farrell, C. M. *et al.* Current status and new features of the Consensus Coding Sequence database. *Nucleic Acids Res* **42**, D865–72 (2014).
37. Volders, P. J. *et al.* An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* **43**, 4363–4364 (2015).
38. Volders, P.-J. *et al.* LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res* **41**, D246–51 (2013).
39. Park, C., Yu, N., Choi, I., Kim, W. & Lee, S. lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics* **30**, 2480–2485 (2014).
40. Li, J.-H., Liu, S., Zhou, H., Qu, L.-H. & Yang, J.-H. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92–7 (2014).
41. Chen, G. *et al.* lncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res* **41**, D983–6 (2013).

Extended Data Figure 5a Source Data

