1 Global changes in patterning, splicing and primate specific lncRNAs in autism brain

2

3 Neelroop N. Parikshak[1,2,*], Vivek Swarup[1,*], T. Grant Belgard[1,2, †,*], Michael Gandal[1,2], Manuel Irimia[5,6],
4 Virpi Leppa[1], Jennifer K. Lowe[1], Robert Johnson[7], Benjamin J. Blencowe[6], Steve Horvath[3-4], Daniel H.
5 Geschwind[1-3]
6
7 1. Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of
8   California, Los Angeles, Los Angeles, CA 90095, USA.
9 2. Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen
10   School of Medicine, University of California Los Angeles, 695 Charles E. Young Drive South, Los
11   Angeles, CA 90095, USA.
12 3. Department of Human Genetics, David Geffen School of Medicine, University of California, Los
13   Angeles, California, USA.
14 4. Department of Biostatistics, David Geffen School of Medicine, University of California, Los Angeles,
15   California, USA.
16 5. EMBL/CRG Research Unit in Systems Biology, Centre for Genomic Regulation (CRG), 88 Dr.
17   Aiguader, Barcelona 08003, Spain.
18 6. Donnelly Centre, University of Toronto, 160 College Street, Toronto, ON M5S 3E1, Canada; Department
19   of Molecular Genetics, University of Toronto, 1 King's College Circle, Toronto, ON M5S 1A8, Canada.
20 7. NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School,
21   Baltimore, Maryland 21201, USA.

22 *These authors contributed equally to this study.
23
24 †Current address: MRC Functional Genomics Unit, Department of Physiology, Anatomy & Genetics,
25 University of Oxford, South Parks Road, Oxford, OX1 3PT, United Kingdom.
26
27

1   Summary

2   We apply transcriptome-wide RNA sequencing in postmortem autism spectrum disorder (ASD)
3   brain and controls and identify convergent alterations in the noncoding transcriptome, including primate
4   specific lncRNA, and transcript splicing in ASD cerebral cortex, but not cerebellum. We characterize an
5   attenuation of patterning between frontal and temporal cortex in ASD and identify *SOX5*, a transcription
6   factor involved in cortical neuron fate specification, as a likely driver of this pattern. We further show that a
7   genetically defined subtype of ASD, Duplication 15q Syndrome, shares the core transcriptomic signature of
8   idiopathic ASD, indicating that observed molecular convergence in autism brain is the likely consequence
9   of manifold genetic alterations. Using co-expression network analysis, we show that diverse forms of
10   genetic risk for ASD affect convergent, independently replicated, biological pathways and provide an
11   unprecedented resource for understanding the molecular alterations associated with ASD in humans.

1   Autism spectrum disorder (ASD) is a neurodevelopmental syndrome characterized by deficits in
2   social communication and mental flexibility[1]. Genetic risk factors contribute substantially to ASD risk, and
3   recent studies support the potential contribution of more than a thousand genes to ASD risk[2-4]. However,
4   given the shared cognitive and behavioral features across the autism spectrum, one hypothesis is that diverse
5   risk factors may converge on common molecular, cellular, and circuit level pathways to result in the shared
6   phenotype[5,6]. Analysis of the transcriptome has been used to identify common molecular pathways in the
7   cerebral cortex (CTX) from postmortem human brain tissue in individuals with ASD[7-11]. However, all
8   transcriptomic studies in ASD to date have been limited to evaluating highly expressed mRNAs
9   corresponding to protein coding genes. Moreover, most lack rigorous replication and do not assess gene
10   expression patterns across brain regions.
11   We used rRNA-depleted RNA-seq (Methods) to evaluate transcriptomes from a large set of ASD
12   and control (CTL) brain samples including neocortex (frontal and temporal) and cerebellum across 79
13   individuals (46 ASD, 33 CTL, 205 samples, Extended Data Fig. 1a-e, Supplementary Table 1). We first
14   compared differential gene expression (DGE) between ASD and CTL individuals in CTX from a previously
15   published[7] microarray study against new, independent gene expression profiles from RNA-seq to evaluate
16   global reproducibility of DGE in ASD. We found a high degree of replication of DGE fold changes between
17   the sample sets, despite evaluation on different gene expression platforms (fold changes at $P < 0.05$ in
18   previously evaluated data correlate with new data with $R^2 = 0.60$, Extended Data Fig. 1f). We observed a
19   much weaker overall signal and replication in cerebellum ($R^2 = 0.033$, Extended Data Fig. 1g). These
20   analyses confirm the existence of a reproducible DGE signature in ASD CTX across different platforms and
21   in independent samples.
22   We next combined samples from all individuals with idiopathic ASD into a covariate-matched
23   "ASD Discovery Set" (Extended Data Fig. 1h) for CTX (106 samples, 26 ASD, 33 CTL individuals) and
24   held out remaining samples for replication ("ASD Replication Set", Methods). For DGE analysis, we used a
25   linear mixed effects model that accounts for biological and technical covariates (Methods) to identify 1156
26   genes differentially expressed in ASD CTX, 582 increased and 574 decreased (Benjamini-Hochberg FDR ≤
27   0.05, Supplementary Table 2). Importantly, DGE analysis with additional covariates or different
28   assumptions about the distribution of the data and test statistics yielded similar results (Extended Data Fig.
29   2a). Additionally this DGE signature clusters over two-thirds of ASD samples together and this clustering is
30   not related to confounding factors such as cortical region, age, sex, and RNA quality (Figure 1a, Extended
31   Data Fig. 2b). The most significantly down-regulated gene was *PVALB* (fold change = 0.53, FDR ≤ 0.05), a
32   marker for GABAergic interneurons. *SST*, a marker for a different subpopulation of GABAergic
33   interneurons, is also among the most downregulated (fold change = 0.61, FDR ≤ 0.05). Other down-
34   regulated genes at FDR ≤ 0.05 include *NEUROD6,* involved in neuronal differentiation (fold change =
35   0.60), multiple ion channels, and *KDM5D*, a lysine demethylase (fold change = 0.66). In contrast, members
36   of the complement cascade implicated in microglial-neuronal interactions (*C4A*, fold change = 1.94; *C1QB*,
37   fold change = 1.65; both FDR ≤ 0.05) are upregulated in ASD CTX. Gene Ontology (GO) term enrichment
38   analysis further supports the involvement of pathways implicated by these genes (Figure 1b), confirming
39   previous findings[7]. Moreover, the upregulated set is enriched for astrocyte and microglia enriched genes,
40   and the down-regulated set is enriched for synaptic genes (Extended Data Fig. 2c), consistent with previous
41   observations[7,11].
42   We next sought to evaluate whether the transcriptional signature identified in the ASD Discovery Set
43   generalizes to the ASD Replication set by assessing the 1st principal component of the DGE set, which
44   summarizes the DGE expression pattern across all cortical samples. The ASD Discovery Set and ASD

1   Replication Set share this pattern, which is significantly different for both sets compared to CTL (Figure
2   1c). Moreover, this pattern is highly associated with ASD diagnosis, but not other biological factors,
3   technical factors, or scores on sub-domains of an ASD diagnostic tool (Figure 1d). These analyses
4   demonstrate that convergent differences in ASD CTX are reproducible in independent samples and are not
5   related to confounding factors.
6   We also detected 2715 lncRNAs expressed in cerebral cortex (after careful filtering for high-
7   confidence transcripts, Supplementary Information), of which 62 were significantly dysregulated between
8   ASD and CTL (33 long intergenic RNAs, lincRNAs; 19 antisense transcripts; and 10 processed transcripts
9   at FDR ≤ 0.05). Similar to the protein coding genes, these transcripts' expression patterns cluster ASD and
10   CTL samples (Figure 1e). Most of these lncRNAs are developmentally regulated[12], have chromatin states
11   indicative of transcription start sites (TSSs) near their 5′ end in brain[13], and are identified in other
12   datasets[12,14] consistent with being valid, functional lncRNAs. Moreover, most (81%) exhibit primate-
13   specific expression patterns in brain[15] (Supplementary Information). For example, Figure 1f depicts two
14   lincRNAs, *LINC00693* and *LINC00689*, which are typically downregulated during development, yet are
15   upregulated in ASD CTX relative to controls (Figure 1g), which we validated by RT-PCR (Extended Data
16   Fig. 2d). *LINC00693* sequence is present, but poorly conserved in mouse, while *LINC00689* is primate-
17   specific (present in macaque and other primates but not in any other species, Supplementary Information,
18   Extended Data Fig. 3 for additional examples). These data indicate that dysregulation of lncRNAs, many of
19   which are primate-specific and involved in brain development, is an important component of transcriptome
20   dysregulation observed in ASD.
21   Previous work suggested that alterations in transcript splicing may contribute to transcriptomic
22   changes in ASD[7,16,17] by evaluating splicing in a targeted manner and pooling samples across
23   individuals[7,16,17]. Given the increased sequencing depth and reduced sequencing bias across transcript length
24   in our dataset, we were able to perform an unbiased genome-wide analysis of differential alternative splicing
25   (AS). We evaluated the percent spliced in (PSI, Extended Data Fig. 4a) for 34,025 AS events in CTX across
26   the ASD Discovery Set, encompassing skipped exons (SE), alternative 5′ splice sites (A5SS), alternative 3′
27   splice sites (A3SS), and mutually exclusive exons (MXE) using the MATS pipeline[18] (Supplementary
28   Information). We first asked whether there was a global signal, finding significant enrichment over
29   background (Extended Data Fig. 4b). We identified 1127 events in 833 genes at FDR ≤ 0.5 in CTX (similar
30   to the number of events at uncorrected $P < 0.005$). Importantly, we obtained similar results with a different
31   splice junction mapping and quantification approach (Extended Data Fig. 4c).
32   We performed PCR validations with nine AS events from the differential splicing set (*ASTN2,
33   MEF2D, ERC2, MED31, SMARCC2, SYNE1, NRCAM, GRIN1, NCAM*) and found that validated changes in
34   splicing patterns were concordant with RNA-seq (Extended Data Fig. 4d-e), demonstrating that our
35   approach identifies alterations in AS with high specificity. Similar to our observations with lncRNA and
36   DGE, AS changes clustered the samples by diagnosis (Figure 2a). The most significantly different event
37   was the inclusion of an exon in *ASTN2* (ΔPSI = 5.8 indicating a mean of 5.8% difference in inclusion in
38   ASD vs CTL; $P = 7.8\times10^{-6}$), a gene implicated by copy number variation (CNV) in ASD and other
39   developmental disorders[19]. GO term analysis of the genes implicated by these pathways indicates
40   involvement of biological processes related to neuronal projection, biological adhesion, and morphogenesis
41   (Figure 2b), pathways where alternative isoforms are critical to specifying interactions between protein
42   products. Moreover, the 1st principal component of the cortex differential splicing signature replicates in the
43   ASD Replication Set and is not associated with other biological or technical factors (Figures 2c-d, Extended
44   Data Fig. 5a). Importantly, many splicing alterations occur in genes that are not differentially expressed

4

1   between ASD and CTL; removing AS events on genes exhibiting even nominal DGE (P < 0.05), still

2   identified a strong difference between ASD and CTL CTX (Extended Data Fig. 5b).

3   A parallel analysis in cerebellum evaluating 32,954 AS events found no differentially regulated

4   events significant at any multiple comparison correction thresholds (Extended Data Fig. 5c, Supplementary

5   Table 3). There was no detectable global overlap between cerebellum and CTX above chance for events

6   significant at P < 0.05 in both comparisons (fold enrichment = 1.1, P = 0.21). This suggests that AS

7   alterations in ASD are largely confined to CTX cell types, consistent with the stronger overall DGE patterns

8   observed in CTX versus cerebellum.

9   To further explore the underlying biology of AS dysregulation, we tested whether the shared splicing

10   signature in ASD might be a product of perturbations in AS factors known to be important to neural

11   development or preferentially expressed in neural tissue. We found that the expression levels of *RBFOX1*,

12   *RBFOX2*, *SRRM4*, *NOVA1*, and *PTBP1* all had high correlations ($R^2 > 0.35$, FDR ≤ 0.05) to AS alterations

13   in CTX (Figure 2e), but not in cerebellum (Figure 2f). Furthermore, enrichment analysis revealed that most

14   changes in cortical AS occur in neuron-specific exons that are excluded in ASD (exons with ΔPSI > 50% in

15   neurons overlap with exons excluded in ASD CTX, fold enrichment = 4.1, P = 1.8×10^{-7}, Extended Data Fig.

16   5d).

17   To validate a regulatory relationship between splicing factors and these events, we evaluated

18   experimental data from knockout, overexpression, and knockdown experiments for Rbfox1[20], SRRM4[21],

19   and PTBP1[22], respectively . We found that exons regulated by each of these splicing factors were

20   significantly enriched in the set of exons excluded in ASD (Figure 2g), while in contrast, there was no

21   enrichment for targets of ESRP[23], a splicing factor involved in epithelial cell differentiation but not

22   expressed in CTX. This shows that alterations in three splicing factors dysregulated in ASD regulate AS of

23   the neuron-specific exons whose inclusion is dysregulated in ASD in CTX and not cerebellum, indicating

24   selective alteration of neuronal splicing in ASD CTX. Remarkably, the expression patterns of these three

25   splicing factors (and others for which appropriate validation experiments were unavailable) results in

26   distinct clusters (Extended Data Fig. 5e), suggesting that subsets of splicing factors act in different

27   individuals to mediate a common downstream AS alteration.

28   Taken together these results indicate global transcriptional alterations in ASD cerebral cortex, but

29   not cerebellum at the level of protein coding transcripts, lncRNA and AS. Therefore, to determine how these

30   different transcriptomic subcategories relate to each other in ASD, we compared the 1st PC for each type of

31   transcriptomic alteration across individuals (Figure 2h). Remarkably, the PCs are highly correlated ($R^2 >$

32   0.8) indicating that the transcriptomic alteration is a unitary phenomenon across protein coding, noncoding,

33   and splicing levels, rather than distinct forms of molecular alteration.

34   Previous analysis with gene expression microarrays in a small cohort suggested that the typical

35   pattern of transcriptional differences between the frontal and temporal cortex may be attenuated in ASD[7].

36   To further test this possibility, we evaluated DGE between CTX regions (Supplementary Information) in 16

37   matched frontal and temporal CTX sample pairs from ASD and CTL subjects and found 551 genes

38   differentially expressed between regions in controls, but only 51 in ASD (FDR ≤ 0.05; Figure 3a). We refer

39   to the set of 523 genes with this pattern in CTL, but not ASD as the "Attenuated Cortical Patterning" set.

40   The attenuation of patterning is evident from the global distribution of test statistics between frontal and

41   temporal CTX in ASD and CTL and genes in this set do not show a greater difference in variability in ASD

42   versus controls compared to other genes (Kolmogorov-Smirnov test, two-tailed P = 0.11, Extended Data

43   Fig. 6a).

1   We complemented this analysis with a machine learning approach using all 123 cortical samples,
2   training a regularized regression model[24] to classify frontal versus temporal CTX with independent gene
3   expression data from BrainSpan[25] (Extended Data Fig. 6b, Supplementary Information). Multiple
4   approaches to training the classifier with BrainSpan can differentiate between frontal and temporal CTX in
5   both CTL and ASD (Extended Data Fig. 6c-e), demonstrating that dissection and sample quality in our
6   samples are of high quality. Loss of classification accuracy in ASD compared to CTL was observed when
7   restricting the model to the genes with the most attenuated patterning in ASD (Extended Data Fig. 6f),
8   demonstrating that attenuation of patterning generalizes across all samples. The Attenuated Cortical
9   Patterning set includes multiple genes known to be involved in cell-cell communication and cortical
10   patterning, such as *PCDH10*, *PCDH17*, *CDH12, MET*, and *PDGFD*, which was recently shown to mediate
11   human specific aspects of cerebral cortical development[26]. GO term enrichment analysis of the Attenuated
12   Cortical Patterning set identified enrichment for G protein coupled signaling, Wnt receptor signaling, and
13   calcium binding, among several developmental processes (Figure 3b), and cell type enrichment analysis did
14   not identify a strong preference for a particular cell type (Extended Data Fig. 6g).
15   To identify potential drivers of the alteration in cortical patterning, we evaluated transcription factor
16   binding site enrichment upstream of genes in the Attenuated Cortical Patterning set (Supplementary
17   Information), and found an enrichment of *SOX5* binding motifs (upstream of 364/523 genes, Figure 3c).
18   Remarkably, *SOX5* itself belongs to the Attenuated Cortical Patterning set: while *SOX5* is differentially
19   expressed between frontal and temporal CTX in CTL, it is not in ASD (Figure 3d). We thus predicted that if
20   *SOX5* regulates cortically patterned genes, its expression should correlate with target gene transcript levels.
21   Consistent with this prediction, we found that genes in the Attenuated Cortical Patterning set are anti-
22   correlated with *SOX5* in CTL CTX, but not in ASD CTX (Figure 3e, top left; Wilcoxon rank sum test of R
23   values, $P = 0.01$), suggesting that the normal role of SOX5 as a transcriptional repressor may be disrupted in
24   ASD. We reasoned that a true loss of SOX5-mediated cortical patterning would be specific to the predicted
25   SOX5 targets. Consistent with this, we find a loss of correlations between *SOX5* and predicted targets, but
26   no difference in correlations between *SOX5* and non-targets in the Attenuated Cortical Patterning set (Figure
27   3e). Taken together, these findings show that a loss of regional patterning downstream of the transcriptional
28   repressor *SOX5*, which plays a crucial role in glutamatergic neuron development[27,28], contributes to the loss
29   of regional identity in ASD.
30   Gene expression changes in postmortem brain may be a consequence of genetic factors,
31   environmental factors, or both. Brain tissue from individuals with ASD that harbor known, penetrant genetic
32   causes are very rare. However, we were able to identify postmortem brain tissue from 8 subjects with one of
33   the more common recurrent forms of ASD, Duplication 15q Syndrome (dup15q, which is present in about
34   0.5-1% of ASD cases, see Extended Data Fig. 7a for characterization of duplications). We performed RNA-
35   seq across frontal and temporal cortex and compared DGE changes in dup15q with those observed in
36   individuals with idiopathic ASD to better understand the extent to which the observed molecular pathology
37   overlaps. As expected, most genes in the 15q11.1-13.2 duplicated region have higher expression in dup15q
38   CTX compared to CTL (Figure 4a), although *SNRPN* and *SNURF* were notably downregulated. Conversely,
39   no significant upregulation of genes in this region were identified in idiopathic ASD or controls. Strikingly,
40   when we assessed genome-wide expression changes, we observed a strong signal of DGE in dup15q that
41   widely overlaps with that of idiopathic ASD (fold changes at FDR ≤ 0.05 in dup15q correlate with
42   idiopathic ASD with $R^2 = 0.79$, Figure 4b). Moreover, the slope of the best-fit line through these changes is
43   2.0, indicating that on average, the transcriptional changes in dup15q CTX are highly similar, but twice the
44   magnitude of those observed in ASD CTX.

1   Next, we sought to evaluate AS changes in dup15q. There is only one significant splicing change in
2   the dup15q region (Supplementary Table 3), consistent with the idea that duplication in this region
3   duplicates all isoforms of the genes, resulting in minimal alteration of transcript structure. Similar to DGE,
4   global AS analysis in dup15q CTX vs to CTL CTX revealed a stronger, but highly overlapping signature
5   with idiopathic ASD CTX (fold changes at FDR ≤ 0.2 in dup15q agree correlate with idiopathic ASD with
6    $R^2 = 0.66$) indicating that splicing changes in dup15q syndrome recapitulate those of idiopathic ASD
7   (Figure 4c). The slope of the best-fit line through the PSI for spliced exons in dup15q CTX compared to
8   those in ASD CTX is 2.5 similar to DGE. Notably, both gene expression and AS changes in dup15q
9   implicated similar pathways as those found in idiopathic ASD (Extended Data Fig. 7c-d). Clustering dup15q
10   samples and CTL samples using both the DGE set and the differential AS set showed that all dup15q
11   samples cluster together (Figure 4d), as opposed to the more variable clustering of idiopathic ASD,
12   supporting the hypothesis that this shared genetic abnormality leads to a more homogeneous molecular
13   phenotype.
14   Next, to test whether this molecular ASD signature may be due to independent of postmortem or
15   reactive effects (Supplementary Information), we compared our data with gene expression profiles from a
16   iPSC-derived neurons (nIPSCs)[29] from dup15q were available, we could use these data to definitively reveal
17   which changes in dup15q CTX are independent of postmortem or reactive effects (Supplementary
18   Information), since such effects are not present *in vitro*. We observe that DGE in the 15q region is
19   concordant with that seen in the nIPSCs (Figure 4e), even though the sample size is small and the analysis is
20   likely underpowered. Upregulated changes in dup15q are also seen in nIPSCs (Figure 4f), consistent with
21   our other statistical analyses showing limited effects of potential confounders. The very immature, fetal state
22   of the nIPSCs[30] likely explains the absence of an enrichment signal for genes downregulated in postnatal
23   ASD brain, which are enriched for genes involved in neurons with more mature synapses.
24   We next applied gene network analysis to construct an organizing framework to understand shared
25   biological functions across idiopathic ASD and dup15q (combining the ASD Discovery Set, ASD
26   Replication Set, and dup15q set). We utilized Weighted Gene Co-expression Network Analysis (WGCNA),
27   which identifies groups of genes with shared expression patterns across samples (modules) from which
28   shared biological function is inferred. Modules identified via WGCNA can than be related to a range of
29   relevant phenotypes and potential confounders[31,32]. We applied signed co-expression analysis and used
30   bootstrapping to ensure the network was robust, and not dependent on any subset of samples
31   (Supplementary Information), while controlling for technical factors and RNA quality ("Adjusted FPKM"
32   levels, Methods). WGCNA identified 16 co-expression modules (Extended Data Fig. 8a, Supplementary
33   Table 2), which are further characterized by their association to ASD (Extended Data Fig. 8b), enrichment
34   for cell-type specific genes (Extended Data Fig. 8c), and enrichment for GO terms (Extended Data Fig. 9).
35   Of the downregulated modules, three are associated with ASD and dup15q (M1/10/17) and one with dup15q
36   only (M11). Five of the upregulated modules are associated with ASD and dup15q (M4/5/6/9/12) and one is
37   specific to dup15q (M13) (Figure 5a, top). Additionally, we identified a module strongly enriched for genes
38   from the Attenuated Cortical Patterning set and *Wnt* signaling that contains *SOX5* (M12; fold enrichment =
39   3.0, P = $3x10^{-8}$), verifying the strong relationship observed between the *Wnt* pathway regulating TF *SOX5*
40   and attenuation of cortical patterning[33].
41   Notably, the modules identified here significantly overlap with previous patterns identified in ASD
42   (asdM12$_{array}$ and asdM16$_{array}$[7]; Figure 5a, middle). We found that the ASD-associated modules identified by
43   our larger sample size and RNA-seq provide significant refinement of previous observations by identifying
44   more discrete biological processes related to cortical development[34], the post-synaptic density[35], and

7

1   lncRNAs (Figure 5a, bottom). For example, M1 overlaps a subset of asdM12$_{array}$ (fold enrichment = 5.7)
2   and developmental modules (devM16 fold enrichment = 3.7), and is enriched for proteins found in the PSD
3   and genes involved in calcium signaling and gated ion channel signaling. Another subset of asdM12$_{array}$,
4   M10 (fold enrichment = 11) overlaps more with a mid-fetal upregulated cortical development module
5   (devM13 fold enrichment = 4.0), and genes involved in secretory pathways and intracellular signaling. A
6   third module, M17 shows the least overlap with asdM12$_{array}$ (fold enrichment = 2.2) and is related to energy
7   metabolism. Notably, these three modules are enriched for neuron-specific genes (Extended Data Fig. 8c),
8   but not all neuronal modules are down regulated in ASD (M3 is not altered in ASD CTX). Taken together,
9   specific neurobiological processes are affected in individuals with ASD related to developmentally
10   regulated neurodevelopmental processes.
11   The most upregulated modules, M5 and M9, both strongly overlap (fold enrichments > 20) with
12   previously identified upregulated co-expression module asdM16$_{array}$. M5 is enriched for microglial cell
13   markers and immune response pathways, whereas M9 is enriched for astrocyte markers and immune-
14   mediated signaling and immune cell activation (Extended Data Fig. 8c, Extended Data Fig. 9). This analysis
15   clearly separates the contributions of the coordinated biological processes of microglial activation and
16   reactive astrocytosis, which were previously not distinguishable as separate modules[7]. Thus, our analysis
17   pinpoints more specific biological pathways in idiopathic ASD than those previously identified and reveals
18   that similar changes occur downstream of the genetic perturbation in dup15q.
19   We evaluated the relationship between the five modules most strongly associated with ASD
20   (M1/5/9/10/17, which are supported by module-trait association analysis and gene set enrichment analysis,
21   Supplementary Information), and found that there was a remarkably high anti-correlation between the
22   eigengene of M5 and downregulated modules, particularly M1 (R^2 = 0.76) (Figure 5b). M1 (Figure 5c) is
23   downregulated in ASD and enriched for genes at the PSD and genes involved in synaptic transmission,
24   while M5 (Figure 5d) is enriched for microglial genes and cytokine activation. This strong anti-correlation
25   between microglial signaling and synaptic signaling in ASD and dup15q provides evidence in humans for
26   dysregulation of microglia-mediated synaptic pruning, as previously suggested[36].
27   Next, to determine the role of causal genetic variation, we evaluated enrichment of both rare genetic
28   variants, focusing on genes affected by ASD associated gene disrupting (LGD) *de novo* mutations[37], and
29   common variants[38,39]. Genes within three modules, M1, M3, and M12, show enrichment for common
30   variation signal for ASD (Figure 5e, Methods). Remarkably, M12 (Figure 5f), which is related to cortical
31   patterning and Wnt signaling, also exhibit GWAS signal enrichment, providing the first evidence that risk
32   conferred by common variation in ASD may affect regionalization of the cortex. Interestingly, M3 is
33   significantly enriched for both schizophrenia (SCZ) and ASD common variants, is related to synaptic
34   transmission, nervous system development, and regulation of ion channel activity (Extended Data Fig. 9),
35   consistent with the notion that ASD and SCZ share common and rare genetic risk[1,40-43].
36   We only identified one module, M2 (Figure 5g), as significantly enriched in protein disrupting
37   (nonsense, splice site, or frameshift) rare *de novo* variants previously associated with SCZ and ASD. M2
38   overlaps with a cortical developmental module implicated in ASD[34] (devM2 fold enrichment = 5.1).
39   Notably, M2 is not differential between ASD and CTL in our dataset, consistent with the observation that
40   these genes are primarily expressed during early neuronal development in fetal brain[34]. Remarkably, M2
41   contains an unusually large fraction of lncRNAs (15% of the genes in M2 are classified as lncRNAs, while
42   other modules are 1-5% lncRNA). We hypothesize that, in addition to protein coding genes involved in
43   transcriptional and chromatin regulation, rare *de novo* variants may also affect lncRNAs in ASD, a
44   prediction that will be testable once large sets of whole genome sequences are available.

1   These combined transcriptomic and genetic analyses reveal that different forms of genetic variation
2   affect biological processes involved in multiple stages of cortical development. Common genetic risk is
3   enriched in M3, M1, and M12, which reflect early glutamatergic neurogenesis, later neuronal function, and
4   cortical patterning, respectively. We also observe that rare *de novo* variation, which is enriched in M2,
5   affects distinct biology related to transcriptional regulation and chromatin modification. These findings are
6   consistent with transcriptomic analyses of early prenatal brain development and ASD risk mutations that
7   implicate chromatin regulation and glutamatergic neuron development[34,44].

8   We provide the first comprehensive picture of largely unexplored aspects of transcription in ASD,
9   lncRNA and alternative splicing, and identify a strong convergent signal in these, as well as protein coding
10   genes[7]. These results will aid in interpreting genetic variation outside of the known exome, as whole
11   genome sequencing supplants current methods. A role of lncRNAs has been previously explored in ASD[45],
12   but only two individuals were evaluated with targeted microarrays. We evaluate lncRNAs in an unbiased
13   manner across many individuals, notably identifying an enrichment of lncRNAs in M2, most of which are
14   uncharacterized in brain and arose on the primate lineage. The involvement of lncRNAs in this early
15   developmental program that is enriched for *de novo* mutations implicated in ASD suggests their study will
16   be particularly relevant to understanding the emergence of primate higher cognition on the mammalian
17   lineage, and by extension human brain evolution[15,46,47].

18   We also provide the first confirmation of an attenuation of genes that typically show differential
19   expression between frontal and temporal lobe in ASD CTX and further identified *SOX5*, known to regulate
20   cortical laminar development[50,51], as a putative regulator of this disruption. That M12, which is enriched for
21   genes exhibiting cortical regionalization and is also enriched in ASD GWAS signal, supports the prediction
22   that attenuation of patterning may be mediated by common genetic variation in or near the *SOX5* target
23   genes. Disruption of cortical lamination by direct effects on glutamatergic neurogenesis and function has
24   been predicted by independent data, including network analyses of rare ASD associated variants identified
25   in exome sequencing studies[34,44].

26   These data, in conjunction with previous studies, reveal a consistent picture of the ASD's emerging
27   postnatal and adult pathology. Specific neuronal signaling and synaptic molecules are downregulated and
28   astrocyte and microglial genes are upregulated in over 2/3 of cases. Microglial infiltration has been observed
29   in ASD cortex with independent methods[52], and normal microglial pruning has been shown to be necessary
30   for brain development[36]. Our findings further suggest that aberrant microglial-neuronal interactions may be
31   pervasive in ASD and related to the gene expression signature seen in a majority of individuals. In our
32   comprehensive AS analysis, we identify three splicing factors upstream of the altered splicing signature
33   observed in ASD CTX. These factors are known to be involved in coordinating sequential processes in
34   neuronal development[17,21] and maintaining neuronal function[48,49]. It may therefore be sufficient to disrupt
35   any one of these factors to induce a similar outcome during brain development, which would be consistent
36   with the shared downstream perturbation observed here.

37   Finally, evaluation of the transcriptome in dup15q supports the enormous value of the "genotype
38   first" approach of studying syndromic forms of ASD, with known penetrant genetic lesions[53]. It is highly
39   unlikely that the shared transcriptional dysregulation in dup15q is due to a shared environmental insult.
40   Thus, the most parsimonious explanation for the convergent transcriptomic pathology seen in all dup15q
41   and over 2/3 of the cases of idiopathic ASD is that it represents an adaptive or maladaptive response to a
42   primary genetic insult, which in most cases of ASD will be genetic[2,54]. As future investigations pursue the
43   full range of causal genetic variation contributing to ASD risk, these analyses and data will be valuable for
44   interpreting genetic and epigenetic studies of ASD as well as those of other neuropsychiatric disorders.
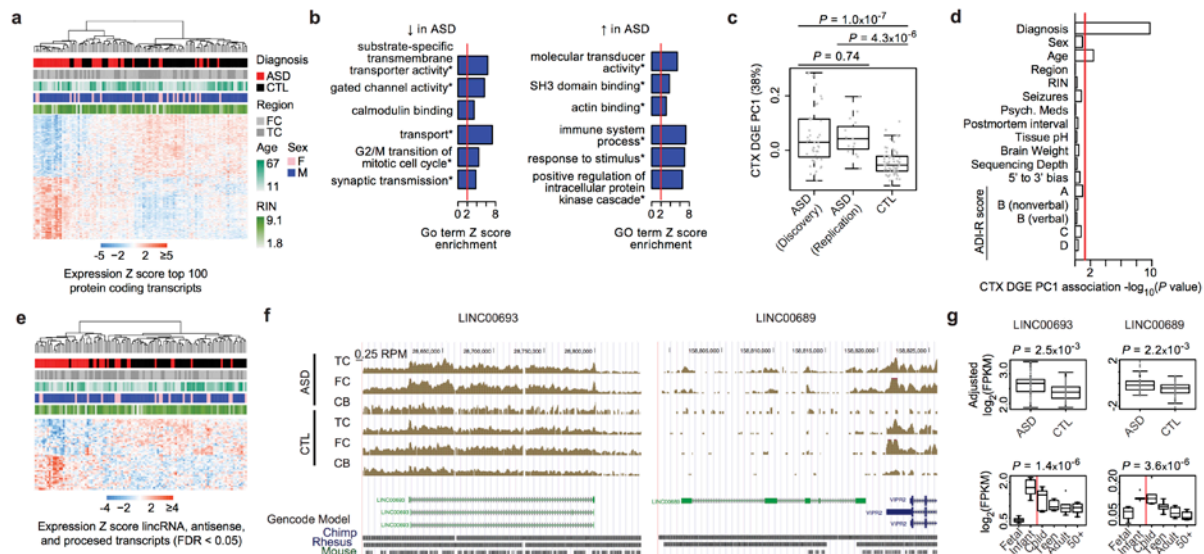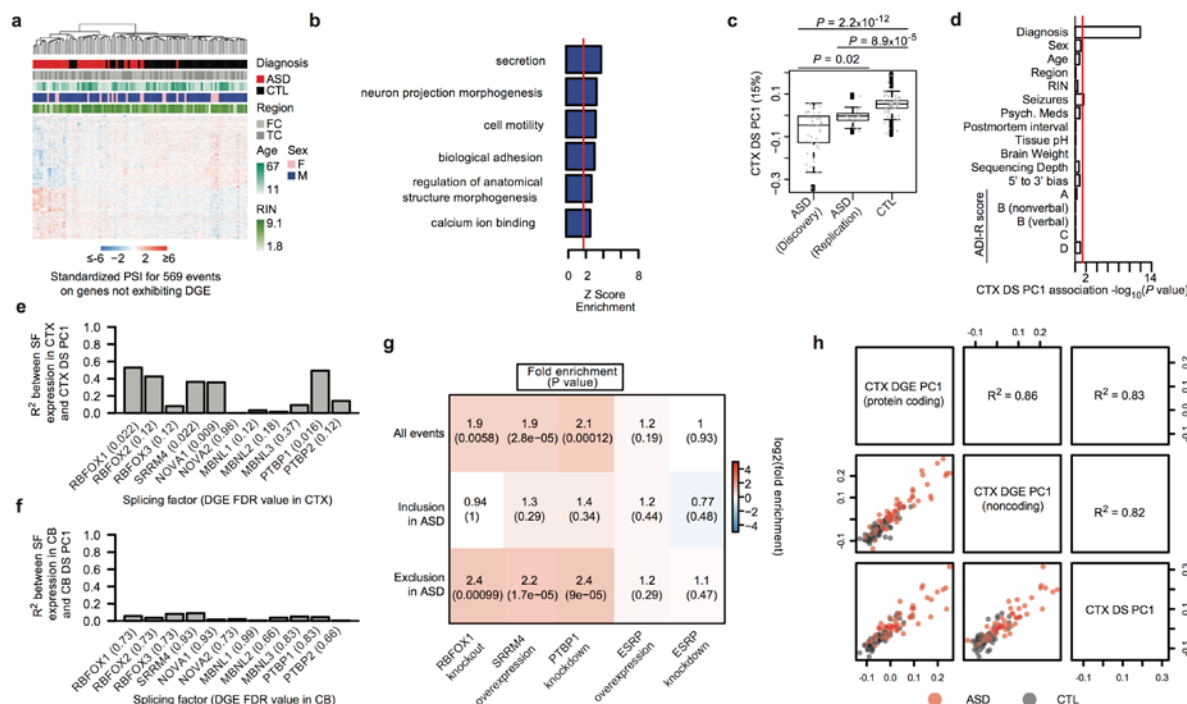
9

1 Figures & Figure Legends
2

3
4 Figure 1 | Transcriptome-wide differential gene expression in ASD. a, Average linkage hierarchical
5 clustering of samples in the ASD Discovery Set using the top 100 upregulated and top 100 downregulated
6 protein coding genes. b, Gene Ontology (GO) term enrichment analysis of upregulated and downregulated
7 genes in ASD. *FDR ≤ 0.05 across all GO terms and gene sets. c, 1st principal component of the CTX DGE
8 set (CTX DGE PC1) is able to distinguish ASD and CTL samples, including independent samples from the
9 ASD Replication Set. d, CTX DGE PC1 is primarily associated with diagnosis, and not other factors. e,
10 Average linkage hierarchical clustering of ASD Discovery Set using all lncRNAs in the DGE set. f, UCSC
11 genome browser track displaying reads per million (RPM) in a representative ASD and CTL sample,
12 superimposed over the gene models and sequence conservation for genomic regions including *LINC00693*
13 and *LINC00689*. g, *LINC00693* and *LINC00689* are upregulated across ASD samples and downregulated
14 during frontal cortex development. Abbreviations: FC, frontal cortex; TC, temporal cortex; RIN, RNA
15 integrity number; ADI-R score, Autism Diagnostic Interview Revised score; FPKM, fragments per kilobase
16 million mapped reads.
17

Figure 2 | Alteration of alternative splicing in ASD. a, Average linkage hierarchical clustering of ASD discovery set using top 100 differentially included and top 100 differentially excluded exons from the differential splicing (DS) set across the ASD Discovery Set. b, Gene Ontology term enrichment analysis of genes with DS in ASD. c, 1st principal component 1 of the CTX differential alternative splicing set (CTX DS PC1) is able to distinguish ASD and CTL samples using independent samples from the ASD Replication Set. d, CTX DS PC1 is primarily associated with diagnosis, and not other factors. e, Correlation between CTX DS PC1 and gene expression of neuronal splicing factors in CTX. f, Correlation between 1st principal component of cerebellum differential splicing (CB DS PC1) and gene expression of neuronal splicing factors in cerebellum. g, Overlap between DS set and splicing events regulated by splicing factors where experimental data was available. h, Scatterplots and correlations between the 1st principal component across the ASD versus CTL DGE sets for different transcriptome subcategories. Abbreviations: FC, frontal cortex; TC, temporal cortex; RIN, RNA integrity number; ADI-R score, Autism Diagnostic Interview Revised score; FPKM, fragments per kilobase million mapped reads.
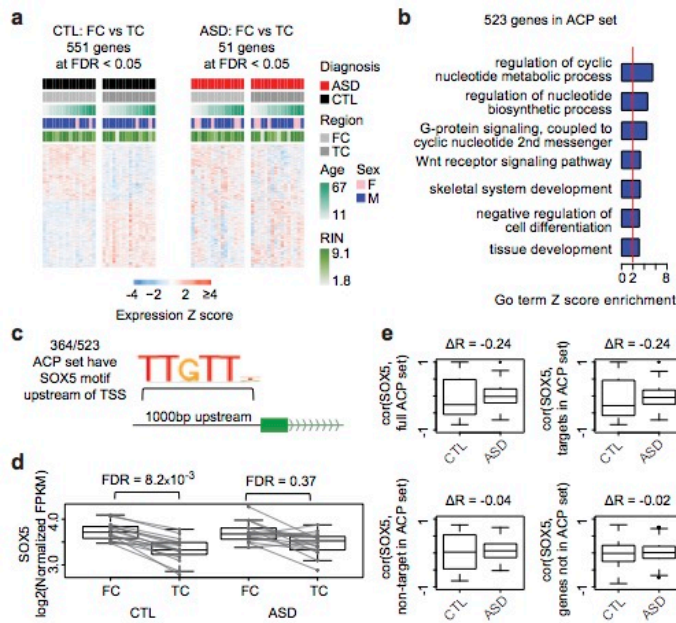
1
2 Figure 3 | Attenuation of cortical patterning in ASD cortex. a, Heatmap of 551 genes exhibiting cortical
3 patterning between frontal cortex (FC) and temporal cortex (TC) in ASD, with samples sorted by
4 diagnostic status and brain region. b, Gene ontology term enrichment analysis of genes exhibiting
5 attenuated cortical patterning (ACP). c, Schematic of transcription factor motif enrichment upstream
6 of genes in the ACP set, with the *SOX5* motif sequence logo. d, The *SOX5* gene exhibits attenuated
7 cortical patterning in ASD CTX compared to CTLs. Lines connect FC-TC pairs that are from the same
8 individual. e, Correlation between *SOX5* gene expression and predicted targets in CTL and ASD, with
9 all ACP genes (top left), SOX5 targets from the ACP set (top right), SOX5 non-targets from the ACP set
10 (bottom left), and all genes not in the ACP set (bottom right). Plots show the difference in correlation
11 between *SOX5* and other genes in ASD and CTL (ΔR).
12

13
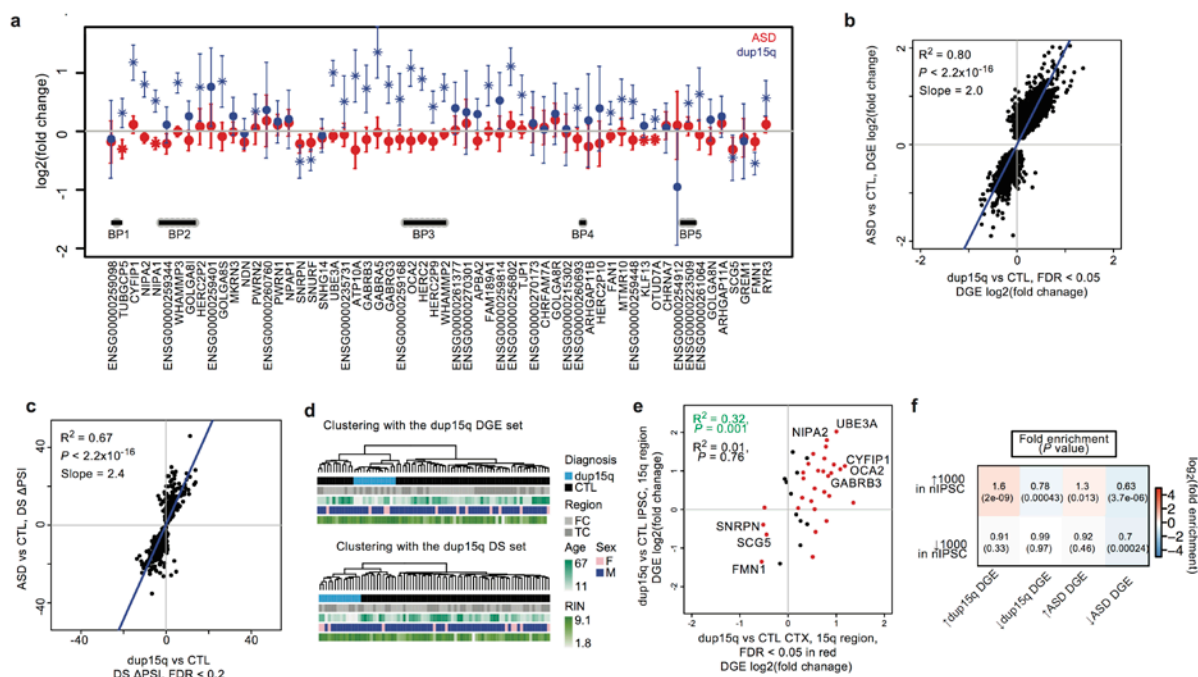
1   Figure 4 | Duplication 15q Syndrome recapitulates transcriptomic changes in idiopathic ASD. a, DGE
2   changes across the 15q11-13.2 region for ASD and dup15q compared to CTL, error bars are +/- 95%
3   confidence intervals for the fold changes. b, Comparison of effect sizes in dup15q vs CTL and ASD vs
4   CTL, with changes in dup15q at FDR ≤ 0.05 highlighted. c, Comparison of differential splicing (DS)
5   changes in dup15q vs CTL and ASD vs CTL, highlighting 402 events at FDR ≤ 0.2 in dup15q. d, Average
6   linkage hierarchical clustering of dup15q samples and controls using the DGE and DS gene sets. e, Plot of
7   fold changes between induced pluripotent stem cells differentiated into neurons (nIPSCs) from dup15q vs
8   CTL and postmortem CTX DGE from dup15q vs CTL in the 15q region. f, Heatmap overlapping the top
9   1000 genes up- and down- regulated in the nIPSC comparison to the up- and down- regulated genes in
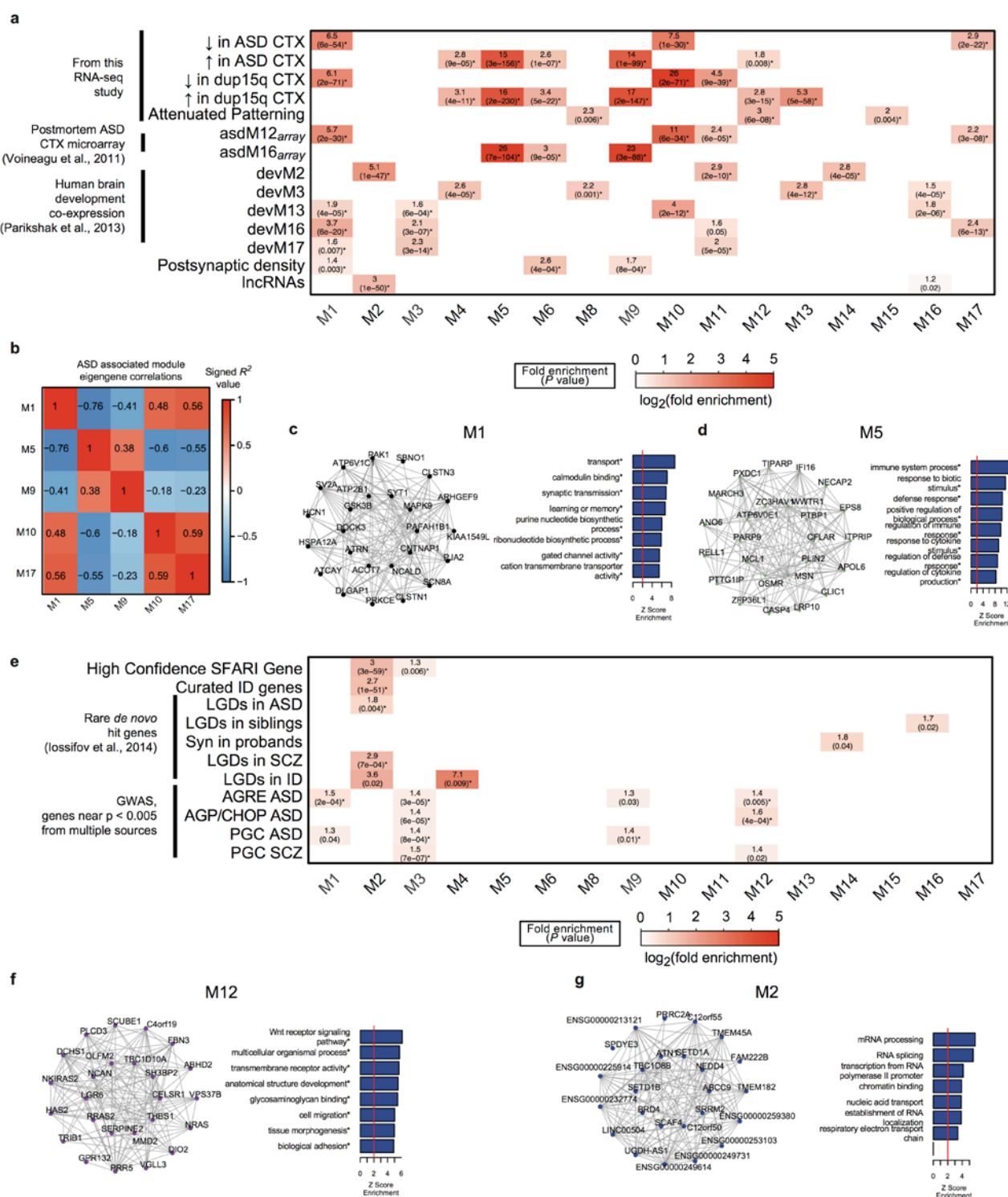10   dup15q and idiopathic ASD CTX.

1

2

Figure 5 | Co-expression network analysis across all ASD and CTL samples in CTX. a, Gene set enrichment analyses comparing the 16 co-expression modules with multiple gene sets from this RNA-seq study, from postmortem ASD CTX microarray, from human brain development, from the postsynaptic density and set of all brain-expressed lncRNAs. b, Comparison of five ASD-associated modules against each other by correlating module eigengenes. c, Module plot of M1 displaying the top 25 hub genes along with the module's Gene Ontology term enrichment. d, similar to c, but for M5. e, Gene set enrichment analysis with genome-wide whole-exome sequencing data (Rare *de novo* hit genes) and genome-wide association study (GWAS) results in ASD, schizophrenia (SCZ), and intellectual disability (ID). Boxes are filled if the odds

14

1 ratio is greater than 0, and the enrichment $P < 0.05$. Asterisks* indicate FDR ≤ 0.05 across all comparisons
2 in a and e. f,g, similar to c, but for M12 and M2, respectively. Abbreviations: LGD, likely gene disrupting,
3 genes affected by nonsense, nonsynonymous, or splice-site mutations or frame-shift indels; AGRE,
4 AGP/CHOP, and PGC refer to consortia that collect genetic data (Supplementary Information for details).
5

6 Methods
7

8 Sample description: Brain tissue for ASD and control individuals was acquired from the Autism Tissue
9 Program (ATP) brain bank at the Harvard Brain and Tissue Bank and the University of Maryland Brain and
10 Tissue Bank (a Brain and Tissue Repository of the NIH NeuroBioBank). Sample acquisition protocols were
11 followed for each brain bank, and samples were de-identified prior to acquisition. Brain sample and
12 individual level metadata is available in Supplementary Table 1.
13

14 RNA-seq methodology: Starting with 1ug of total RNA, samples were rRNA depleted (RiboZero Gold,
15 Illumina) and libraries were prepared using the TruSeq v2 kit (Illumina) to construct unstranded libraries
16 with a mean fragment size of 150bp (range 100-300bp) that underwent 50bp paired end sequencing on an
17 Illumina HiSeq 2000 or 2500 machine. Paired-end reads were mapped to hg19 using Gencode v18
18 annotations[55] via Tophat2[56]. Gene expression levels were quantified using union exon models with
19 HTSeq[57]. For additional and information on sequencing and read alignment parameters, please see
20 Supplementary Information.
21

22 Sample sets for analysis: For differential gene expression and splicing analysis, we defined an age matched
23 set, referred to as the ASD Discovery Set (106 samples in CTX, 51 in cerebellum) of idiopathic ASD and
24 control samples for the discovery set, and held out younger or unmatched samples as the ASD Discovery
25 Set (17 in CTX, 8 in cerebellum). Dup15q individuals were analysed separately, utilizing the full set of
26 controls from the ASD Discovery Set. For co-expression network analysis, we combined the discovery set,
27 replication set, and dup15q individuals for a total of 137 CTX samples and 59 cerebellum samples.
28

29 Differential Gene Expression (DGE): DGE analysis was performed with expression levels adjusted for gene
30 length, library size, and G+C content (referred to as "Normalized FPKM") Supplementary Information.
31 CTX samples (frontal and temporal) were analyzed separately from cerebellum samples. A linear mixed
32 effects model framework was used to assess differential expression in log2(Normalized FPKM) values for
33 each gene for cortical regions (as multiple brain regions were available from the same individuals) and a
34 linear model was used for cerebellum (where one brain region was available in each individual, with a
35 handful of technical replicates removed). Individual brain ID was treated as a random effect, while age, sex,
36 brain region (except in the case of cerebellum, where there is only one region), and diagnoses were treated
37 as fixed effects. We also used technical covariates accounting for RNA quality, library preparation, and
38 batch effects as fixed effects into this model (Supplementary Information).
39

40 Reproducibility analyses: We assessed replication between datasets by evaluating the concordance between
41 independent sample sets by comparing the squared correlation (R^2) of fold changes of genes in each sample
42 set at a non-stringent P value threshold. This general approach has been shown to be effective for
43 identifying reproducible gene expression patterns[58], and we modify it such that the P value threshold is set
44 in one sample set (the x axis in the scatterplots), and the R^2 with fold changes in these genes are evaluated in
45 an independent sample set (the y axis in the scatterplots).
46

47 Differential Splicing Analysis: Alternative splicing was quantified using the percent spliced in (PSI) metric
48 using Multivariate Analysis of Transcript Splicing (MATS, v3.08)[18]. For each event, MATS reports counts
49 supporting the inclusion (I) or exclusion (E) of a splicing event. To reduce spurious events due to low

1   counts, we required at least 80% of samples to have I + S >= 10. For these events, the percent spliced in is
2   calculated as PSI = I / (I + S) (Extended Data Fig. 4a). Statistical analysis for differential splicing was
3   performed utilizing the linear mixed effects model regression framework as described above for DGE. This
4   approach is advantageous over existing methods as it allows modeling of covariates and takes into
5   consideration the variability in PSI across samples when assessing event significance with ASD
6   (Supplementary Information).
7

8   Genotyping dup15q: For Dup15q samples, the type of duplication and copy number in the breakpoint 2-3
9   region were available for these brains[59]. To expand this to the regions between each of the recurrent
10 breakpoint in these samples, 7/8 dup15q brains were genotyped (one was not genotyped due to limitations in
11 tissue availability). The number of copies between each of the breakpoints is reported in Extended Data Fig.
12 7a.
13

14 Co-expression network analysis: The R package weighted gene co-expression network analysis (WGCNA)
15 was used to construct co-expression networks using the technical variation normalized data[31,60] (referred to
16 as "Adjusted FPKM"). We used the biweight midcorrelation to assess correlations between
17 log2(Normalized FPKM) and parameters for network analysis are described in Supplementary Information.
18 Notably, we utilized a modified version of WGCNA that involves bootstrapping the underlying dataset 100
19 times and constructing 100 networks. The consensus of these networks (50th percentile across all edges) was
20 then used as the final network [32], ensuring that a handful of samples do not determine the network structure.
21 For module-trait analyses, 1st principal component of each module (eigengene) was related to ASD
22 diagnosis, age, sex, and brain region in a linear mixed effects framework as above, only replacing the
23 expression values of each gene with the eigengene.
24

25 Enrichment analysis of gene sets and GWAS: Enrichment analyses were performed either with Fisher's
26 exact test (cell type and splicing factor enrichments) or logistic regression (all enrichment analyses in Figure
27 5). We used logistic regression in the latter case to control for gene length or other biases that may influence
28 enrichment analysis (Supplementary Information). All GO term enrichment analysis was performed using
29 GO Elite[61] with 10,000 permutations. We focused on molecular function and biological process terms for
30 display purposes.
31

32 Extended Data Figures & Legends
33
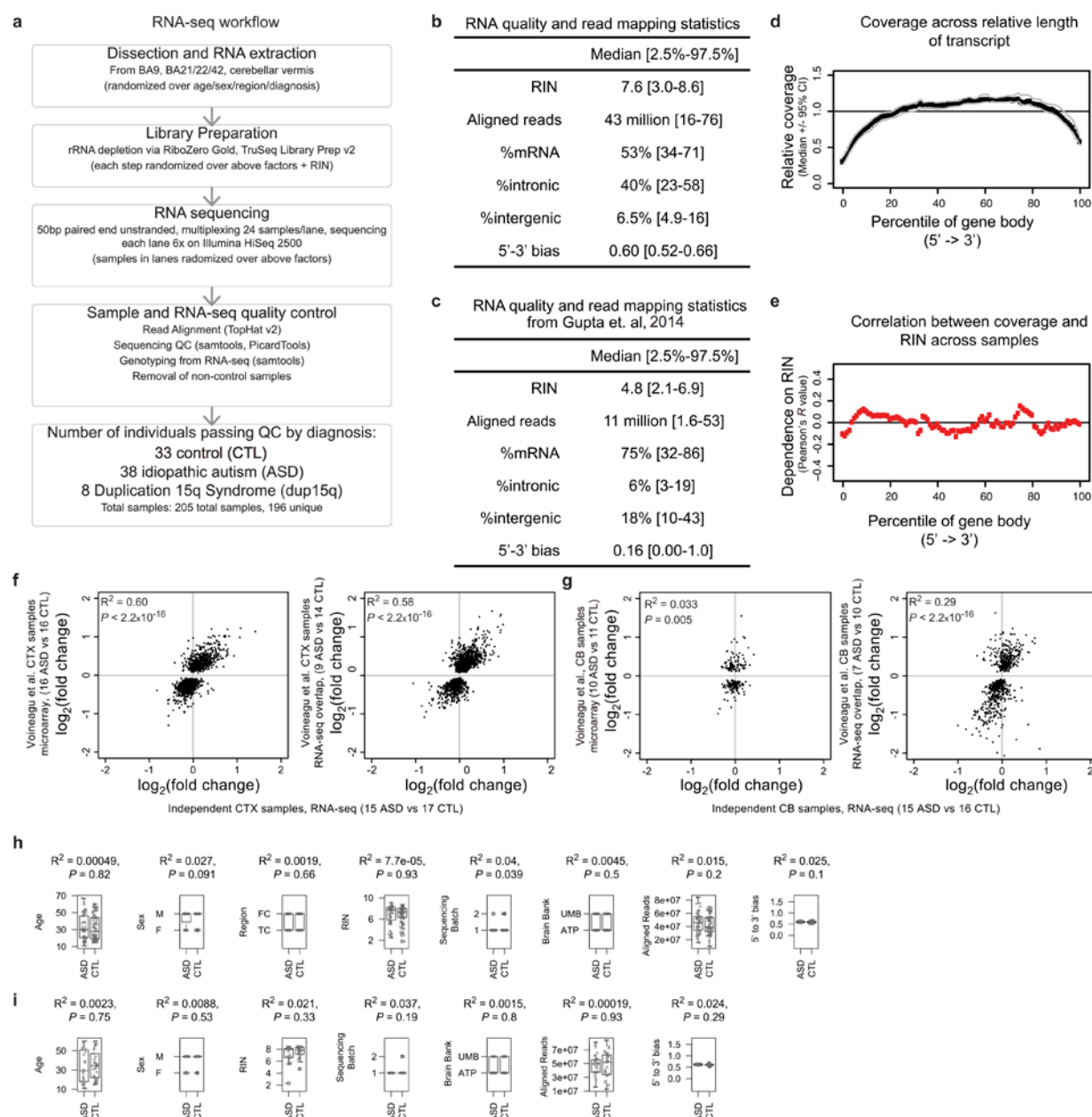
1
2 Extended Data Figure 1 | Methodology, quality control, and differential expression replication analysis. a,
3 RNA-seq workflow, including RNA extraction, library preparation, sequencing, read alignment, and quality
4 control. b, RNA-seq quality and alignment statistics from this study, including RNA integrity number
5 (RIN), number of aligned reads, proportion of reads mapping to different genomic features (mRNA,
6 intronic, intergenic), and bias in coverage from the 5' to the 3' end of the top 1000 expressed transcripts
7 (statistics compiled using PicardTools). c, Similar statistics as in b for another RNA-seq study that utilized
8 polyA tail selection mRNA-seq to evaluate the transcriptome in ASD cortex[11] (primarily BA19, visual
9 cortex, but also including some BA10/44 samples, frontal cortex). d, RNA-seq read coverage relative to
10 normalized gene length across transcripts from the 5' to the 3' end in this study. e, Dependence between
11 coverage and RIN across gene body (correlation between RIN and coverage in d across samples). f,
12 Correlation of ASD vs CTL fold changes between previously evaluated and new ASD samples in CTX by
13 microarray (left) and RNA-seq (right) using genes that were at $P < 0.05$ the samples from Voineagu et al.,

1   2011. g, Correlation between effect sizes as in f, but for cerebellum (CB) samples. h,i, Correlation between
2   covariates and ASD vs CTL status in CTX (h) and CB (i) in the ASD Discovery Set.
3
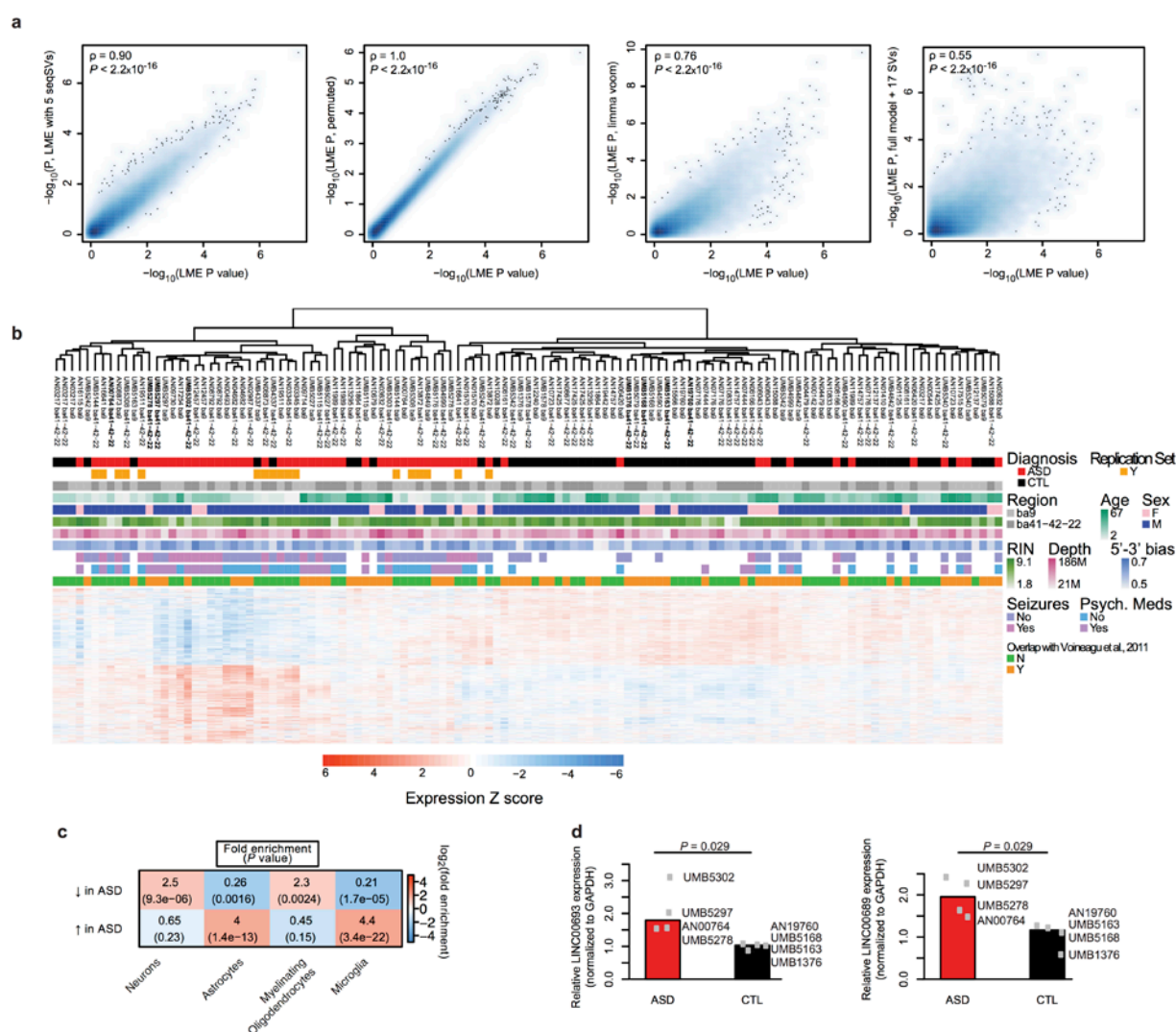
4
5   Extended Data Figure 2 | Transcriptome-wide differential gene expression (DGE) analysis in CTX. a,
6   Comparison of P value rankings across different methods for DGE with Spearman's correlation. From left
7   to right: removal of three additional principal components of sequencing statistics (Supplementary
8   Information) related to RNA-sequencing quality, application of a permutation analysis for DGE P value
9   computation, application of variance-weighted linear regression for DGE[62], and using surrogate variable
10 analysis for DGE[63]. b, Average linkage hierarchical clustering heatmap using all genes DGE in the ASD
11 Discovery Set, but including all idiopathic ASD frontal cortex (FC) and temporal cortex (TC) samples
12 across 123 samples, combining the ASD Discovery set and the ASD Replication set. Bolded samples in the
13 dendrogram are used for validation in d. c, Enrichment analysis of cell-type specific gene sets (5-fold
14 enriched in the cell type compared to all other cells) with genes decreased and increased in ASD. d, RT-
15 PCR validation of the two lincRNAs shown in Figure 1f-g, P values are computed with the Wilcoxon rank-
16 sum test.
17

18

1
2 Extended Data Figure 3 | Gene browser tracks for selected primate-specific lncRNAs. For each lncRNA,
3 expression for representative samples for ASD vs CTL (top) in human, macaque (middle), and mouse
4 (bottom) are shown. The genome location for macaque and mouse displayed is syntenic to the human
5 region, with the expected location of the lncRNA highlighted.
6
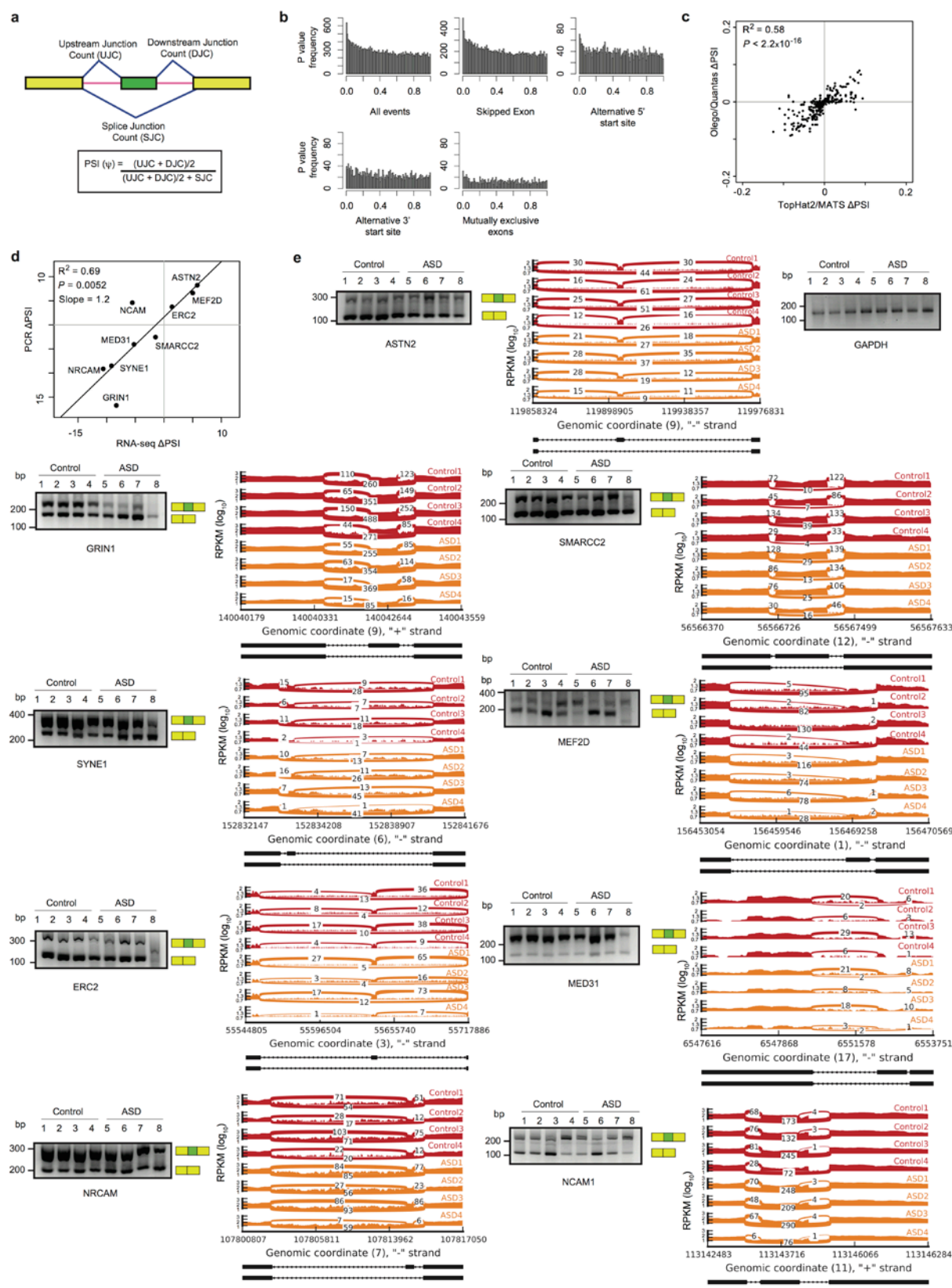
1
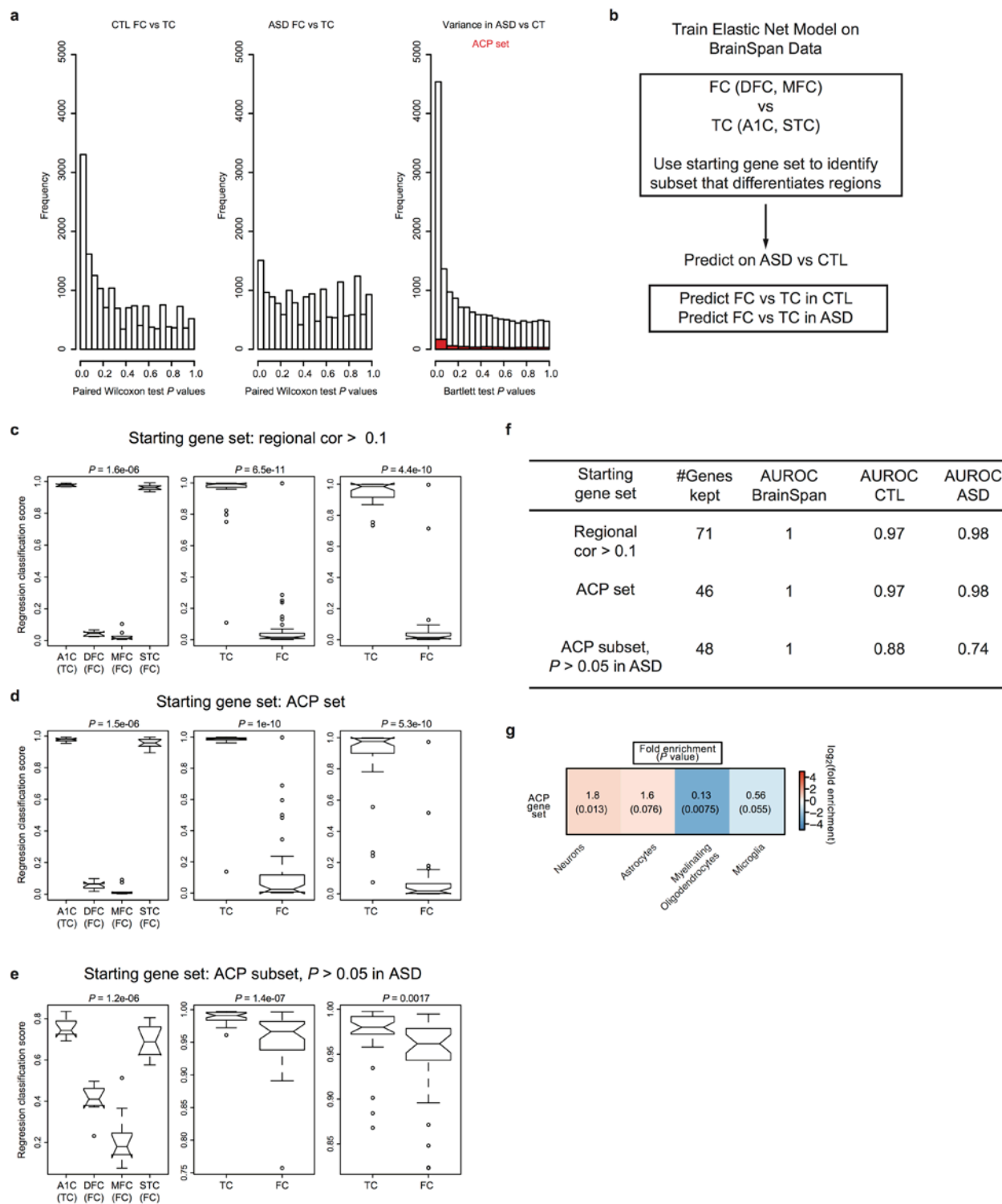2 Extended Data Figure 4 | Splicing analyses and validation in ASD. a, Schematic describing how the percent
3 spliced in (PSI) metric is computed. b, Distribution of P values for changes in the PSI between ASD and
4 CTL in CTX for all events (left) and event subtypes (SE, spiced exon; A5SS, alternative 5' splice site;
5 A3SS, alternative 3' splice site; MXE, mutually exclusive exons). c, Comparison of the CTX splicing

20

1 analyses in when using PSI values obtained via read alignment by TopHat2[64] followed by the MATS[18]
2 pipeline (used throughout this study) against read alignment by OLego followed by Quantas[65]. d,
3 Comparison of ΔPSI values in nine splicing events between PCR and RNA-seq. e, PCR validation and
4 sashimi plots for the nine splicing events delineated in d, from the samples highlighted in Extended Data
5 Fig. 5a.
6
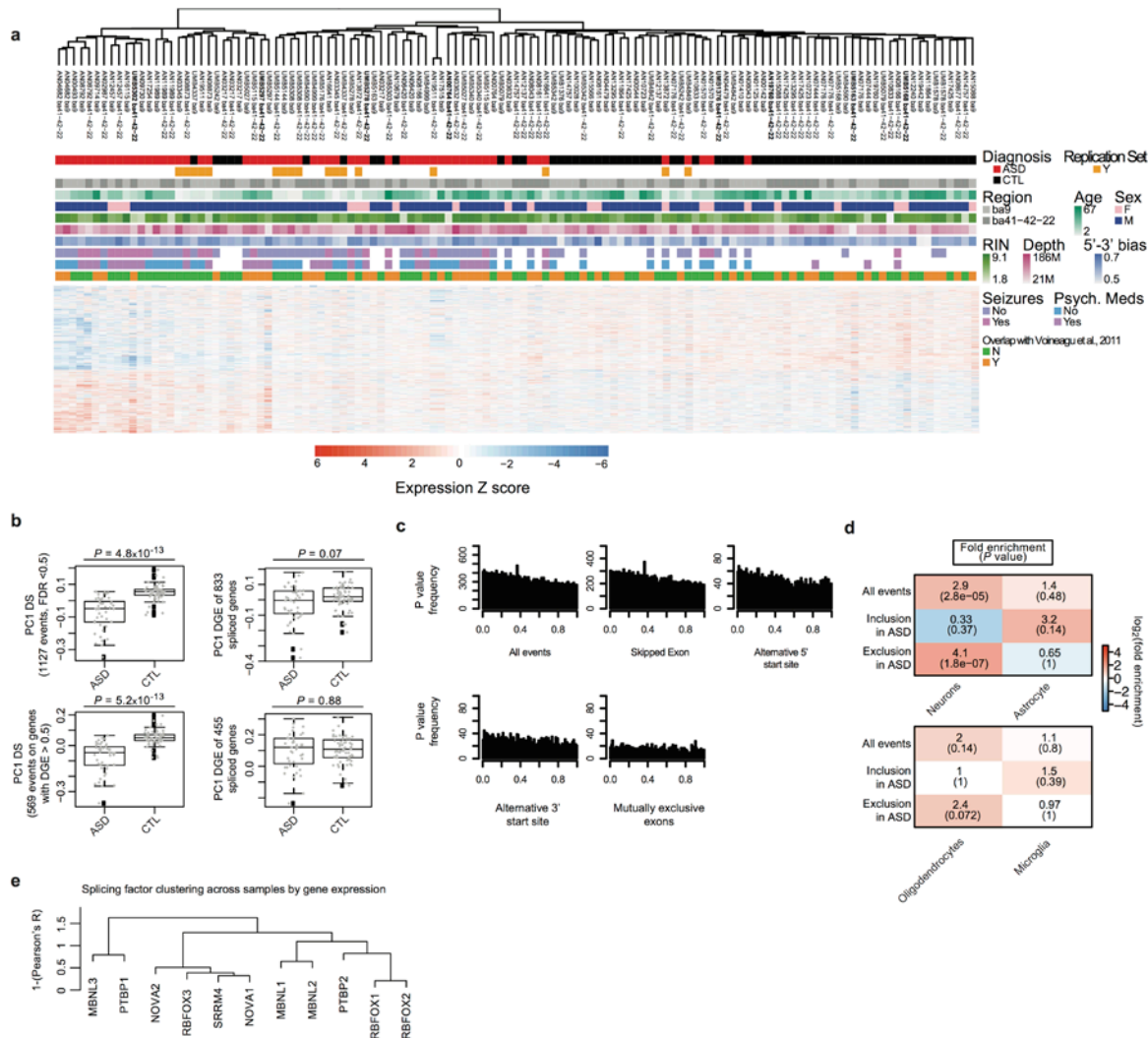
7
8 Extended Data Figure 5 | Additional splicing analyses in ASD. a, Average linkage hierarchical clustering
9 heatmap using all differentially spiced (DS) events from the ASD Discovery Set, but including all idiopathic
10 ASD neocortical samples (FC and TC) across 123 samples, combining the ASD Discovery set and the ASD
11 Replication set. Bolded samples in the dendrogram were used for PCR validation in Extended Data Fig. 4.
12 b, Top: difference between ASD and CTL in the DS set based on PC1 of the DS set at the PSI level, and
13 PC1 of the gene expression levels of genes in the DS set. Bottom: Same comparison after differentially
14 expressed genes (p < 0.05) are removed. c, Distribution of P values for changes in the PSI between ASD and
15 CTL in cerebellum. d, Cell-type enrichment analysis of splicing events from CTX. e, Average-linkage
16 hierarchical clustering using 1-(Pearson's correlation) to compare the gene expression patterns of the
17 splicing factors investigated in Figure 2.

21

| Starting gene set	#Genes kept	AUROC BrainSpan	AUROC CTL	AUROC ASD
Regional cor > 0.1	71	1	0.97	0.98
ACP set	46	1	0.97	0.98
ACP subset, $P > 0.05$ in ASD	48	1	0.88	0.74

1

Extended Data Figure 6 | Attenuation of cortical patterning in ASD. a, Histograms of P values from paired
Wilcoxon rank-sum test differential gene expression between 16 frontal cortex (FC) and 16 temporal cortex
(TC) in CTL and ASD and a histogram of Bartlett's test P values for differences in gene expression variance
between ASD and CTL for all genes (white) and genes in the Attenuated Cortical Patterning (ACP) set
(red). c, Approach to training the elastic net model on BrainSpan and application of the model on 123
cortical samples in this study. c-e, Results of learned cortical region classifications with different starting
gene sets, with the BrainSpan training set (left), CTL samples (middle), and ASD samples (right) in each

1   panel and the Wilcoxon rank-sum test P value of FC vs TC difference for each comparison. f, Summary of

2   results form c-e. g, Cell type enrichment analysis for genes in the ACP set. Abbreviations: A1C, primary

3   auditory cortex; DFC, dorsolateral prefrontal cortex; MFC, medial prefrontal cortex; STC, superior temporal

4   cortex; FC, frontal cortex; TC, temporal cortex; AUROC, area under the receiver-operator characteristic
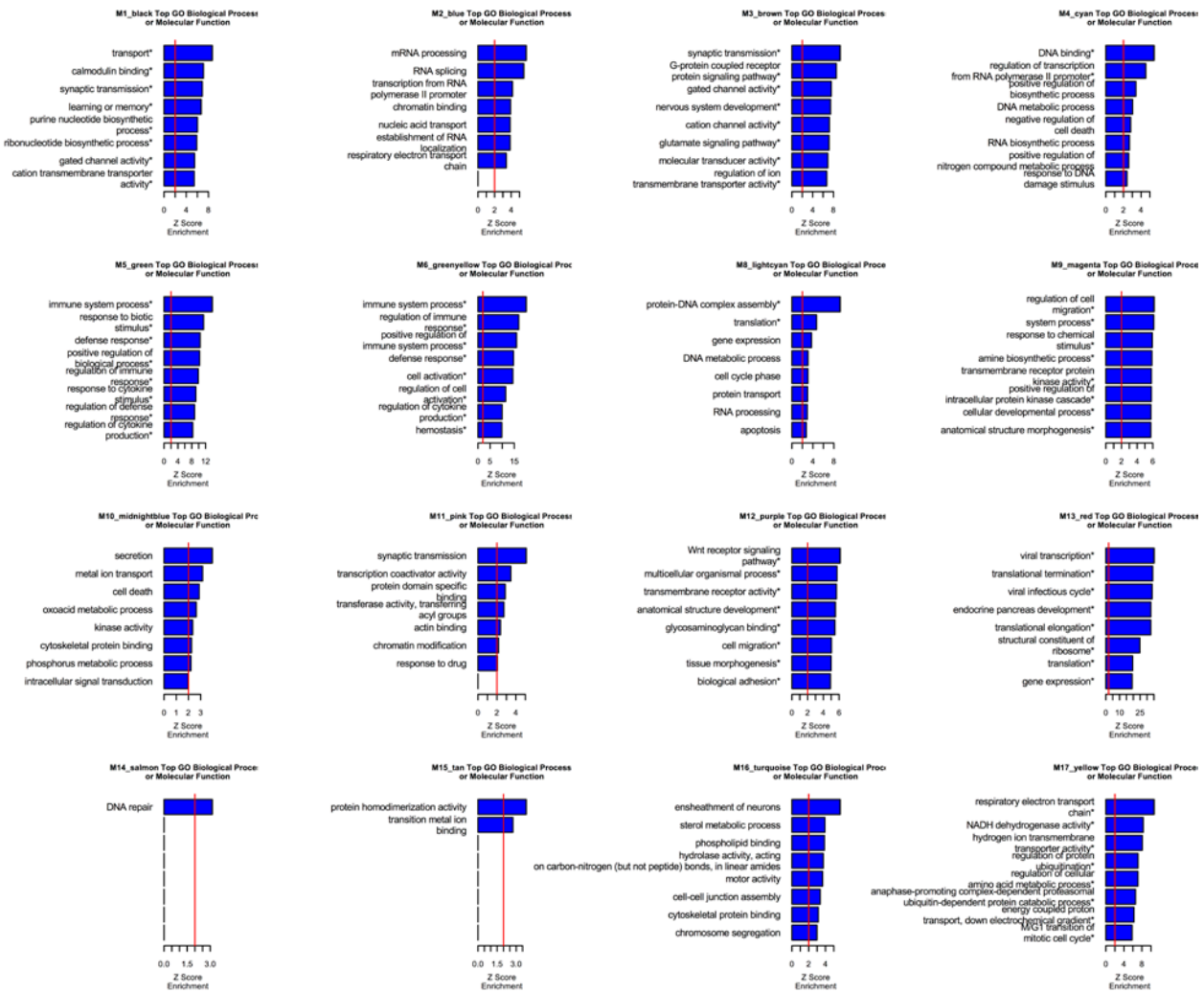
5   curve.

1
2 Extended Data Figure 7 | Dup15q syndrome analyses. a, Copy number between breakpoints (BP) in the 15q
3 region. Genome-wide CNV analysis allowed evaluation of copy number in additional regions from previous

1 studies[59,66]. b, Differential expression across the 15q region of interest in dup15q vs CTL and ASD vs CTL
2 cerebellum, note only 3 samples were available for dup15q cerebellum so additional analyses were not
3 pursued. c, Gene Ontology term enrichment analysis for the dup15q CTX differential expression set. d,
4 Gene Ontology term enrichment analysis for the dup15q CTX differential splicing (DS) set. e, Hierarchical
5 clustering of iPSC-derived neurons from dup15q, Angelman syndrome, and a control[29].
6
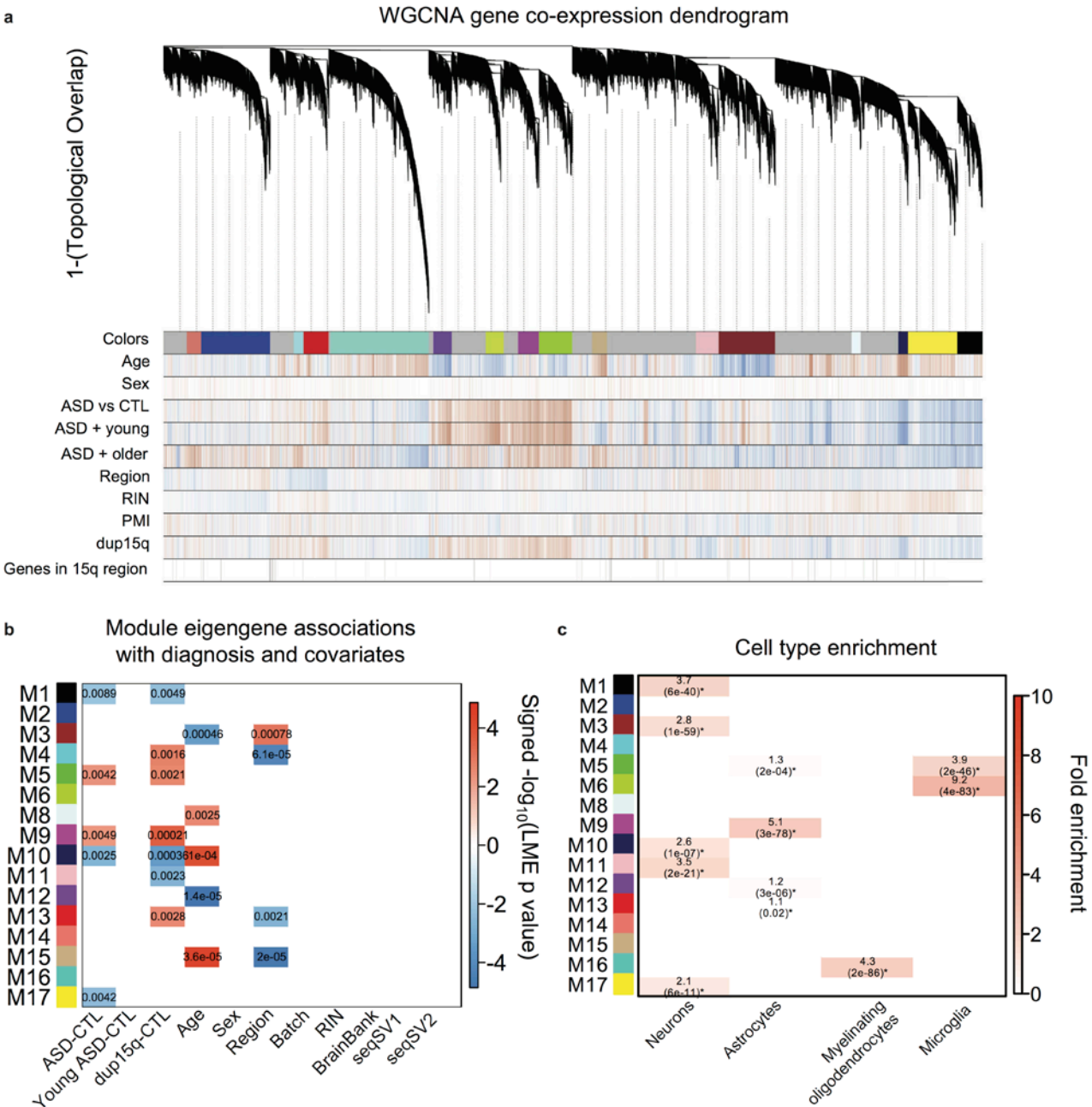
7
8 Extended Data Figure 8 | Co-expression network analysis in ASD CTX. a, Modules identified from a
9 dendrogram constructed from a consensus of 100 bootstrapped datasets using the 137 CTX samples.
10 Correlations for each gene to each measured factor are delineated below the dendrogram (blue = negative,
11 red = positive correlation). b, Module-trait associations as computed by a linear mixed effects model with all
12 factors on the x-axis used as covariates. All P values are displayed where the coefficient passed $p < 0.01$.
13 Note that this alternative approach to module-trait association agrees with the Fisher's exact test used in
14 Figure 5a when the fold enrichment for module overlap with DGE sets is > 2.8, and we use an intersection

1 of both methods for the modules focused on in Figure 5b. c, Module enrichments for cell type specific gene
2 expression patterns.
3

4
5 Extended Data Figure 9 | GO term enrichments for all modules. *FDR < 0.05 across all GO enrichments
6 across all modules.
7

1 References

2

3 1. Geschwind, D. H. Genetics of autism spectrum disorders. *Trends Cogn. Sci. (Regul. Ed.)* 15, 409–416
4   (2011).
5 2. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* 46, 881–885
6   (2014).
7 3. Geschwind, D. H. & State, M. W. Gene hunting in autism spectrum disorder: on the path to precision
8   medicine. *Lancet Neurol* (2015). doi:10.1016/S1474-4422(15)00044-7
9 4. Gratten, J., Wray, N. R., Keller, M. C. & Visscher, P. M. Large-scale genomics unveils the genetic
10   architecture of psychiatric disorders. *Nat. Neurosci.* 17, 782–790 (2014).
11 5. Chen, J. A., Peñagarikano, O., Belgard, T. G., Swarup, V. & Geschwind, D. H. The emerging picture
12   of autism spectrum disorder: genetics and pathology. *Annu Rev Pathol* 10, 111–144 (2015).
13 6. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new
14   neurobiology. *Nat Rev Genet* 9, 341–355 (2008).
15 7. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology.
16   *Nature* 474, 380–384 (2011).
17 8. Purcell, A. E., Jeon, O. H., Zimmerman, A. W., Blue, M. E. & Pevsner, J. Postmortem brain
18   abnormalities of the glutamate neurotransmitter system in autism. *Neurology* 57, 1618–1628 (2001).
19 9. Garbett, K. *et al.* Immune transcriptome alterations in the temporal cortex of subjects with autism.
20   *Neurobiology of Disease* 30, 303–311 (2008).
21 10. Chow, M. L. *et al.* Age-Dependent Brain Gene Expression and Copy Number Anomalies in Autism
22   Suggest Distinct Pathological Processes at Young Versus Mature Ages. *PLoS Genet.* 8, e1002592
23   (2012).
24 11. Gupta, S. *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and
25   neuronal activity-dependent genes in autism. *Nat Comms* 5, 5748 (2014).
26 12. Jaffe, A. E. *et al.* Developmental regulation of human cortex transcription and its clinical relevance at
27   single base resolution. *Nature Publishing Group* 18, 154–161 (2015).
28 13. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–
29   330 (2015).
30 14. Iyer, M. K. *et al.* The landscape of long noncoding RNAs in the human transcriptome. *Nat Genet* 47,
31   199–208 (2015).
32 15. Necsulea, A. *et al.* The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*
33   505, 635–640 (2014).
34 16. Weyn-Vanhentenryck, S. M. *et al.* HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-
35   Regulatory Network Linked to Brain Development and Autism. *Cell Reports* 6, 1139–1152 (2014).
36 17. Irimia, M. *et al.* A highly conserved program of neuronal microexons is misregulated in autistic
37   brains. *Cell* 159, 1511–1523 (2014).
38 18. Shen, S. *et al.* MATS: a Bayesian framework for flexible detection of differential alternative splicing
39   from RNA-Seq data. *Nucleic Acids Res* 40, e61–e61 (2012).
40 19. Lionel, A. C. *et al.* Disruption of the ASTN2/TRIM32 locus at 9q33.1 is a risk factor in males for
41   autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Human Molecular*
42   *Genetics* 23, 2752–2768 (2014).
43 20. Lovci, M. T. *et al.* Rbfox proteins regulate alternative mRNA splicing through evolutionarily
44   conserved RNA bridges. *Nat Struct Mol Biol* 20, 1434–1442 (2013).
45 21. Raj, B. *et al.* A Global Regulatory Mechanism for Activating an Exon Network Required for
46   Neurogenesis. *Molecular Cell* 56, 90–103 (2014).
47 22. Gueroussov, S. *et al.* An alternative splicing event amplifies evolutionary differences between
48   vertebrates. *Science* 349, 868–873 (2015).
49 23. Dittmar, K. A. *et al.* Genome-wide determination of a broad ESRP-regulated posttranscriptional
50   network by high-throughput sequencing. *Molecular and Cellular Biology* 32, 1468–1482 (2012).

24. Tibshirani, R., Johnstone, I., Hastie, T. & Efron, B. Least angle regression. *The Annals of Statistics* 32, 407–499 (2004).

25. Sunkin, S. M. *et al.* Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res* 41, D996–D1008 (2013).

26. Lui, J. H. *et al.* Radial glia require PDGFD–PDGFRβ signalling in human but not mouse neocortex. *Nature* 515, 264–268 (2014).

27. Lai, T. *et al.* SOX5 Controls the Sequential Generation of Distinct Corticofugal Neuron Subtypes. *Neuron* 57, 232–247 (2008).

28. Kwan, K. Y. *et al.* SOX5 postmitotically regulates migration, postmigratory differentiation, and projections of subplate and deep-layer neocortical neurons. *Proc. Natl. Acad. Sci. U.S.A.* 105, 16021–16026 (2008).

29. Germain, N. D. *et al.* Gene expression analysis of human induced pluripotent stem cell-derived neurons carrying copy number variants of chromosome 15q11-q13.1. *Mol Autism* 5, 44 (2014).

30. Stein, J. L. *et al.* A Quantitative Framework to Evaluate Modeling of Cortical Development by Neural Stem Cells. *Neuron* 83, 69–86 (2014).

31. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology* 4, Article17 (2005).

32. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* 1, 54 (2007).

33. Morales, P. L. M., Quiroga, A. C., Barbas, J. A. & Morales, A. V. SOX5 controls cell cycle progression in neural progenitors by interfering with the WNT–β-catenin pathway. *EMBO reports* 11, 466–472 (2010).

34. Parikshak, N. N. *et al.* Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. *Cell* 155, 1008–1021 (2013).

35. Bayés, À. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* 14, 19–21 (2010).

36. Schafer, D. P. *et al.* Microglia Sculpt Postnatal Neural Circuits in an Activity and Complement-Dependent Manner. *Neuron* 74, 691–705 (2012).

37. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221 (2014).

38. Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorders. *Human Molecular Genetics* 21, 4781–4792 (2012).

39. Wang, K. *et al.* Common genetic variants on 5p14.1 associate with autism spectrum disorders. 459, 528–533 (2009).

40. Cross-Disorder Group of the Psychiatric Genomics Consortium *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 45, 984–994 (2013).

41. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–190 (2014).

42. Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E. E. The discovery of integrated gene networks for autism and related disorders. *Genome Res* 25, 142–154 (2015).

43. Gilman, S. R. *et al.* Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* 15, 1723–1728 (2012).

44. Willsey, A. J. *et al.* Coexpression Networks Implicate Human Midfetal Deep Cortical Projection Neurons in the Pathogenesis of Autism. *Cell* 155, 997–1007 (2013).

45. Ziats, M. N. & Rennert, O. M. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. *J Mol Neurosci* 49, 589–593 (2012).

46. Geschwind, D. H. & Rakic, P. Cortical Evolution: Judge the Brain by Its Cover. *Neuron* 80, 633–647 (2013).

47. Zhang, Y. E., Landback, P., Vibranovski, M. D. & Long, M. Accelerated Recruitment of New Brain Development Genes into the Human Genome. *PLoS Biol* 9, e1001179 (2011).

48. Fogel, B. L. *et al.* RBFOX1 regulates both splicing and transcriptional networks in human neuronal

28

1   development. *Human Molecular Genetics* 21, 4171–4186 (2012).

2   49.   Gehman, L. T. *et al.* The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the
3   mammalian brain. *Nat Genet* 43, 706–711 (2011).

4   50.   Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic
5   of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci* 14, 755–
6   769 (2013).

7   51.   Srinivasan, K. *et al.* A network of genetic repression and derepression specifies projection fates in the
8   developing neocortex. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19071–19078 (2012).

9   52.   Morgan, J. T. *et al.* Abnormal microglial–neuronal spatial organization in the dorsolateral prefrontal
10   cortex in autism. *Brain Research* 1456, 72–81 (2012).

11   53.   Stessman, H. A., Bernier, R. & Eichler, E. E. A Genotype-First Approach to Defining the Subtypes of
12   a Complex Disease. *Cell* 156, 872–877 (2014).

13   54.   Ronemus, M., Iossifov, I., Levy, D. & Wigler, M. The role of de novo mutations in the genetics of
14   autism spectrum disorders. *Nat Rev Genet* 15, 133–141 (2014).

15   55.   Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol* 7, S4–9
16   (2006).

17   56.   Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat*
18   *Biotechnol* 31, 46–53 (2012).

19   57.   Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput
20   sequencing data. *Bioinformatics* 31, 166–169 (2015).

21   58.   Shi, L. *et al.* The balance of reproducibility, sensitivity, and specificity of lists of differentially
22   expressed genes in microarray studies. *BMC Bioinformatics* 9, S10 (2008).

23   59.   Scoles, H. A., Urraca, N., Chadwick, S. W., Reiter, L. T. & LaSalle, J. M. Increased copy number for
24   methylated maternal 15q duplications leads to changes in gene and protein expression in human
25   cortical samples. *Mol Autism* 2, 19 (2011).

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

28   61.   Zambon, A. C. *et al.* GO-Elite: a flexible solution for pathway and ontology over-representation.
29   *Bioinformatics* 28, 2209–2210 (2012).

30   62.   Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis
31   tools for RNA-seq read counts. *Genome Biol* 15, R29 (2014).

32   63.   Leek, J. T. & Storey, J. D. Capturing Heterogeneity in Gene Expression Studies by Surrogate
33   Variable Analysis. *PLoS Genet.* 3, e161 (2007).

34   64.   Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with
35   TopHat and Cufflinks. *Nat Protoc* 7, 562–578 (2012).

36   65.   Wu, J., Anczuków, O., Krainer, A. R., Zhang, M. Q. & Zhang, C. OLego: fast and sensitive mapping
37   of spliced mRNA-Seq reads using small seeds. *Nucleic Acids Res* 41, 5149–5163 (2013).

38   66.   Wintle, R. F. *et al.* A genotype resource for postmortem brain samples from the Autism Tissue
39   Program. *Autism Res* 4, 89–97 (2011).

40

41

42   End Notes

43   None.

44   Author Contributions

45   NNP, VS, and TGB performed dissections and RNA-seq analyses and differential gene expression
46   analysis. NNP and VS performed splicing and co-expression network analysis. NNP and TGB
47   performed analyses with Duplication 15q Syndrome individuals. NNP and MG reviewed clinical
48   information and performed meta-analysis of ASD gene expression studies. VL and JKL performed

1   genotyping and CNV analysis on dup15q samples. VS performed validation experiments for gene and
2   splicing level alterations in ASD. MI and BJB assisted with splicing analyses. RJ performed dissections.
3   SH provided guidance on differential gene expression and co-expression analyses. DHG provided
4   guidance on all experiments and analyses. NNP and DHG wrote the manuscript. All authors
5   contributed to revising and finalizing the manuscript.