

1 **Title:** Atypical genomic patterning of the cerebral cortex in autism with poor early language
2 outcome

3 **Authors:** Michael V. Lombardo^{1,2*}, Lisa Eyler^{3,4}, Tiziano Prampero⁵, Vahid H. Gazestani⁵,
4 Donald J. Hagler Jr.^{6,7}, Chi-Hua Chen³, Anders M. Dale^{6,7,8}, Jakob Seidlitz^{9,10}, Richard A. I.
5 Bethlehem^{2,11}, Natasha Bertelsen^{1,12}, Cynthia Carter Barnes⁵, Linda Lopez⁵, Kathleen
6 Campbell^{5,13}, Nathan E. Lewis^{14,15,16}, Karen Pierce⁵, & Eric Courchesne⁵

7 **Affiliations:**

8 ¹ Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and
9 Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy

10 ² Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge,
11 United Kingdom

12 ³ Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA

13 ⁴ VISN 22 Mental Illness Research, Education, and Clinical Center, VA San Diego Healthcare
14 System, San Diego, CA, USA

15 ⁵ Autism Center of Excellence, Department of Neurosciences, University of California, San Diego,
16 La Jolla, CA, USA

17 ⁶ Center for Multimodal Imaging and Genetics, University of California, San Diego, La Jolla, CA,
18 USA

19 ⁷ Department of Radiology, University of California, San Diego, La Jolla, CA, USA

20 ⁸ Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

21 ⁹ Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of
22 Philadelphia, Philadelphia PA USA

23 ¹⁰ Department of Psychiatry, University of Pennsylvania, Philadelphia, PA USA

24 ¹¹ Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

25 ¹² Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy

26 ¹³ Department of Pediatrics, University of Utah, Salt Lake City, UT, USA

27 ¹⁴ Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA

28 ¹⁵ Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

29 ¹⁶ Novo Nordisk Foundation Center for Biosustainability at the University of California, San
30 Diego, La Jolla, CA, USA

31 *Correspondence to: michael.lombardo@iit.it

32

33

34

35 **Abstract:** Cortical regional identities develop through anterior-posterior (A-P) and dorsal-ventral
36 (D-V) prenatal genomic patterning gradients. Here we find that A-P and D-V genomic patterning
37 of cortical surface area (SA) and thickness (CT) is intact in typically developing and autistic
38 toddlers with good language outcome, but is absent in autistic toddlers with poor early language
39 outcome. Genes driving this effect are prominent in midgestational A-P and D-V gene expression
40 gradients and prenatal cell types driving SA and CT variation (e.g., progenitor cells versus
41 excitatory neurons). These genes are also important for vocal learning, human-specific evolution,
42 and prenatal co-expression networks enriched for high-penetrance autism risk genes. Autism with
43 poor early language outcome may be linked to atypical genomic cortical patterning starting in
44 prenatal periods and which impacts later development of regional functional specialization and
45 circuit formation.

46

47 **One Sentence Summary:** Genomic patterning of the cortex is atypical in autistic toddlers with
48 poor early language outcome.

49

50

51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96

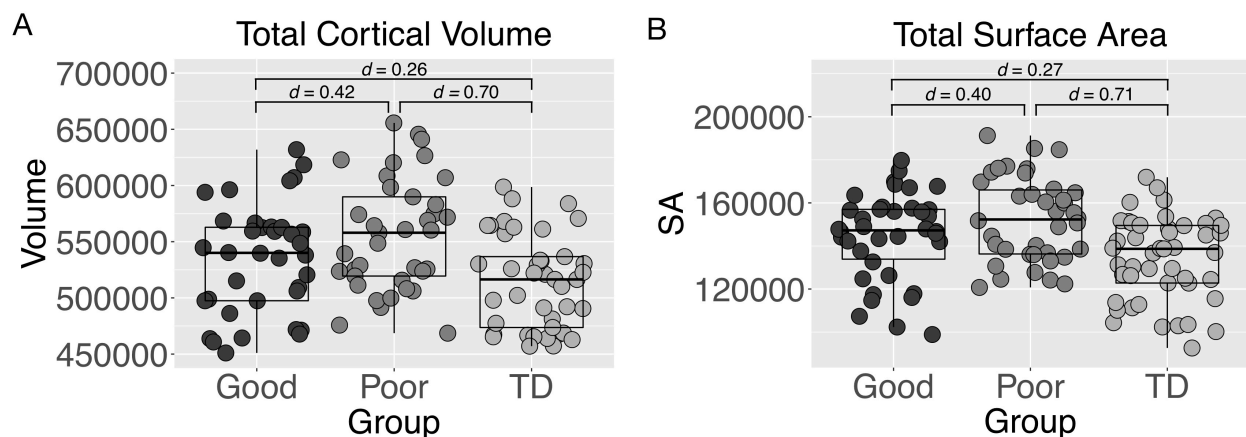
It is widely accepted that the autisms (ASD) arise in large part due to complex genetic mechanisms (1, 2). The major priorities for the field are to develop an individualized understanding of how complex genetic mechanisms cascade over development to cause phenotypic differentiation at multiple scales and link these mechanisms to clinical outcomes with high real-world impact and relevance (3–5). At the nexus of these priorities, our prior functional imaging (fMRI) work showed that large-scale activity in blood leukocyte gene co-expression modules differentially relates to language-relevant functional neural phenotypes measured in typically developing (TD) and ASD toddlers with good (ASD Good) versus poor (ASD Poor) early language outcome (6). This result indicates that the atypical language-relevant functional neural phenotypes typically seen in ASD Poor (7) are driven by different underlying functional genomic mechanisms.

The genes of importance that differentially relate to language-relevant functional neural phenotypes in ASD Poor (6) are an omnigenic (8) array of genes that are typically broadly expressed across many tissues including the brain. Broadly expressed genes tend to be one of the most important classes of ASD-risk genes and in the brain they show peak expression during early prenatal periods, when cell proliferation, differentiation, neurogenesis, and migration are the primary biological processes (9, 10). If these genes operate at early prenatal periods to affect proliferation, differentiation, neurogenesis, and migration, this suggests that structural features of the developing cerebral cortex such as surface area (SA) and cortical thickness (CT) may be substantially altered in the ASD Poor subtype.

The action of broadly expressed genes during prenatal periods may also be important for how the cortex is genomically patterned. It is well established that during prenatal periods the cortex is patterned by gene expression gradients that follow anterior-posterior (A-P) and dorsal-ventral (D-V) axes (11–17). This prenatal genomic patterning is the beginning of cortical arealization processes that allow different cortical regions to develop their own cellular, functional, and circuit identities (11, 13, 15, 16). Cortical arealization or patterning may be atypical in ASD. Prior evidence from case-control comparisons of post-mortem cortical tissue has found dysregulation of cortical patterning genes and attenuation of gene expression differences in frontal versus temporal cortex (18–20). WNT-signaling is known to affect cortical patterning (13, 15, 16, 21) and WNT-signaling abnormalities are also identified in ASD (19, 20, 22–24), particularly within broadly expressed ASD-risk genes (10). Therefore, if broadly expressed genes in early prenatal periods impact the ASD Poor subtype, could this also implicate a disturbance of prenatal genomic patterning of the cortex?

In the current work we examined these questions in a sample of n=123 toddlers (12-50 months) with and without ASD (ASD Good n=38, ASD Poor = n=38, and TD n = 47). With T1-weighted structural MRI images, we used Freesurfer (<http://surfer.nmr.mgh.harvard.edu>) to extract SA and CT measures from 12 cortical regions that parcellate the cortex by hierarchical genetic similarity (25–27). This cortical parcellation, known as GCLUST, was chosen in order to maximize sensitivity for detecting genetic relationships (28). GCLUST is also sensitive to the SA and CT genetic similarity gradients that fall along A-P and D-V axes and therefore, also maximizes sensitivity for detecting such genetically sensitive A-P and D-V gradient effects (25–27) (see Methods for more details).

97 Since one of the most robust findings on early structural brain development in ASD is the
98 on-average effect of early brain overgrowth in the first years of life (4, 29–31), we started by
99 examining whether there are subtype differences on global measures such as total cortical volume
100 (CV), SA and mean CT. Statistical models controlling for age and sex identified a group effect on
101 total CV ($F(2,193) = 14.30, p = 2.74e-6, \eta^2 = 0.075$) that is driven by the ASD Poor subtype having
102 on-average larger CV than the other groups (ASD Good vs ASD Poor $t(125) = 1.88, p = 0.06$,
103 *Cohen's d* = -0.42; TD vs ASD Poor $t(132) = -2.98, p = 0.003$, *Cohen's d* = -0.70; TD vs ASD
104 Good $t(127) = -1.10, p = 0.27$, *Cohen's d* = -0.26) (Fig. 1A). A group effect also emerged for total
105 SA ($F(2,193) = 15.39, p = 1.14e-6, \eta^2 = 0.072$) and was again driven by on-average increases in
106 ASD Poor relative to the other groups (ASD Good vs ASD Poor $t(125) = 1.79, p = 0.07$, *Cohen's d*
107 $d = -0.40$; TD vs ASD Poor $t(132) = -2.84, p = 0.005$, *Cohen's d* = -0.71; TD vs ASD Good $t(127)$
108 $= -1.28, p = 0.20$, *Cohen's d* = -0.27) (Fig. 1B). In contrast, no group differences were identified
109 for mean CT ($F(2,193) = 2.80, p = 0.06, \eta^2 = 0.002$; ASD Good vs ASD Poor $t(125) = -0.24, p =$
110 0.80 , *Cohen's d* = 0.03; TD vs ASD Poor $t(132) = 0.17, p = 0.86$, *Cohen's d* = 0.11; TD vs ASD
111 Good $t(127) = 0.37, p = 0.71$, *Cohen's d* = 0.07). These effects illustrate that the ASD Poor subtype
112 drives the on-average effect of early brain overgrowth in autism. Differences in CV and total SA,
113 but not mean CT, is compatible with other work showing that early brain overgrowth is largely
114 driven by expansion of cortical SA rather than CT (32, 33). We next examined regional level SA
115 or CT effects when adjusting for global differences using the GCLUST parcellation (see Methods).
116 Here we find no evidence of SA or CT group differences for any of the 12 GCLUST regions,
117 indicating that the primary overall group differences in brain size are restricted to global effects in
118 CV and SA, rather than localized regional effects after adjusting for such global effects.
119



120
121 **Fig. 1: Subtype differences in total cortical volume (A) and total surface area (B).** Standardized
122 effect sizes (*Cohen's d*) are shown for each pairwise group comparison.
123

124 We next examined large-scale associations between gene expression and regional SA or
125 CT from the GCLUST parcellation. To examine gene expression, leukocyte cells were extracted
126 from blood samples and microarrays were used to quantify expression from 14,426 protein coding
127 genes. This set 14,426 genes was then reduced to 21 gene co-expression modules using weighted
128 gene co-expression network analysis (WGCNA) (34). We then we used partial least squares (PLS)
129 analysis to test for large-scale associations between blood leukocyte co-expression modules and
130 SA or CT phenotypes from the GCLUST parcellation (see Methods for more details). For SA, we
131 identified one statistically significant latent variable (LV) pair (SA LV1: $d = 3.99, p = 0.0001$),
132 which explains 36% of the covariance between SA and gene expression. To decompose how this

l33 multivariate relationship manifests across co-expression modules and groups, in Fig. 2D we show
l34 which co-expression modules have ‘non-zero’ relationships in each group. These ‘non-zero
l35 modules’ have 95% confidence intervals (CIs) estimated by bootstrapping that do not include a
l36 correlation of 0 and are thus the most important co-expression modules driving the SA LV1
l37 relationship. In contrast, co-expression modules that we dub as ‘zero modules’ are those whereby
l38 the 95% CIs include a correlation of 0 and thus do not reliably contribute to the overall SA LV1
l39 relationship. Non-zero modules for SA LV1 account for a good majority (68%) of all genes
l40 examined and this effect is compatible with ideas about omnigenic effects on complex traits such
l41 as imaging phenotypes in ASD subtypes (6, 8). Fig. 2D also shows that non-zero modules are
l42 highly similar for ASD Good and TD groups, whereas hardly any non-zero modules are present
l43 for ASD Poor. This similarity between ASD Good and TD can be quantified as a significant
l44 positive correlation in the PLS correlations values for these groups (Fig. 2D-E) ($r = 0.55$, $p =$
l45 0.008). This result indicates that the SA LV1 relationship manifests similarly in TD and ASD Good
l46 groups. In contrast, there is a lack of correlation between ASD Poor and the other groups (ASD
l47 Poor-ASD Good: $r = 0.27$, $p = 0.23$; ASD Poor-TD: $r = -0.41$, $p = 0.06$). Therefore, SA LV1 can
l48 be described as a large-scale SA-gene expression relationship that likely reflects a normative
l49 phenomenon present in TD and which is also preserved in the ASD Good subtype. However, this
l50 normative SA-gene expression relationship is absent in the ASD Poor subtype.

l51
l52 PLS analysis applied to CT data isolated 2 statistically significant LV pairs (CT LV1: $d =$
l53 4.30 , $p = 0.0001$; CT LV2: $d = 3.09$, $p = 0.0001$), explaining 37% and 19% of the covariance
l54 between CT and gene expression respectively. Similar to SA LV1, non-zero modules for CT LV1
l55 comprise a large majority of all genes examined (65%) and are highly similar for ASD Good and
l56 TD, but not ASD Poor (Fig. 2D, F, G). These results indicate that CT LV1 mostly pertains to a
l57 normative relationship preserved across TD and ASD Good, but which is absent in ASD Poor. In
l58 contrast to CT LV1, the non-zero modules for CT LV2 are almost exclusively relevant for the
l59 ASD Poor subtype, comprise about 48% of all genes examined, and do not show strong
l60 correlations between groups (Fig. 2D, F, G). These results indicate that CT LV2 captures a
l61 relationship that is specific to ASD Poor.

l62
l63 Given that our prior work discovered that PLS non-zero modules related to language-
l64 relevant functional neural phenotypes are highly enriched for broadly expressed genes (6), we next
l65 asked if SA and CT non-zero modules were similarly enriched. Indeed, SA LV1 non-zero modules
l66 are highly enriched in broadly expressed genes (enrichment odds ratio (OR)= 3.48, $p = 1.90e-71$)
l67 but not brain-specific genes (OR = 1.67, $p = 0.23$), while no enrichments were present for zero
l68 modules (broadly expressed, OR = 1.10, $p = 0.99$; brain-specific, OR = 0.94, $p = 0.99$). CT LV1
l69 non-zero modules are also highly enriched in broadly expressed genes (OR= 2.96, $p = 4.43e-43$)
l70 but not brain-specific genes (OR = 1.56, $p = 0.55$), while zero modules were not enriched in either
l71 broadly expressed (OR = 1.10, $p = 0.99$) or brain-specific genes (OR = 0.94, $p = 0.99$). In contrast,
l72 CT LV2 showed enrichments for broadly expressed genes in both non-zero (OR= 1.90, $p = 1.31e-$
l73 7) and zero modules (OR = 2.43, $p = 1.34e-28$), but no enrichments for brain-specific genes (non-
l74 zero modules OR = 1.21, $p=0.98$; zero modules OR = 1.40, $p = 0.57$). These results show that SA
l75 and CT LV1 results are largely driven by the class broadly expressed genes, while for CT LV2 the
l76 enrichment for broadly expressed genes is present, but not specific to non-zero modules.

l77

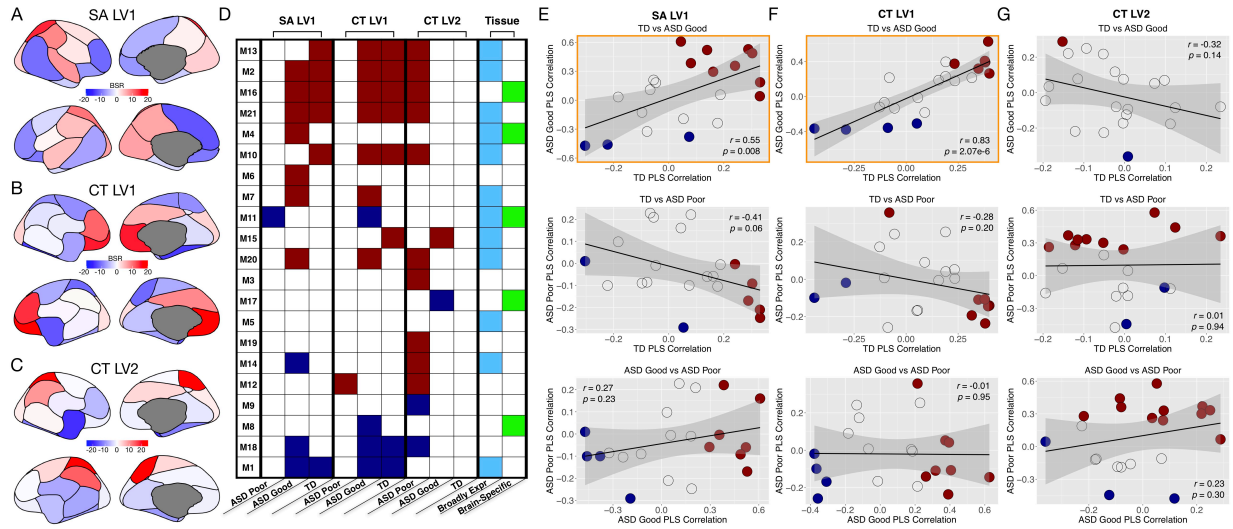


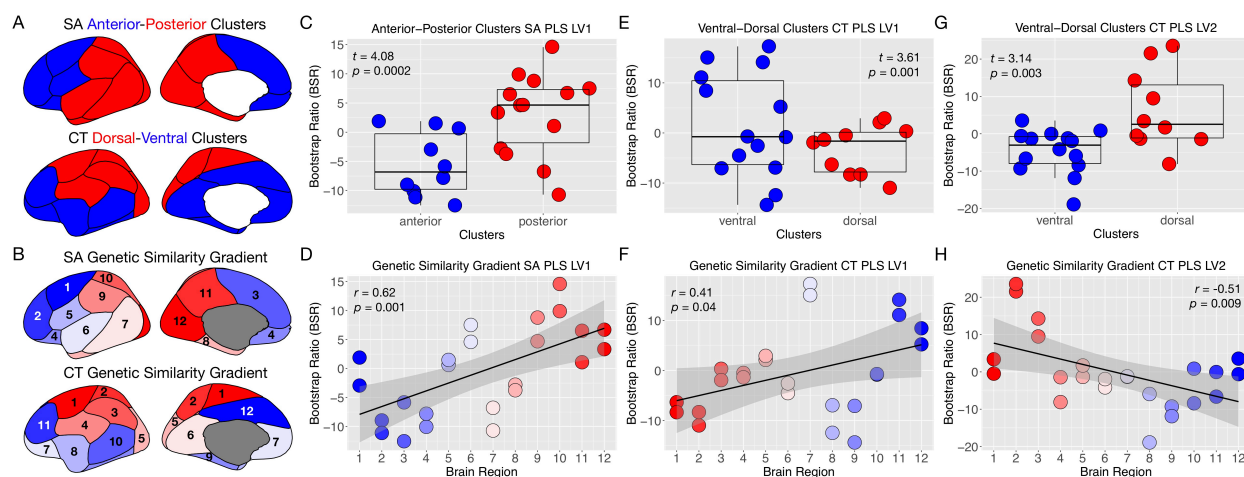
Fig. 2: Multivariate gene co-expression relationships with SA and CT. Panels A-C show brain bootstrap ratios (BSR) for SA LV1 (A), CT LV1 (B), and CT LV2 (C) for all 12 regions from the GCLUST SA and CT parcellations. Regions that are increasingly colored red and blue are regions that most reliably contribute to the PLS relationship. Panel D shows which co-expression modules are ‘non-zero’ modules (dark red or dark blue) or ‘zero’ modules (white). Non-zero modules are co-expression modules where the correlation between gene expression and SA or CT is significantly non-zero, as indicated by 95% bootstrap confidence intervals not encompassing a correlation of 0. These non-zero modules are the strongest contributors to the PLS relationship. All white cells indicate ‘zero’ modules that are not sufficiently correlated in a non-zero way (e.g., 95% bootstrap confidence intervals include a correlation of 0). Non-zero modules in dark red can be interpreted as positive correlations with brain regions in panels A-C colored in red. However, for brain regions colored in blue, the correlations in non-zero modules colored in dark red are interpreted as negative correlations. These interpretations about the directionality of the correlation are reversed when it comes to non-zero modules colored in dark blue. The final two columns show which modules are enriched for broadly expressed or brain-specific genes. Panels E-G show similarity in PLS correlations for all pairwise comparisons for SA LV1 (E), CT LV1 (F), and CT LV2 (G). In these scatterplots each dot is a co-expression module and the x and y-axes indicate the PLS correlations for different groups. Dots colored in dark red and dark blue indicate the non-zero modules, while grey dots indicate zero modules. Scatterplots with the orange outline indicate similar relationships for TD and ASD Good for SA LV1 and CT LV1.

We next investigated how genomic variability patterns SA and CT cortical phenotypes. The patterning of PLS brain bootstrap ratios (BSR) shown in Fig. 2A-C can be used to answer this question. BSRs indicate the directionality through which gene expression is associated with SA and CT and can also show how these relationships manifest similarly or differently across brain regions. It is visually evident from Fig. 2A-C that BSR patterning is not uniform across cortical regions and varies considerably along A-P and D-V axes. With a 2-cluster solution previously identified by Chen and colleagues (25–27) to be the genetically parcellated A-P and D-V axes of SA and CT (Fig. 3A), we confirm that BSRs highly differ along these A-P and D-V clusters (Fig. 3C, E, G). This indicates that the relationship between gene expression and SA or CT at one pole of the A-P or D-V axes is different relative to the other pole.

211 Perhaps even more striking than these differences between binary A-P and D-V partitions
 212 is that BSRs also covary along continuous A-P and D-V genetic similarity gradients. After ordering
 213 regions by genetic similarity gradients discovered by Chen and colleagues (25–27) (Fig. 3B) we
 214 find that BSRs are highly correlated with the ordering along this axis of genetic similarity between
 215 regions (Fig. 3D, F, H). This indicates that large-scale blood leukocyte gene co-expression
 216 relationships with SA and CT reveal how the cortex is genomically patterned to promote the
 217 development of cortical regionalization and areal identity (13). Because SA LV1 and CT LV1 are
 218 normative effects primarily relevant for TD and ASD Good, but not ASD Poor, these results
 219 indicate that normative genomic patterning of the cortex does not occur in the ASD Poor subtype.
 220 Conversely, CT in ASD Poor subtype may be patterned in a completely different way given that
 221 CT LV2 was primarily relevant to this subtype and given that the BSR patterning is reversed for
 222 CT LV2 compared to CT LV1 (Fig. 3E and G versus Fig. 3F and H). Given evidence of focal
 223 laminar patches throughout the cortex in ASD (35), it will be important for future work to
 224 investigate further how such phenomena may be relevant to atypical CT patterning, particularly in
 225 the ASD Poor subtype.

227 In contrast to these effects of genomic patterning along A-P and D-V gradients, we also
 228 examined if the effect size of SA or CT difference between ASD subtypes and TD would similarly
 229 follow A-P and D-V gradients. Prior work using lobar parcellations has suggested that case-control
 230 differences in cortical size may follow an A-P gradient (36). However, effect sizes do not seem to
 231 follow either the 2-cluster A-P and D-V partitions or continuous genetic similarity gradients (Fig.
 232 S1). This result suggests that these cortical patterning effects are not simply effects that can be
 233 seen as on-average group differences in SA or CT and point more towards the specific importance
 234 of how the underlying genomic mechanisms act to pattern SA and CT across the cortex.

235



236
 237 **Fig. 3: Cortical patterning along genetic similarity gradients.** Panel A shows the coarse 2-cluster
 238 anterior-posterior (A-P) and dorsal-ventral (D-V) genetic similarity partitions identified by Chen
 239 and colleagues (25–27). Panel B shows the rank ordering of regions by hierarchical genetic
 240 similarity gradients discovered by Chen and colleagues (25–27). These two parcellations were
 241 utilized to examine how brain BSRs may vary along these genetic similarity gradients. Panels C-
 242 D show A-P and genetic similarity gradients for SA LV1. Panels E-H show D-V (E, G) and genetic
 243 similarity gradients (F, H) for CT LV1 (E, F) and CT LV2 (G, H).

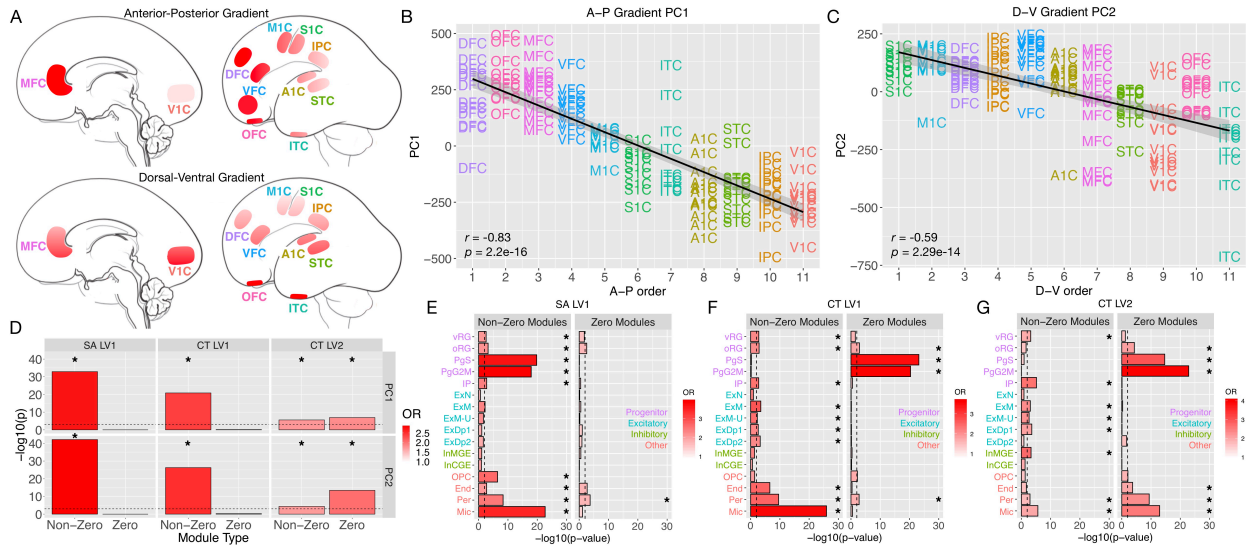
244

245 Because cortical regionalization begins in early prenatal periods from A-P and D-V
246 gradient patterning of gene expression (12–14, 17), we next assessed whether genes from SA and
247 CT non-zero modules are the same genes that play important prenatal roles in the genomic gradient
248 patterning of the cortex. Using the Development PsychENCODE dataset, we used sparse PCA
249 (37) to identify A-P (PC1) and D-V (PC2) gene expression gradients and the most important genes
250 contributing to those gradients from 12 regions of prenatal cortical tissue sampled from 12-24
251 weeks post-conception (e.g., midgestation) (Fig. 4A-C). Remarkably, non-zero module SA LV1
252 and CT LV1 gene sets are highly enriched for genes that drive the prenatal A-P and D-V gradients
253 (Fig. 4D). CT LV2 genes were also enriched for A-P and D-V prenatal gradients, but unlike SA
254 LV1 and CT LV1, the enrichments were apparent for both zero and non-zero modules (Fig. 4D).
255 These results suggest that the genes responsible for the normative SA LV1 and CT LV1
256 relationships are also genes in prenatal periods that act to initialize the regionalization and
257 patterning of cortex along A-P and D-V axes. Since SA LV1 and CT LV1 relationships are largely
258 absent in the ASD Poor subtype, this result suggests that the atypical genomic patterning of SA
259 and CT in this subtype could stem from perturbations in earlier prenatal development.

261 The evidence that SA and CT non-zero modules are enriched for genes that are important
262 for midgestational A-P and D-V expression gradients leaves open the question of what prenatal
263 cell types might drive such effects. The radial unit hypothesis (12) suggests that symmetric cell
264 division in progenitor cell types (e.g., radial glia) in the ventricular zone leads to a substantial
265 proliferation of radial units that then each become their own cortical column and thus, leads to
266 substantial expansion of SA. Variation in this proliferative process in different parts of the
267 ventricular zone protomap regulates regional differences in SA (12, 13, 38). Programmed cell
268 death could also be another mechanism regulating SA (13) and could implicate microglia
269 involvement. In contrast, CT is likely regulated by asymmetric cell division leading to more
270 neurons within particular cortical columns (12) as well as intermediate progenitor cell types (13).
271 CT is also heavily influenced by dendritic arborization (39). While arborization changes over
272 development due to a variety of factors such as experience-dependent pruning, CT and the
273 trajectory it follows over development is also known to be heavily influenced by genetic factors
274 even in middle-aged adults, suggesting that individual differences in CT have a genetic and
275 neurodevelopmental origin (40, 41). Given that cell type markers from midgestational periods are
276 available (42), we next asked if specific prenatal cell type markers are enriched for genes from SA
277 and CT non-zero modules. In striking agreement with prenatal mechanisms hypothesized to affect
278 SA expansion (12, 13), we find that SA LV1 non-zero modules show enrichments for all progenitor
279 cells types - ventricular and outer radial glia (vRG, oRG), cycling progenitors in S and G2M phases
280 of cell cycle (PgS, PgG2M), and intermediate progenitors (IP). In contrast, SA LV1 non-zero
281 modules are devoid of enrichments in later differentiated excitatory (ExM, ExN, ExM-U, ExDp1,
282 ExDp2) and inhibitory (InCGE, InMGE) neurons. Several non-neuronal cells also show SA LV1
283 enrichments, including oligodendrocyte precursors (OPC), endothelial cells (End), and microglia
284 (Mic) (Fig. 4E; Table S1). Similar to SA LV1, CT LV1 and LV2 share enrichments for vRG
285 progenitors. IP cell types are the only other progenitor cell type enriched for CT LV1 and CT LV2
286 non-zero modules, and this effect is compatible with hypothesized effects of IP cells on CT (13).
287 However, CT LV1 and LV2 are differentiated from the enrichment profile of SA LV1 by the
288 presence of enrichments with several types of excitatory neurons (Fig. 4F-G; Table S1). This result
289 indicates a striking contrast between the SA LV1 enrichment profile of primarily progenitor cell
290 types and are compatible with the radial unit and protomap hypotheses (12), differential SA and

291 CT GWAS enrichments (40), and other viewpoints regarding contributors to CT (39). These results
 292 also highlight effects of non-neuronal cell types such as microglia cells. Microglia enrichments are
 293 present and particularly strong for SA LV1 and CT LV1 non-zero modules. This effect may have
 294 implications for programmed cell death and pruning explanations (43) and which may be relevant
 295 to ideas behind ASD-relevant broadly expressed genes and their particularly strong effects on non-
 296 neuronal cell types such as microglia (10).

297



298

299

300 **Fig. 4: Enrichment between PLS non-zero modules and genes involved in prenatal A-P and D-**
 301 **V expression gradients and prenatal cell types.** Panels A shows cortical brain areas sampled from
 302 12-24 weeks post-conception from the Development PsychENCODE RNA-seq dataset from Li and
 303 colleagues (14). AC-PCA (37) was utilized to isolate anterior-posterior (A-P) (PC1, panel B) and
 304 dorsal-ventral (D-V) (PC2, panel C) expression gradients. Panel D shows $-\log_{10}$ p-values for
 305 enrichment tests of non-zero and zero modules for SA LV1, CT LV1, and CT LV2 for genes isolated
 306 from PC1 and PC2. Panels E-F show enrichments in prenatal cell types for SA LV1 (E), CT LV1
 307 (F), and CT LV2 (G). Abbreviations: A-P, anterior-posterior; D-V dorsal-ventral; PC, principal
 308 component; OR, enrichment odds ratio; vRG, ventricular radial glia; oRG, outer radial glia; PgS,
 309 cycling progenitors (S phase); PgG2M, cycling progenitors (G2/M phase); IP, intermediate
 310 progenitors; ExM, maturing excitatory; ExN, migrating excitatory; ExM-U, maturing excitatory
 311 upper enriched; ExDp1, excitatory deep layer 1; ExDp2, excitatory deep layer 2; InCGE,
 312 interneuron caudal ganglion eminence; InMGE, interneuron medial ganglion eminence; OPC,
 313 oligodendrocyte precursor cells; End, endothelial cells; Per, pericytes; Mic, microglia.

314

315

316

317

318

319

320

321

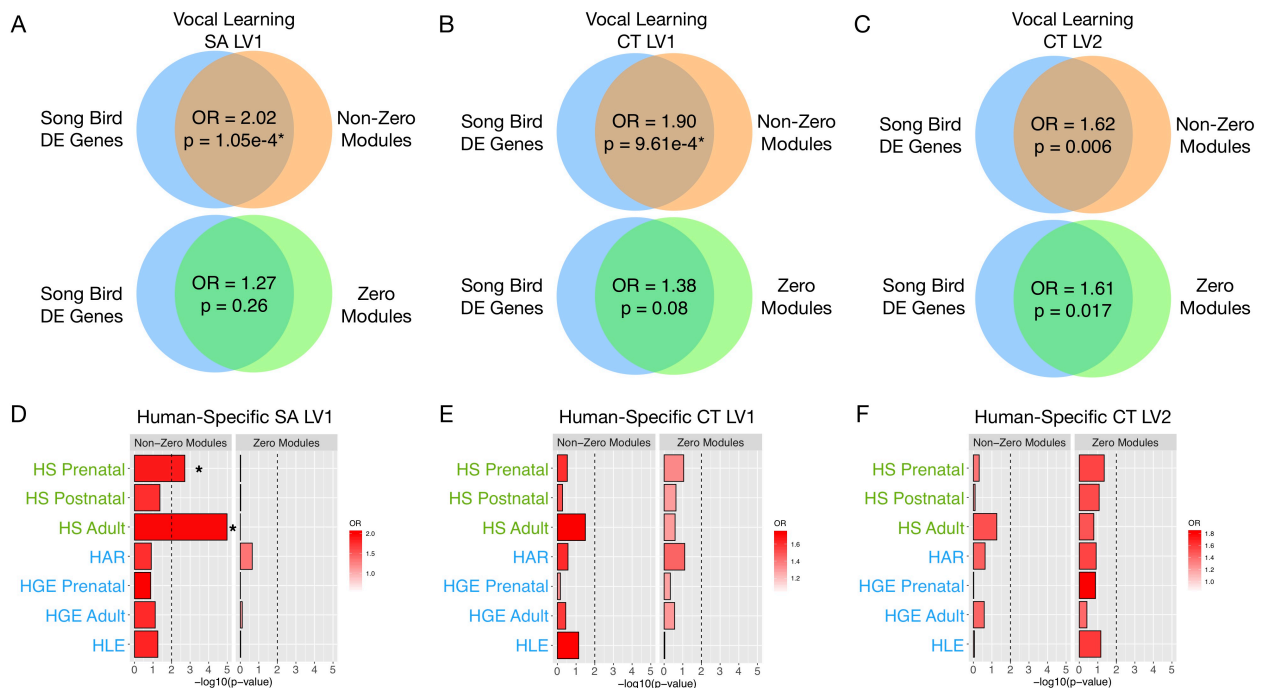
322

323

The results so far suggest that SA and CT non-zero modules are highly prenatally relevant for establishing cortical patterning and regionalization and implicate several cell types that may be of mechanistic importance to different ASD early language outcome subtypes. However, are the SA and CT non-zero modules also functionally relevant for processes that are essential for language development? Our prior work showed that PLS non-zero modules associated with speech-related fMRI response (6) were highly enriched for differentially expressed genes in Area X from a songbird model of vocal learning (44). To test if similar enrichments held up for SA and CT non-zero modules we ran enrichment tests with vocal learning DE genes from Hilliard and colleagues (44). Remarkably, we find similar types of enrichments between DE songbird vocal learning genes and PLS non-zero modules in SA LV1 (OR = 2.02, $p = 1.05e-4$) and CT LV1 (OR

324 = 1.90, $p = 9.61e-4$), but not zero modules ($p > 0.08$) (Fig. 5A-C; Table S1). For CT LV2,
 325 enrichments were present at FDR $q < 0.05$ (but not FDR $q < 0.01$) for both non-zero (OR = 1.62, $p =$
 326 0.006) and zero modules (OR = 1.61, $p = 0.017$). These effects suggest that many genes responsible
 327 for vocal learning in songbirds are conserved and highly represented specifically within SA and
 328 CT non-zero modules that are relevant for groups with relatively intact language (e.g., TD and
 329 ASD Good).
 330

331 Language is a uniquely human ability and there is some evidence that genes implicated in
 332 human-specific evolution are also relevant for autism (45–48). In prior work we found that PLS
 333 non-zero modules associated with speech-related fMRI response (6) were enriched for
 334 differentially expressed genes in the cortex of humans versus non-human primates (i.e. ‘human-
 335 specific’ genes). Given that cortical SA is a phenotype that is dramatically expanded in human
 336 evolution, and much moreso than CT, we investigated the hypothesis of whether SA non-zero
 337 modules would be specifically enriched for human-specific genes. Using 3 lists of human
 338 differentially expressed genes in prenatal, early postnatal, and adulthood periods (47), we find that
 339 SA LV1 non-zero modules are specifically enriched for prenatal and adulthood human-specific
 340 genes (prenatal OR = 1.86, $p = 1.93e-3$; adulthood OR = 1.97, $p = 1.02e-5$) (Fig. 5D; Table S1).
 341 In contrast, no such enrichments are found with genes relevant to CT LV1 or LV2 (Fig. 5E-F). In
 342 addition to differentially expressed genes we also examined genes that are targets of human-
 343 accelerated regions (HAR) or human-gained (HGE) or lossed enhancer (HLE) regions (48).
 344 However, no enrichments for SA or CT were identified for HAR, HGE and HLE genes (Fig. 5D-
 345 F). These results expand on the notion that human-specific genes are of relevance to ASD by
 346 showing that the normative genomic mechanisms associated to SA are also genes of importance
 347 for human-specific evolution. Given that the SA LV1 relationship is absent in ASD Poor, this
 348 suggests that the loss of such normative associations may allow for early SA expansion and
 349 possibly early brain overgrowth for ASD Poor.
 350



351

352 **Fig. 5: Enrichments between PLS non-zero modules and songbird vocal learning or human-**
353 **specific genes.** Panels A-C indicate enrichments between differentially expressed songbird vocal
354 learning genes and non-zero and zero modules for SA LV1 (A), CT LV1 (B), AND CT LV2 (C).
355 Panels D-F indicate enrichments between human-specific genes and non-zero and zero modules
356 for SA LV1 (D), CT LV1 (E), and CT LV2 (F). Asterisks marks enrichments at FDR $q < 0.01$.
357 Abbreviations: DE, differentially expressed; OR, enrichment odds ratio; SA, surface area; CT,
358 cortical thickness; LV, latent variable pair; HS, human-specific; HAR, human-accelerated region;
359 HGE, human-gained enhancer; HLE, human-located enhancer.

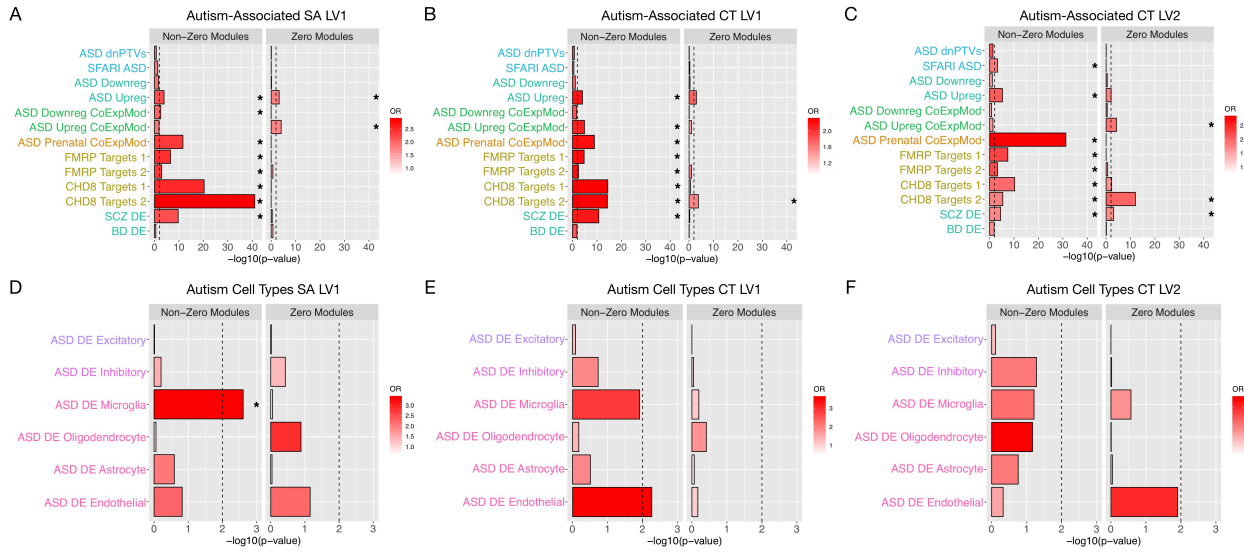
360
361 Next, we asked whether SA and CT non-zero modules were relevant for known autism-
362 associated genomic mechanisms. SA LV1 and CT LV1 non-zero or zero modules are not enriched
363 for rare de novo protein truncating variants (49) or other genes that are annotated as autism-
364 associated in SFARI Gene (50). However, CT LV2 non-zero modules were enriched for SFARI
365 ASD genes (Table S1). Thus, at the level of ASD-risk gene mutations, CT LV2 was the only
366 feature showing enrichments with non-zero modules. This could be compatible with the nature of
367 CT LV2 being mostly specific to the ASD Poor subtype.

368
369 At the level of genes with evidence of ASD-dysregulated expression from post-mortem
370 cortical tissue, we find that both CT LV1 and LV2 non-zero modules were enriched for ASD
371 upregulated genes (51). In contrast, genes from cortically downregulated co-expression modules
372 (19) were highly enriched with genes from SA LV1 non-zero modules (Table S1). This result
373 shows an interesting contrast between CT and genes that show upregulated expression versus SA
374 and genes that show downregulated expression in ASD.

375
376 Non-zero modules from SA LV1, CT LV1, and CT LV2 are also enriched for co-
377 expression modules that are highly transcriptionally active during prenatal periods and which
378 contain many high-penetrance ASD-related mutations (Fig. 6A-C; Table S1). This is compatible
379 with the idea that broadly expressed genes can interact and impact key ASD-risk genes,
380 particularly in prenatal periods (9, 10). Downstream targets of highly penetrant genes like *FMRI*
381 and *CHD8* were also enriched in non-zero modules from SA LV1, CT LV1, and CT LV2.
382 However, not all of these enrichments are specific to autism-associated genes. Genes differentially
383 expressed in schizophrenia (51) were also significantly enriched in non-zero modules across SA
384 LV1, CT LV1, and CT LV2.

385
386 Finally, we examined enrichments with cell type specific differentially expressed genes in
387 autism (52). Here we found that only SA LV1 non-zero modules are enriched for differentially
388 expressed genes in microglia cells (Fig. 6D). No other comparisons for DE cell types were
389 statistically significant. See Fig. 6 and Table S1 for a summary of autism-associated enrichments.
390 The fact that non-zero modules are devoid of enrichments in most DE genes from specific cell
391 types is compatible with the notion that these genes are of primary relevance for early prenatal
392 periods and will not be a highly discoverable DE signal in post-mortem ASD tissue.

393



394
 395 **Fig. 6: Enrichment between PLS non-zero modules and autism-associated genes.** Panels A-C
 396 indicate enrichments between different autism-associated gene lists and non-zero and zero
 397 modules for SA LV1 (A), CT LV1 (B), AND CT LV2 (C). Panels D-F indicate enrichments between
 398 differentially expressed genes in specific cell types in autism and non-zero and zero modules for
 399 SA LV1 (D), CT LV1 (E), and CT LV2 (F). Asterisks marks enrichments at FDR $q < 0.01$.
 400 Abbreviations: DE, differentially expressed; OR, enrichment odds ratio; SA, surface area; CT,
 401 cortical thickness; LV, latent variable pair; dnPTVs, de novo protein truncating variants.
 402

403 To aid future work examining specific genes of interest, we focused on identifying high-
 404 confidence ASD-risk genes (annotated as the ‘high-confidence’ category 1 list in SFARI Gene)
 405 that are also SA-and prenatally-relevant progenitor and A-P patterning genes (i.e. the intersection
 406 of SFARI ASD, SA non-zero modules, PC1 A-P genes, prenatal progenitor cell types, and ASD
 407 prenatal co-expression modules). *SON* and *BAZ2B* were identified and these genes play roles in
 408 splicing, cell cycle, transcriptional regulation, and chromatin remodeling. For CT LV1 genes, we
 409 next searched for high-confidence ASD-risk genes that were also prenatally-relevant excitatory
 410 and D-V patterning genes (i.e. the intersection of SFARI ASD, CT LV1 non-zero modules, PC2
 411 D-V genes, prenatal excitatory cell types, and ASD prenatal co-expression modules). Here we find
 412 ASD ‘high-confidence’ genes of *ATRX*, *AUTS2*, and *BCL11A*. In a similar search within CT LV2
 413 non-zero modules of prenatal relevance to excitatory neurons and D-V patterning, we identified
 414 *ATRX*, *AUTS2*, *BCL11A*, *CACNA1E*, and *MEIS2* as high-confidence ASD-risk genes. A common
 415 theme of all these CT-relevant genes is their role in chromatin modification and remodeling (with
 416 the exception of *CACNA1E*) and their links to syndromes causing intellectual disability.
 417 Additionally, with the exceptions of *BCL11A* and *CACNA1E*, all SA- and CT-relevant high-
 418 confidence genes listed here fall into the broadly expressed gene list, highlighting the importance
 419 of these high-impact genes in ASD biology (9).
 420

421 These findings represent a significant enhancement to the mechanistic and clinical
 422 precision of our understanding of the early brain basis behind the autisms (1). Along with our prior
 423 work (6, 7), this work showcases that the ASD Poor subtype indeed a distinct subtype with
 424 multiscale differentiation across development, behavior, and underlying neural systems. While
 425 prior work showed a biological distinction in this subtype with fMRI (6, 7), this work shows that
 426 the biology is also distinct when examining structural neural phenotypes like SA and CT. While

127 on-average early brain overgrowth is one of the most robust findings in the literature on
128 neurodevelopment in ASD (4, 31), this work shows that the effect is driven by a subtype of ASD
129 toddlers with poor early language development and outcome.
130

131 This work also uncovers an altogether new discovery behind how cortical SA and CT
132 phenotypes are atypically genomically patterned in the ASD Poor subtype. Genomic patterning of
133 SA and CT occurs along A-P and D-V gradients, thus enabling the development of cortical areal
134 and circuit identities (11, 13, 25–27). This A-P and D-V genomic patterning of SA and CT is intact
135 in TD and ASD Good, but absent in ASD Poor (e.g., Fig. 2D). Atypical genomic patterning of the
136 cortex in ASD Poor could be the key neural explanation behind why these individuals have much
137 more pervasive and more severe behavioral difficulties and poor outcomes. Prior work has shown
138 that molecular identity defined by gene expression affects cell type specific neurophysiological
139 response (53). Thus, without intact genomic patterning of the cortex it may be that development
140 of regional or circuit level identities may be perturbed in ASD Poor and this may help explain the
141 phenotypic difficulties in complex information processing in domains like language and social
142 communication.
143

144 The results also shed insight into the developmental and mechanistic origins at the root of
145 the ASD Poor subtype. Evidence suggests that these SA and CT-relevant genes are the same genes
146 responsible in early prenatal periods for establishing these A-P and D-V gene expression gradients
147 across the cortex. The genes responsible for this atypical prenatal genomic patterning are massive
148 in scale, encompassing a large majority of the genes examined. This result is compatible with ideas
149 from the omnigenic model of complex traits (8). The omnigenic model also proposed that broadly
150 expressed genes should have large impact on complex traits encompassed by neuropsychiatric
151 phenotypes and diagnoses. Indeed, broadly expressed genes manifest in this and other studies (6)
152 and are a key class of ASD genetic risk that operates at early prenatal timepoints (9, 10). In future
153 work it will be important to explore how genomic patterning of the cortex can affect other ASD
154 subtypes. Additionally, it will be important for future work to investigate how omnigenic and
155 broadly expressed genes such as those identified here may play roles in other atypical multiscale
156 phenomena in ASD.

157
158

159

160

References and Notes:

161

1. D. H. Geschwind, P. Levitt, Autism spectrum disorders: developmental disconnection syndromes. *Curr. Opin. Neurobiol.* **17**, 103–111 (2007).

162

163

2. M.-C. Lai, M. V. Lombardo, S. Baron-Cohen, Autism. *Lancet.* **383**, 896–910 (2014).

164

3. M. V. Lombardo, M.-C. Lai, S. Baron-Cohen, Big data approaches to decomposing heterogeneity across the autism spectrum. *Mol. Psychiatry.* **24**, 1435–1450 (2019).

165

166

4. E. Courchesne, T. Pramparo, V. H. Gazestani, M. V. Lombardo, K. Pierce, N. E. Lewis, The ASD Living Biology: from cell proliferation to clinical phenotype. *Mol. Psychiatry.* **24**, 88–107 (2019).

167

168

169

5. N. N. Parikshak, M. J. Gandal, D. H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015).

170

171

6. M. V. Lombardo, T. Pramparo, V. Gazestani, V. Warriar, R. A. I. Bethlehem, C. Carter Barnes, L. Lopez, N. E. Lewis, L. Eyler, K. Pierce, E. Courchesne, Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. *Nat. Neurosci.* **21**, 1680–1688 (2018).

172

173

174

175

7. M. V. Lombardo, K. Pierce, L. T. Eyler, C. Carter Barnes, C. Ahrens-Barbeau, S. Solso, K. Campbell, E. Courchesne, Different functional neural substrates for good and poor language outcome in autism. *Neuron.* **86**, 567–577 (2015).

176

177

178

8. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell.* **169**, 1177–1186 (2017).

179

180

9. E. Courchesne, V. H. Gazestani, N. E. Lewis, Prenatal Origins of ASD: The When, What, and How of ASD Development. *Trends Neurosci.* **43**, 326–342 (2020).

181

182

10. V. H. Gazestani, A. W. Chiang, E. Courchesne, N. E. Lewis, “Autism genetics perturb prenatal neurodevelopment through a hierarchy of broadly-expressed and brain-specific genes” (preprint, Systems Biology, 2020), , doi:10.1101/2020.05.23.112623.

183

184

185

11. D. D. M. O’Leary, S.-J. Chou, S. Sahara, Area patterning of the mammalian cortex. *Neuron.* **56**, 252–269 (2007).

186

187

12. P. Rakic, Specification of cerebral cortical areas. *Science.* **241**, 170–176 (1988).

188

13. C. R. Cadwell, A. Bhaduri, M. A. Mostajo-Radji, M. G. Keefe, T. J. Nowakowski, Development and Arealization of the Cerebral Cortex. *Neuron.* **103**, 980–1004 (2019).

189

190

14. M. Li, G. Santpere, Y. Imamura Kawasawa, O. V. Evgrafov, F. O. Gulden, S. Pochareddy, S. M. Sunkin, Z. Li, Y. Shin, Y. Zhu, A. M. M. Sousa, D. M. Werling, R. R. Kitchen, H. J. Kang, M. Pletikos, J. Choi, S. Muchnik, X. Xu, D. Wang, B. Lorente-Galdos, S. Liu, P. Giusti-

191

192

- 5193 Rodríguez, H. Won, C. A. de Leeuw, A. F. Pardiñas, BrainSpan Consortium, PsychENCODE
5194 Consortium, PsychENCODE Developmental Subgroup, M. Hu, F. Jin, Y. Li, M. J. Owen, M.
5195 C. O'Donovan, J. T. R. Walters, D. Posthuma, M. A. Reimers, P. Levitt, D. R. Weinberger, T.
5196 M. Hyde, J. E. Kleinman, D. H. Geschwind, M. J. Hawrylycz, M. W. State, S. J. Sanders, P.
5197 F. Sullivan, M. B. Gerstein, E. S. Lein, J. A. Knowles, N. Sestan, Integrative functional
5198 genomic analysis of human brain development and neuropsychiatric risks. *Science*. **362**
5199 (2018), doi:10.1126/science.aat7615.
- 5200 15. M. Sur, J. L. R. Rubenstein, Patterning and plasticity of the cerebral cortex. *Science*. **310**, 805–
5201 810 (2005).
- 5202 16. E. S. Monuki, C. A. Walsh, Mechanisms of cerebral cortical patterning in mice and humans.
5203 *Nat. Neurosci.* **4 Suppl**, 1199–1206 (2001).
- 5204 17. M. B. Johnson, Y. I. Kawasawa, C. E. Mason, Z. Krsnik, G. Coppola, D. Bogdanović, D. H.
5205 Geschwind, S. M. Mane, M. W. State, N. Sestan, Functional and evolutionary insights into
5206 human brain development through global transcriptome analysis. *Neuron*. **62**, 494–509 (2009).
- 5207 18. I. Voineagu, X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, J. Mill, R. M. Cantor, B.
5208 J. Blencowe, D. H. Geschwind, Transcriptomic analysis of autistic brain reveals convergent
5209 molecular pathology. *Nature*. **474**, 380–384 (2011).
- 5210 19. N. N. Parikshak, V. Swarup, T. G. Belgard, M. Irimia, G. Ramaswami, M. J. Gandal, C. Hartl,
5211 V. Leppa, L. de la T. Ubieta, J. Huang, J. K. Lowe, B. J. Blencowe, S. Horvath, D. H.
5212 Geschwind, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns
5213 in autism. *Nature*. **540**, 423–427 (2016).
- 5214 20. M. L. Chow, T. Pramparo, M. E. Winn, C. C. Barnes, H.-R. Li, L. Weiss, J.-B. Fan, S. Murray,
5215 C. April, H. Belinson, X.-D. Fu, A. Wynshaw-Boris, N. J. Schork, E. Courchesne, Age-
5216 dependent brain gene expression and copy number anomalies in autism suggest distinct
5217 pathological processes at young versus mature ages. *PLoS Genet.* **8**, e1002592 (2012).
- 5218 21. J. L. Freese, D. Pino, S. J. Pleasure, Wnt signaling in development and disease. *Neurobiol.*
5219 *Dis.* **38**, 148–153 (2010).
- 5220 22. V. H. Gazestani, T. Pramparo, S. Nalabolu, B. P. Kellman, S. Murray, L. Lopez, K. Pierce, E.
5221 Courchesne, N. E. Lewis, A perturbed gene network containing PI3K–AKT, RAS–ERK and
5222 WNT– β -catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat*
5223 *Neurosci.* **22**, 1624–1634 (2019).
- 5224 23. M. C. Marchetto, H. Belinson, Y. Tian, B. C. Freitas, C. Fu, K. Vadodaria, P. Beltrao-Braga,
5225 C. A. Trujillo, A. P. D. Mendes, K. Padmanabhan, Y. Nunez, J. Ou, H. Ghosh, R. Wright, K.
5226 Brennand, K. Pierce, L. Eichenfield, T. Pramparo, L. Eyler, C. C. Barnes, E. Courchesne, D.
5227 H. Geschwind, F. H. Gage, A. Wynshaw-Boris, A. R. Muotri, Altered proliferation and
5228 networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry*. **22**, 820–
5229 835 (2017).

- 530 24. T. N. Turner, K. Sharma, E. C. Oh, Y. P. Liu, R. L. Collins, M. X. Sosa, D. R. Auer, H. Brand,
531 S. J. Sanders, D. Moreno-De-Luca, V. Pihur, T. Plona, K. Pike, D. R. Soppet, M. W. Smith,
532 S. W. Cheung, C. L. Martin, M. W. State, M. E. Talkowski, E. Cook, R. Haganir, N. Katsanis,
533 A. Chakravarti, Loss of δ -catenin function in severe autism. *Nature*. **520**, 51–56 (2015).
- 534 25. C.-H. Chen, M. S. Panizzon, L. T. Eyler, T. L. Jernigan, W. Thompson, C. Fennema-Notestine,
535 A. J. Jak, M. C. Neale, C. E. Franz, S. Hamza, M. J. Lyons, M. D. Grant, B. Fischl, L. J.
536 Seidman, M. T. Tsuang, W. S. Kremen, A. M. Dale, Genetic influences on cortical
537 regionalization in the human brain. *Neuron*. **72**, 537–544 (2011).
- 538 26. C.-H. Chen, E. D. Gutierrez, W. Thompson, M. S. Panizzon, T. L. Jernigan, L. T. Eyler, C.
539 Fennema-Notestine, A. J. Jak, M. C. Neale, C. E. Franz, M. J. Lyons, M. D. Grant, B. Fischl,
540 L. J. Seidman, M. T. Tsuang, W. S. Kremen, A. M. Dale, Hierarchical genetic organization of
541 human cortical surface area. *Science*. **335**, 1634–1636 (2012).
- 542 27. C.-H. Chen, M. Fiecas, E. D. Gutiérrez, M. S. Panizzon, L. T. Eyler, E. Vuoksima, W. K.
543 Thompson, C. Fennema-Notestine, D. J. Hagler, T. L. Jernigan, M. C. Neale, C. E. Franz, M.
544 J. Lyons, B. Fischl, M. T. Tsuang, A. M. Dale, W. S. Kremen, Genetic topography of brain
545 morphology. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17089–17094 (2013).
- 546 28. D. van der Meer, O. Frei, T. Kaufmann, C.-H. Chen, W. K. Thompson, K. S. O’Connell, J.
547 Monereo Sánchez, D. E. J. Linden, L. T. Westlye, A. M. Dale, O. A. Andreassen, Quantifying
548 the Polygenic Architecture of the Human Cerebral Cortex: Extensive Genetic Overlap between
549 Cortical Thickness and Surface Area. *Cereb. Cortex* (2020), doi:10.1093/cercor/bhaa146.
- 550 29. E. Courchesne, R. Carper, N. Akshoomoff, Evidence of brain overgrowth in the first year of
551 life in autism. *JAMA*. **290**, 337–344 (2003).
- 552 30. E. Courchesne, P. R. Mouton, M. E. Calhoun, K. Semendeferi, C. Ahrens-Barbeau, M. J.
553 Hallet, C. C. Barnes, K. Pierce, Neuron number and size in prefrontal cortex of children with
554 autism. *JAMA*. **306**, 2001–2010 (2011).
- 555 31. R. Sacco, S. Gabriele, A. M. Persico, Head circumference and brain size in autism spectrum
556 disorder: A systematic review and meta-analysis. *Psychiatry Res.* **234**, 239–251 (2015).
- 557 32. H. C. Hazlett, M. D. Poe, G. Gerig, M. Styner, C. Chappell, R. G. Smith, C. Vachet, J. Piven,
558 Early brain overgrowth in autism associated with an increase in cortical surface area before
559 age 2 years. *Arch. Gen. Psychiatry*. **68**, 467–476 (2011).
- 560 33. H. C. Hazlett, H. Gu, B. C. Munsell, S. H. Kim, M. Styner, J. J. Wolff, J. T. Ellison, M. R.
561 Swanson, H. Zhu, K. N. Botteron, D. L. Collins, J. N. Constantino, S. R. Dager, A. M. Estes,
562 A. C. Evans, V. S. Fonov, G. Gerig, P. Kostopoulos, R. C. McKinstry, J. Pandey, S. Paterson,
563 J. R. Pruett, R. T. Schultz, D. W. Shaw, L. Zwaigenbaum, J. Piven, IBIS Network, Clinical
564 Sites, Data Coordinating Center, Image Processing Core, Statistical Analysis, Early brain
565 development in infants at high risk for autism spectrum disorder. *Nature*. **542**, 348–351 (2017).
- 566 34. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis.
567 *BMC Bioinformatics*. **9**, 559 (2008).

- 568 35. R. Stoner, M. L. Chow, M. P. Boyle, S. M. Sunkin, P. R. Mouton, S. Roy, A. Wynshaw-Boris,
569 S. A. Colamarino, E. S. Lein, E. Courchesne, Patches of disorganization in the neocortex of
570 children with autism. *N. Engl. J. Med.* **370**, 1209–1219 (2014).
- 571 36. R. A. Carper, P. Moses, Z. D. Tigue, E. Courchesne, Cerebral lobes in autism: early
572 hyperplasia and abnormal age effects. *Neuroimage.* **16**, 1038–1051 (2002).
- 573 37. Z. Lin, C. Yang, Y. Zhu, J. Duchi, Y. Fu, Y. Wang, B. Jiang, M. Zamanighomi, X. Xu, M. Li,
574 N. Sestan, H. Zhao, W. H. Wong, Simultaneous dimension reduction and adjustment for
575 confounding variation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14662–14667 (2016).
- 576 38. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo,
577 M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X.
578 Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene
579 expression trajectories reveal developmental hierarchies of the human cortex. *Science.* **358**,
580 1318–1323 (2017).
- 581 39. K. Wagstyl, J. P. Lerch, in *Brain Morphometry*, G. Spalletta, F. Piras, T. Gili, Eds. (Springer
582 New York, New York, NY, 2018; http://link.springer.com/10.1007/978-1-4939-7647-8_3),
583 vol. 136 of *Neuromethods*, pp. 35–49.
- 584 40. K. L. Grasby, N. Jahanshad, J. N. Painter, L. Colodro-Conde, J. Bralten, D. P. Hibar, P. A.
585 Lind, F. Pizzagalli, C. R. K. Ching, M. A. B. McMahon, N. Shatkhina, L. C. P. Zsembik, S.
586 I. Thomopoulos, A. H. Zhu, L. T. Strike, I. Agartz, S. Alhusaini, M. A. A. Almeida, D. Alnæs,
587 I. K. Amlien, M. Andersson, T. Ard, N. J. Armstrong, A. Ashley-Koch, J. R. Atkins, M.
588 Bernard, R. M. Brouwer, E. E. L. Buimer, R. Bülow, C. Bürger, D. M. Cannon, M.
589 Chakravarty, Q. Chen, J. W. Cheung, B. Couvy-Duchesne, A. M. Dale, S. Dalvie, T. K. de
590 Araujo, G. I. de Zubicaray, S. M. C. de Zwarte, A. den Braber, N. T. Doan, K. Dohm, S.
591 Ehrlich, H.-R. Engelbrecht, S. Erk, C. C. Fan, I. O. Fedko, S. F. Foley, J. M. Ford, M.
592 Fukunaga, M. E. Garrett, T. Ge, S. Giddaluru, A. L. Goldman, M. J. Green, N. A. Groenewold,
593 D. Grotegerd, T. P. Gurholt, B. A. Gutman, N. K. Hansell, M. A. Harris, M. B. Harrison, C.
594 C. Haswell, M. Hauser, S. Herms, D. J. Heslenfeld, N. F. Ho, D. Hoehn, P. Hoffmann, L.
595 Holleran, M. Hoogman, J.-J. Hottenga, M. Ikeda, D. Janowitz, I. E. Jansen, T. Jia, C. Jockwitz,
596 R. Kanai, S. Karama, D. Kasperaviciute, T. Kaufmann, S. Kelly, M. Kikuchi, M. Klein, M.
597 Knapp, A. R. Knodt, B. Krämer, M. Lam, T. M. Lancaster, P. H. Lee, T. A. Lett, L. B. Lewis,
598 I. Lopes-Cendes, M. Luciano, F. Macciardi, A. F. Marquand, S. R. Mathias, T. R. Melzer, Y.
599 Milaneschi, N. Mirza-Schreiber, J. C. V. Moreira, T. W. Mühleisen, B. Müller-Myhsok, P.
600 Najt, S. Nakahara, K. Nho, L. M. Olde Loohuis, D. P. Orfanos, J. F. Pearson, T. L. Pitcher, B.
601 Pütz, Y. Quidé, A. Ragothaman, F. M. Rashid, W. R. Reay, R. Redlich, C. S. Reinbold, J.
602 Repple, G. Richard, B. C. Riedel, S. L. Risacher, C. S. Rocha, N. R. Mota, L. Salminen, A.
603 Saremi, A. J. Saykin, F. Schlag, L. Schmaal, P. R. Schofield, R. Seclin, C. Y. Shapland, L.
604 Shen, J. Shin, E. Shumskaya, I. E. Sønderby, E. Sprooten, K. E. Tansey, A. Teumer, A.
605 Thalamuthu, D. Tordesillas-Gutiérrez, J. A. Turner, A. Uhlmann, C. L. Vallerga, D. van der
606 Meer, M. M. J. van Donkelaar, L. van Eijk, T. G. M. van Erp, N. E. M. van Haren, D. van
607 Rooij, M.-J. van Tol, J. H. Veldink, E. Verhoef, E. Walton, M. Wang, Y. Wang, J. M.
608 Wardlaw, W. Wen, L. T. Westlye, C. D. Whelan, S. H. Witt, K. Wittfeld, C. Wolf, T. Wolfers,
609 J. Q. Wu, C. L. Yasuda, D. Zaremba, Z. Zhang, M. P. Zwiers, E. Artiges, A. A. Assareh, R.

- 510 Ayesa-Arriola, A. Belger, C. L. Brandt, G. G. Brown, S. Cichon, J. E. Curran, G. E. Davies,
511 F. Degenhardt, M. F. Dennis, B. Dietsche, S. Djurovic, C. P. Doherty, R. Espiritu, D. Garijo,
512 Y. Gil, P. A. Gowland, R. C. Green, A. N. Häusler, W. Heindel, B.-C. Ho, W. U. Hoffmann,
513 F. Holsboer, G. Homuth, N. Hosten, C. R. Jack, M. Jang, A. Jansen, N. A. Kimbrel, K. Kolskår,
514 S. Koops, A. Krug, K. O. Lim, J. J. Luykx, D. H. Mathalon, K. A. Mather, V. S. Mattay, S.
515 Matthews, J. Mayoral Van Son, S. C. McEwen, I. Melle, D. W. Morris, B. A. Mueller, M.
516 Nauck, J. E. Nordvik, M. M. Nöthen, D. S. O’Leary, N. Opel, M.-L. P. Martinot, G. B. Pike,
517 A. Preda, E. B. Quinlan, P. E. Rasser, V. Ratnakar, S. Reppermund, V. M. Steen, P. A. Tooney,
518 F. R. Torres, D. J. Veltman, J. T. Voyvodic, R. Whelan, T. White, H. Yamamori, H. H. H.
519 Adams, J. C. Bis, S. Dobbie, C. Decarli, M. Fornage, V. Gudnason, E. Hofer, M. A. Ikram, L.
520 Launer, W. T. Longstreth, O. L. Lopez, B. Mazoyer, T. H. Mosley, G. V. Roshchupkin, C. L.
521 Satizabal, R. Schmidt, S. Seshadri, Q. Yang, Alzheimer’s Disease Neuroimaging Initiative,
522 CHARGE Consortium, EPIGEN Consortium, IMAGEN Consortium, SYS Consortium,
523 Parkinson’s Progression Markers Initiative, M. K. M. Alvim, D. Ames, T. J. Anderson, O. A.
524 Andreassen, A. Arias-Vasquez, M. E. Bastin, B. T. Baune, J. C. Beckham, J. Blangero, D. I.
525 Boomsma, H. Brodaty, H. G. Brunner, R. L. Buckner, J. K. Buitelaar, J. R. Bustillo, W. Cahn,
526 M. J. Cairns, V. Calhoun, V. J. Carr, X. Caseras, S. Caspers, G. L. Cavalleri, F. Cendes, A.
527 Corvin, B. Crespo-Facorro, J. C. Dalrymple-Alford, U. Dannlowski, E. J. C. de Geus, I. J.
528 Deary, N. Delanty, C. Depondt, S. Desrivières, G. Donohoe, T. Espeseth, G. Fernández, S. E.
529 Fisher, H. Flor, A. J. Forstner, C. Francks, B. Franke, D. C. Glahn, R. L. Gollub, H. J. Grabe,
530 O. Gruber, A. K. Håberg, A. R. Hariri, C. A. Hartman, R. Hashimoto, A. Heinz, F. A.
531 Henskens, M. H. J. Hillegers, P. J. Hoekstra, A. J. Holmes, L. E. Hong, W. D. Hopkins, H. E.
532 Hulshoff Pol, T. L. Jernigan, E. G. Jönsson, R. S. Kahn, M. A. Kennedy, T. T. J. Kircher, P.
533 Kochunov, J. B. J. Kwok, S. Le Hellard, C. M. Loughland, N. G. Martin, J.-L. Martinot, C.
534 McDonald, K. L. McMahon, A. Meyer-Lindenberg, P. T. Michie, R. A. Morey, B. Mowry, L.
535 Nyberg, J. Oosterlaan, R. A. Ophoff, C. Pantelis, T. Paus, Z. Pausova, B. W. J. H. Penninx, T.
536 J. C. Polderman, D. Posthuma, M. Rietschel, J. L. Roffman, L. M. Rowland, P. S. Sachdev, P.
537 G. Sämann, U. Schall, G. Schumann, R. J. Scott, K. Sim, S. M. Sisodiya, J. W. Smoller, I. E.
538 Sommer, B. St Pourcain, D. J. Stein, A. W. Toga, J. N. Trollor, N. J. A. Van der Wee, D. van
539 ’t Ent, H. Völzke, H. Walter, B. Weber, D. R. Weinberger, M. J. Wright, J. Zhou, J. L. Stein,
540 P. M. Thompson, S. E. Medland, Enhancing NeuroImaging Genetics through Meta-Analysis
541 Consortium (ENIGMA)—Genetics working group, The genetic architecture of the human
542 cerebral cortex. *Science*. **367** (2020), doi:10.1126/science.aay6690.
- 543 41. A. M. Fjell, H. Grydeland, S. K. Krogstad, I. Amlie, D. A. Rohani, L. Ferschmann, A. B.
544 Storsve, C. K. Tamnes, R. Sala-Llonch, P. Due-Tønnessen, A. Bjørnerud, A. E. Salsnes, A. K.
545 Håberg, J. Skranes, H. Bartsch, C.-H. Chen, W. K. Thompson, M. S. Panizzon, W. S. Kremen,
546 A. M. Dale, K. B. Walhovd, Development and aging of cortical thickness correspond to genetic
547 organization patterns. *Proc Natl Acad Sci USA*. **112**, 15462–15467 (2015).
- 548 42. D. Polioudakis, L. de la Torre-Ubieta, J. Langerman, A. G. Elkins, X. Shi, J. L. Stein, C. K.
549 Vuong, S. Nichterwitz, M. Gevorgian, C. K. Opland, D. Lu, W. Connell, E. K. Ruzzo, J. K.
550 Lowe, T. Hadzic, F. I. Hinz, S. Sabri, W. E. Lowry, M. B. Gerstein, K. Plath, D. H. Geschwind,
551 A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-
552 gestation. *Neuron*. **103**, 785-801.e8 (2019).

- 553 43. G. Tang, K. Gudsnuk, S.-H. Kuo, M. L. Cotrina, G. Rosoklija, A. Sosunov, M. S. Sonders, E.
554 Kanter, C. Castagna, A. Yamamoto, Z. Yue, O. Arancio, B. S. Peterson, F. Champagne, A. J.
555 Dwork, J. Goldman, D. Sulzer, Loss of mTOR-Dependent Macroautophagy Causes Autistic-
556 like Synaptic Pruning Deficits. *Neuron*. **83**, 1131–1143 (2014).
- 557 44. A. T. Hilliard, J. E. Miller, E. R. Fraley, S. Horvath, S. A. White, Molecular microcircuitry
558 underlies functional specification in a basal ganglia circuit dedicated to vocal learning.
559 *Neuron*. **73**, 537–552 (2012).
- 560 45. G. Konopka, T. Friedrich, J. Davis-Turak, K. Winden, M. C. Oldham, F. Gao, L. Chen, G.-Z.
561 Wang, R. Luo, T. M. Preuss, D. H. Geschwind, Human-specific transcriptional networks in
562 the brain. *Neuron*. **75**, 601–617 (2012).
- 563 46. X. Liu, D. Han, M. Somel, X. Jiang, H. Hu, P. Guijarro, N. Zhang, A. Mitchell, T. Halene, J.
564 J. Ely, C. C. Sherwood, P. R. Hof, Z. Qiu, S. Pääbo, S. Akbarian, P. Khaitovich, Disruption of
565 an Evolutionarily Novel Synaptic Expression Pattern in Autism. *PLoS Biol.* **14**, e1002558
566 (2016).
- 567 47. Y. Zhu, A. M. M. Sousa, T. Gao, M. Skarica, M. Li, G. Santpere, P. Esteller-Cucala, D. Juan,
568 L. Ferrández-Peral, F. O. Gulden, M. Yang, D. J. Miller, T. Marques-Bonet, Y. Imamura
569 Kawasawa, H. Zhao, N. Sestan, Spatiotemporal transcriptomic divergence across human and
570 macaque brain development. *Science*. **362** (2018), doi:10.1126/science.aat8077.
- 571 48. H. Won, J. Huang, C. K. Opland, C. L. Hartl, D. H. Geschwind, Human evolved regulatory
572 elements modulate genes involved in cortical expansion and neurodevelopmental disease
573 susceptibility. *Nat Commun.* **10**, 2396 (2019).
- 574 49. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R.
575 Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B.
576 Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero,
577 C. Dias, B. Aleksic, R. Anney, M. Barbosa, S. Bishop, A. Brusco, J. Bybjerg-Grauholm, A.
578 Carracedo, M. C. Y. Chan, A. G. Chiochetti, B. H. Y. Chung, H. Coon, M. L. Cuccaro, A.
579 Curró, B. Dalla Bernardina, R. Doan, E. Domenici, S. Dong, C. Fallerini, M. Fernández-Prieto,
580 G. B. Ferrero, C. M. Freitag, M. Fromer, J. J. Gargus, D. Geschwind, E. Giorgio, J. González-
581 Peñas, S. Guter, D. Halpern, E. Hansen-Kiss, X. He, G. E. Herman, I. Hertz-Picciotto, D. M.
582 Hougaard, C. M. Hultman, I. Ionita-Laza, S. Jacob, J. Jamison, A. Jugessur, M. Kaartinen, G.
583 P. Knudsen, A. Kolvezon, I. Kushima, S. L. Lee, T. Lehtimäki, E. T. Lim, C. Lintas, W. I.
584 Lipkin, D. Lopergolo, F. Lopes, Y. Ludena, P. Maciel, P. Magnus, B. Mahjani, N. Maltman,
585 D. S. Manoach, G. Meiri, I. Menashe, J. Miller, N. Minshew, E. M. S. Montenegro, D. Moreira,
586 E. M. Morrow, O. Mors, P. B. Mortensen, M. Mosconi, P. Muglia, B. M. Neale, M. Nordentoft,
587 N. Ozaki, A. Palotie, M. Parellada, M. R. Passos-Bueno, M. Pericak-Vance, A. M. Persico, I.
588 Pessah, K. Puura, A. Reichenberg, A. Renieri, E. Riberi, E. B. Robinson, K. E. Samocha, S.
589 Sandin, S. L. Santangelo, G. Schellenberg, S. W. Scherer, S. Schlitt, R. Schmidt, L. Schmitt,
590 I. M. W. Silva, T. Singh, P. M. Siper, M. Smith, G. Soares, C. Stoltenberg, P. Suren, E. Susser,
591 J. Sweeney, P. Szatmari, L. Tang, F. Tassone, K. Teufel, E. Trabetti, M. del P. Trelles, C. A.
592 Walsh, L. A. Weiss, T. Werge, D. M. Werling, E. M. Wigdor, E. Wilkinson, A. J. Willsey, T.
593 W. Yu, M. H. C. Yu, R. Yuen, E. Zachy, E. Agerbo, T. D. Als, V. Appadurai, M. Bækvad-

- 594 Hansen, R. Belliveau, A. Buil, C. E. Carey, F. Cerrato, K. Chambert, C. Churchhouse, S.
595 Dalsgaard, D. Demontis, A. Dumont, J. Goldstein, C. S. Hansen, M. E. Hauberg, M. V.
596 Hollegaard, D. P. Howrigan, H. Huang, J. Maller, A. R. Martin, J. Martin, M. Mattheisen, J.
597 Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen,
598 S. Ripke, A. J. Schork, W. K. Thompson, P. Turley, R. K. Walters, C. Betancur, E. H. Cook,
599 L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A.
700 E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D.
701 Buxbaum, Large-Scale Exome Sequencing Study Implicates Both Developmental and
702 Functional Changes in the Neurobiology of Autism. *Cell* (2020),
703 doi:10.1016/j.cell.2019.12.036.
- 704 50. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I.
705 Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: a community-driven
706 knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. **4**, 36 (2013).
- 707 51. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel,
708 M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran
709 Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof,
710 J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M.
711 Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-
712 level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. **362**, eaat8127
713 (2018).
- 714 52. D. Velmeshev, L. Schirmer, D. Jung, M. Haeussler, Y. Perez, S. Mayer, A. Bhaduri, N. Goyal,
715 D. H. Rowitch, A. R. Kriegstein, Single-cell genomics identifies cell type-specific molecular
716 changes in autism. *Science*. **364**, 685–689 (2019).
- 717 53. S. Mayer, J. Chen, D. Velmeshev, A. Mayer, U. C. Eze, A. Bhaduri, C. E. Cunha, D. Jung, A.
718 Arjun, E. Li, B. Alvarado, S. Wang, N. Lovegren, M. L. Gonzales, L. Szpankowski, A. Leyrat,
719 J. A. A. West, G. Panagiotakos, A. Alvarez-Buylla, M. F. Paredes, T. J. Nowakowski, A. A.
720 Pollen, A. R. Kriegstein, Multimodal Single-Cell Analysis Reveals Physiological Maturation
721 in the Developing Human Neocortex. *Neuron*. **102**, 143-158.e7 (2019).
- 722 54. T. Pramparo, K. Pierce, M. V. Lombardo, C. C. Barnes, S. Marinero, C. Ahrens-Barbeau, S.
723 S. Murray, L. Lopez, R. Xu, E. Courchesne, Prediction of Autism by Translation and
724 Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices.
725 *JAMA Psychiatry*. **72**, 386–394 (2015).
- 726 55. T. Pramparo, M. V. Lombardo, K. Campbell, C. C. Barnes, S. Marinero, S. Solso, J. Young,
727 M. Mayo, A. Dale, C. Ahrens-Barbeau, S. S. Murray, L. Lopez, N. Lewis, K. Pierce, E.
728 Courchesne, Cell cycle networks link gene expression dysregulation, mutation, and brain
729 maldevelopment in autistic toddlers. *Mol. Syst. Biol.* **11**, 841 (2015).
- 730 56. M. V. Lombardo, L. Eyler, A. Moore, M. Datko, C. Carter Barnes, D. Cha, E. Courchesne, K.
731 Pierce, Default mode-visual network hypoconnectivity in an autism subtype with pronounced
732 social visual engagement difficulties. *Elife*. **8** (2019), doi:10.7554/eLife.47427.

- 733 57. K. Pierce, D. Conant, R. Hazin, R. Stoner, J. Desmond, Preference for geometric patterns early
734 in life as a risk factor for autism. *Arch. Gen. Psychiatry*. **68**, 101–109 (2011).
- 735 58. K. Pierce, S. Marinero, R. Hazin, B. McKenna, C. C. Barnes, A. Malige, Eye Tracking Reveals
736 Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism
737 Spectrum Disorder Subtype Associated With Increased Symptom Severity. *Biol. Psychiatry*.
738 **79**, 657–666 (2016).
- 739 59. K. Pierce, V. H. Gazestani, E. Bacon, C. C. Barnes, D. Cha, S. Nalabolu, L. Lopez, A. Moore,
740 S. Pence-Stophaeros, E. Courchesne, Evaluation of the Diagnostic Stability of the Early
741 Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. *JAMA*
742 *Pediatr*. **173**, 578–587 (2019).
- 743 60. K. Pierce, C. Carter, M. Weinfeld, J. Desmond, R. Hazin, R. Bjork, N. Gallagher, Detecting,
744 studying, and treating autism early: the one-year well-baby check-up approach. *J. Pediatr*.
745 **159**, 458-465.e1–6 (2011).
- 746 61. A. Wetherby, B. Prizant, *Communication and Symbolic Behavior Scales Developmental*
747 *Profile, First Normed Edition* (Paul H. Brookes, Baltimore, 2002).
- 748 62. A. M. Wetherby, S. Brosnan-Maddox, V. Peace, L. Newton, Validation of the Infant-Toddler
749 Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age.
750 *Autism*. **12**, 487–511 (2008).
- 751 63. C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, M.
752 Rutter, The autism diagnostic observation schedule-generic: a standard measure of social and
753 communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. **30**, 205–
754 223 (2000).
- 755 64. E. M. Mullen, *Mullen scales of early learning*. (American Guidance Service, Inc, Circle Pine,
756 MN, 1995).
- 757 65. S. Sparrow, D. Cicchetti, Balla, D., *Vineland-II scales of adaptive behavior: survey form*
758 *manual*. (American Guidance Service Inc, Circle Pines, MN, 2005).
- 759 66. A. Schroeder, O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W.
760 Menzel, M. Granzow, T. Ragg, The RIN: an RNA integrity number for assigning integrity
761 values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
- 762 67. P. Du, W. A. Kibbe, S. M. Lin, lumi: a pipeline for processing Illumina microarray.
763 *Bioinformatics*. **24**, 1547–1548 (2008).
- 764 68. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L.
765 Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch,
766 C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang,
767 J. Zhang, Bioconductor: open software development for computational biology and
768 bioinformatics. *Genome Biol.* **5**, R80 (2004).

- 769 69. M. Chikina, E. Zaslavsky, S. C. Sealfon, CellCODE: a robust latent variable approach to
770 differential expression analysis for heterogeneous cell populations. *Bioinformatics*. **31**, 1584–
771 1591 (2015).
- 772 70. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and
773 reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
- 774 71. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and
775 surface reconstruction. *Neuroimage*. **9**, 179–194 (1999).
- 776 72. B. Fischl, A. Liu, A. M. Dale, Automated manifold surgery: constructing geometrically
777 accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med*
778 *Imaging*. **20**, 70–80 (2001).
- 779 73. F. Ségonne, J. Pacheco, B. Fischl, Geometrically accurate topology-correction of cortical
780 surfaces using nonseparating loops. *IEEE Trans Med Imaging*. **26**, 518–529 (2007).
- 781 74. B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, A. M.
782 Dale, Sequence-independent segmentation of magnetic resonance images. *Neuroimage*. **23**
783 **Suppl 1**, S69-84 (2004).
- 784 75. B. Fischl, A. M. Dale, Measuring the thickness of the human cerebral cortex from magnetic
785 resonance images. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11050–11055 (2000).
- 786 76. X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco,
787 M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, B. Fischl,
788 Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of
789 field strength, scanner upgrade and manufacturer. *Neuroimage*. **32**, 180–194 (2006).
- 790 77. B. Fischl, M. I. Sereno, A. M. Dale, Cortical surface-based analysis. II: Inflation, flattening,
791 and a surface-based coordinate system. *Neuroimage*. **9**, 195–207 (1999).
- 792 78. S. S. Ghosh, S. Kakunoori, J. Augustinack, A. Nieto-Castanon, I. Kovelman, N. Gaab, J. A.
793 Christodoulou, C. Triantafyllou, J. D. E. Gabrieli, B. Fischl, Evaluating the validity of volume-
794 based and surface-based brain image registration for developmental cognitive neuroscience
795 studies in children 4 to 11 years of age. *Neuroimage*. **53**, 85–93 (2010).
- 796 79. T. L. Jernigan, T. T. Brown, D. J. Hagler, N. Akshoomoff, H. Bartsch, E. Newman, W. K.
797 Thompson, C. S. Bloss, S. S. Murray, N. Schork, D. N. Kennedy, J. M. Kuperman, C. McCabe,
798 Y. Chung, O. Libiger, M. Maddox, B. J. Casey, L. Chang, T. M. Ernst, J. A. Frazier, J. R.
799 Gruen, E. R. Sowell, T. Kenet, W. E. Kaufmann, S. Mostofsky, D. G. Amaral, A. M. Dale,
300 Pediatric Imaging, Neurocognition and Genetics Study, The Pediatric Imaging,
301 Neurocognition, and Genetics (PING) Data Repository. *Neuroimage*. **124**, 1149–1154 (2016).
- 302 80. J. Levman, P. MacDonald, A. R. Lim, C. Forgeron, E. Takahashi, A pediatric structural MRI
303 analysis of healthy brain development from newborns to young adults. *Hum Brain Mapp*. **38**,
304 5931–5942 (2017).

- 305 81. A. Krishnan, L. J. Williams, A. R. McIntosh, H. Abdi, Partial Least Squares (PLS) methods
306 for neuroimaging: a tutorial and review. *Neuroimage*. **56**, 455–475 (2011).
- 307 82. A. R. McIntosh, N. J. Lobaugh, Partial least squares analysis of neuroimaging data:
308 applications and advances. *Neuroimage*. **23 Suppl 1**, S250-263 (2004).
- 309 83. GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis:
310 multitissue gene regulation in humans. *Science*. **348**, 648–660 (2015).
- 311 84. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis
312 Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx
313 (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA,
314 Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI,
315 Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain
316 Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser
317 Data Integration & Visualization—EBI, Genome Browser Data Integration & Visualization—
318 UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory,
319 Data Analysis & Coordinating Center (LDACC):, NIH program management:, Biospecimen
320 collection:, Pathology:, eQTL manuscript working group:, A. Battle, C. D. Brown, B. E.
321 Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues.
322 *Nature*. **550**, 204–213 (2017).
- 323 85. G. Konopka, T. F. Roberts, Insights into the Neural and Genetic Basis of Vocal
324 Communication. *Cell*. **164**, 1269–1276 (2016).
- 325 86. A. R. Pfenning, E. Hara, O. Whitney, M. V. Rivas, R. Wang, P. L. Roulhac, J. T. Howard, M.
326 Wirthlin, P. V. Lovell, G. Ganapathy, J. Mouncastle, M. A. Moseley, J. W. Thompson, E. J.
327 Soderblom, A. Iriki, M. Kato, M. T. P. Gilbert, G. Zhang, T. Bakken, A. Bongaarts, A.
328 Bernard, E. Lein, C. V. Mello, A. J. Hartemink, E. D. Jarvis, Convergent transcriptional
329 specializations in the brains of humans and song-learning birds. *Science*. **346**, 1256846 (2014).
- 330 87. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers
331 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*
332 *Res*. **43**, e47 (2015).
- 333 88. J. D. Storey, A direct approach to false discovery rates. *Journal of the Royal Statistical Society:*
334 *Series B (Statistical Methodology)*. **64**, 479–498 (2002).
- 335 89. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H.
336 Geschwind, Integrative functional genomic analyses implicate specific molecular pathways
337 and circuits in autism. *Cell*. **155**, 1008–1021 (2013).
- 338 90. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L.
339 Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-
340 Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L.
341 Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin,
342 N. Sestan, M. W. State, Coexpression networks implicate human midfetal deep cortical
343 projection neurons in the pathogenesis of autism. *Cell*. **155**, 997–1007 (2013).

- 344 91. E. Eising, A. Carrion-Castillo, A. Vino, E. A. Strand, K. J. Jakielski, T. S. Scerri, M. S.
345 Hildebrand, R. Webster, A. Ma, B. Mazoyer, C. Francks, M. Bahlo, I. E. Scheffer, A. T.
346 Morgan, L. D. Shriberg, S. E. Fisher, A set of regulatory genes co-expressed in embryonic
347 human brain is implicated in disrupted speech development. *Mol. Psychiatry*. **24**, 1065–1078
348 (2019).
- 349 92. J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley,
350 J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S.
351 Caldejon, A. Carey, C. Cuhaciyan, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. C. Facer, D.
352 Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M.
353 Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N.
354 Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J.
355 Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman,
356 M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M.
357 Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A.
358 Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W.
359 Phillips, N. Sestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, E. S. Lein,
360 Transcriptional landscape of the prenatal human brain. *Nature*. **508**, 199–206 (2014).
- 361 93. J. Cotney, R. A. Muhle, S. J. Sanders, L. Liu, A. J. Willsey, W. Niu, W. Liu, L. Klei, J. Lei, J.
362 Yin, S. K. Reilly, A. T. Tebbenkamp, C. Bichsel, M. Pletikos, N. Sestan, K. Roeder, M. W.
363 State, B. Devlin, J. P. Noonan, The autism-associated chromatin modifier CHD8 regulates
364 other autism risk genes during human neurodevelopment. *Nat Commun*. **6**, 6404 (2015).
- 365 94. A. Sugathan, M. Biagioli, C. Golzio, S. Erdin, I. Blumenthal, P. Manavalan, A. Ragavendran,
366 H. Brand, D. Lucente, J. Miles, S. D. Sheridan, A. Stortchevoi, M. Kellis, S. J. Haggarty, N.
367 Katsanis, J. F. Gusella, M. E. Talkowski, CHD8 regulates neurodevelopmental pathways
368 associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci USA*. **111**,
369 E4468–E4477 (2014).
- 370 95. J. C. Darnell, S. J. Van Driesche, C. Zhang, K. Y. S. Hung, A. Mele, C. E. Fraser, E. F. Stone,
371 C. Chen, J. J. Fak, S. W. Chi, D. D. Licatalosi, J. D. Richter, R. B. Darnell, FMRP stalls
372 ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. **146**, 247–
373 261 (2011).
- 374 96. E. J. Greenblatt, A. C. Spradling, Fragile X mental retardation 1 gene enhances the translation
375 of large autism-related proteins. *Science*. **361**, 709–712 (2018).

376

377

378

379 **Acknowledgments:** We thank all participants and their families for participating in this study.

380

381 **Funding:** This project has received funding from the European Research Council (ERC) under the
382 European Union's Horizon 2020 research and innovation programme under grant agreement No
383 755816 (ERC Starting Grant to MVL). This work was also supported by the following grants to
384 EC, KP, LE, and NEL - NIMH R01-MH080134 (KP), NIMH R01-MH104446 (KP), NFAR grant
385 (KP), NIMH Autism Center of Excellence grant P50-MH081755 (EC, KP), NIMH R01-
386 MH036840 (EC), NIMH R01-MH110558 (EC, NEL), NIMH U01-MH108898 (EC), NIDCD
387 R01-DC016385 (EC, KP, LE, MVL), CDMRP AR130409 (EC), and the Simons Foundation
388 176540 (EC). KC was supported by the Utah Stimulating Access to Research in Residency
389 Transition Scholar (StARRTS) under Award Number 1R38HL143605-01.

390

391 **Author contributions:** Conceptualization: MVL, EC, KP, LE TP. Methodology: MVL, TP, VH, G,
392 NEL, EC, KP, LE. Software: MVL, DJH, AMD, VH, G. Formal analysis: MVL, VH, G, DJH, TP,
393 JS, RAIB, NB. Investigation: LE, KC, CCB, LL, KP, EC. Data curation: EC, KP, LE, TP, VH, G,
394 MVL. Writing - original draft preparation: MVL, EC. Writing - review and editing: MVL, EC,
395 KP, LE, JS RAIB, NB, KC, NEL, VH, G, DJH, AMD. Visualization: MVL. Supervision: EC, KP,
396 LE, NEL. Project administration: EC, KP, LE, NEL. Funding acquisition: MVL, EC, KP, LE,
397 NEL.

398

399 **Competing interests:** None of the authors have any biomedical financial interests or potential
400 conflicts of interest to report.

401

402 **Data and materials availability:** Analysis code is available at [https://github.com/IIT-](https://github.com/IIT-LAND/genomic_cortical_patterning_autisms)
403 [LAND/genomic_cortical_patterning_autisms](https://github.com/IIT-LAND/genomic_cortical_patterning_autisms). Data are publicly available from the NIH National
404 Database for Autism Research (NDAR). Raw and normalized blood gene expression data are also
405 deposited in Gene Expression Omnibus (GSE42133; GSE111175). RNA-seq data from the
406 Development PsychENCODE dataset can be found here: <http://development.psychencode.org>.
407 GTEx data can be found here: <https://www.gtexportal.org>. Microarray data from the songbird
408 vocal learning model can be found in Gene Expression Omnibus (GSE34819).

409

410

011

012 **Supplementary Materials:**

013

014 ***Materials and Methods***

015

016 ***Participants***

017

018 This study was approved by the Institutional Review Board at University of California, San
019 Diego. Parents provided written informed consent according to the Declaration of Helsinki and
020 were paid for their participation. Identical to the approach used in our earlier studies (6, 7, 22, 54–
021 59) toddlers were recruited through two mechanisms: community referrals (e.g., website) or a
022 general population-based screening method called the 1-Year Well-Baby Check-Up Approach
023 (60) that allowed for the prospective study of ASD beginning at 12 months based on a toddler's
024 failure of the CSBS-DP Infant-Toddler Checklist (61, 62). All toddlers were tracked from an intake
025 assessment around 12 months and followed roughly every 12 months until 3–4 years of age. All
026 toddlers, including normal control subjects, participated in a series of tests collected longitudinally
027 across all visits, including the Autism Diagnostic Observation Schedule (ADOS; Module T, 1, or
028 2) (63), the Mullen Scales of Early Learning (64), and the Vineland Adaptive Behavior Scales
029 (65). All testing occurred at the University of California, San Diego Autism Center of Excellence
030 (ACE). No randomization procedures were implemented as part of the data collection process.
031 Data collection and analyses were not performed blind to the conditions of the experiment.

032

033 Stratification of ASD Poor versus ASD Good was made on the basis of Mullen EL and RL
034 T-scores. An ASD toddler was classified as ASD Poor if both Mullen EL and RL T-scores at the
035 final outcome assessment was below 1 standard deviation of the T-score norm of 50 (i.e. $T < 40$).
036 ASD Good labels were made if the toddler had either Mullen EL or RL T-scores within 1 standard
037 deviation or above the normative T-score of 50 (i.e. $T \geq 40$). A total of $n=123$ toddlers had T1
038 structural MRI and gene expression data available. From these 123 toddlers, $n=76$ ASD individuals
039 were examined and were split into the 2 language outcome subtypes - ASD Poor $n=38$ (32 male,
040 6 female; mean age at MRI scan = 29.01 months, SD at fMRI scan = 7.22, range = 12-50 months),
041 ASD Good $n=38$ (28 male, 10 female; mean age at fMRI scan = 29.02 months, SD at fMRI scan
042 = 9.55, range = 14-46 months) and TD $n=47$ (25 male, 22 female; mean age at fMRI scan = 25.91
043 months, SD at fMRI scan = 10.44, range = 13-46 months). ASD subtypes and TD did not
044 statistically differ in age at the time of scanning ($F(2, 120) = 1.62, p = 0.20$). For more demographic
045 and phenotypic information, please see Table S2.

046

047 ***Blood Sample Collection, RNA extraction, quality control and samples preparation***

048

049 Four to six milliliters of blood was collected into EDTA-coated tubes from toddlers on
050 visits when they had no fever, cold, flu, infections or other illnesses, or use of medications for
051 illnesses 72 hours prior blood draw. Blood samples were passed over a LeukoLOCK™ filter
052 (Ambion, Austin, TX, USA) to capture and stabilize leukocytes and immediately placed in a -20°C
053 freezer. Total RNA was extracted following standard procedures and manufacturer's instructions
054 (Ambion, Austin, TX, USA). LeukoLOCK disks (Ambion Cat #1933) were freed from RNA-later

155 and Tri-reagent (Ambion Cat #9738) was used to flush out the captured lymphocyte and lyse the
156 cells. RNA was subsequently precipitated with ethanol and purified through washing and cartridge-
157 based steps. The quality of mRNA samples was quantified by the RNA Integrity Number (RIN),
158 values of 7.0 or greater were considered acceptable (66), and all processed RNA samples passed
159 RIN quality control. Quantification of RNA was performed using Nanodrop (Thermo Scientific,
160 Wilmington, DE, USA). Samples were prepped in 96-well plates at the concentration of 25 ng/ μ l.
161

162 ***Gene expression and data processing***

163
164 RNA was assayed at Scripps Genomic Medicine (La Jolla, CA, USA) for labeling,
165 hybridization, and scanning using the Illumina BeadChips pipeline (Illumina, San Diego, CA,
166 USA) per the manufacturer's instruction. All arrays were scanned with the Illumina BeadArray
167 Reader and read into Illumina GenomeStudio software (version 1.1.1). Raw data was exported
168 from Illumina GenomeStudio, and data pre-processing was performed using the lumi package (67)
169 for R (<http://www.R-project.org>) and Bioconductor (<https://www.bioconductor.org>) (68). Raw and
170 normalized data are part of larger sets deposited in the Gene Expression Omnibus database
171 (GSE42133; GSE111175).
172

173 A larger primary dataset of blood leukocyte gene expression was available from 383
174 samples from 314 toddlers with the age range of 1-to-4 years old. The samples were assayed using
175 the Illumina microarray platform on three batches. The datasets were combined by matching the
176 Illumina Probe ID and probe nucleotide sequences. The final set included a total of 20,194 gene
177 probes. Quality control analysis was performed to identify and remove 23 outlier samples from
178 the dataset. Samples were marked as outlier if they showed low signal intensity (average signal
179 two standard deviations lower than the overall mean), deviant pairwise correlations, deviant
180 cumulative distributions, deviant multi-dimensional scaling plots, or poor hierarchical clustering,
181 as described elsewhere (55). The high-quality dataset included 360 samples from 299 toddlers.
182 High reproducibility was observed across technical replicates (mean Spearman correlation of 0.97
183 and median of 0.98). Thus, we randomly removed one of each of two technical replicates from the
184 primary dataset. From the subjects in the larger primary dataset, n=123 also had MRI data and thus
185 a total of n=105 from the Illumina HT12 platform along with n=18 from the Illumina WG6
186 platform were used in this study. Batch was not asymmetrically distributed across one subgroup
187 more than another, as chi-square analyses on the contingency table between subgroup and batch
188 show no effect ($\chi^2(4) = 0.84, p = 0.93$). ASD subtypes and TD toddlers also did not statistically
189 differ in age at the time of blood sampling ($F(2,120) = 1.27, p = 0.28$). The 20,194 probes were
190 then collapsed to 14,426 genes based on picking the probe with maximal mean expression across
191 samples. Data were quantile normalized and then adjusted for batch effects, sex, and RIN. This
192 batch, sex, and RIN adjusted data were utilized in all further downstream analyses. We also
193 checked for differences in proportion estimates of different leukocyte cell types (i.e. neutrophils,
194 B cells, T cells, NK cells, and monocytes) using the CellCODE deconvolution method (69), but
195 found no evidence of differences across groups for any cell type (see Table S3).
196

197 ***Weighted Gene Co-Expression Network Analysis***

198
199 We reduced the number of features in the gene expression dataset from 14,426 genes down
200 to 21 modules of tightly co-expressed genes. This data reduction step was achieved using weighted

)01 gene co-expression network analysis (WGCNA), implemented within the WGCNA library in R
)02 (34). Correlation matrices estimated with the robust correlation measure of biweight
)03 midcorrelation were computed and then converted into adjacency matrices that retain the sign of
)04 the correlation. These adjacency matrices were then raised to a soft power of 16 (Fig. S2). This
)05 soft power was chosen by finding the first soft power where a measure of R^2 scale-free topology
)06 model fit saturates. The soft power thresholded adjacency matrix was then converted into a
)07 topological overlap matrix (TOM) and then a TOM dissimilarity matrix (e.g., 1-TOM). The TOM
)08 dissimilarity matrix was then input into agglomerative hierarchical clustering using the average
)09 linkage method. Gene modules were defined from the resulting clustering tree, and branches were
)10 cut using a hybrid dynamic tree cutting algorithm (deepSplit parameter = 4) (Fig. S2). Modules
)11 were merged at a cut height of 0.2, and the minimum module size was set to 100. Only genes with
)12 a module membership of $r > 0.2$ were retained within modules. For each gene module, a summary
)13 measure called the module eigengene (ME) was computed as the first principal component of the
)14 scaled (standardized) module expression profiles. We also computed module membership for each
)15 gene and module. Module membership indicates the correlation between each gene and the module
)16 eigengene (see Table S4). Genes that could not be clustered into any specific module are left within
)17 the M0 module, and this module was not considered in any further analyses. Further WGCNA
)18 analyses were run separately within each group in order to check for preservation of detected
)19 modules across groups at a soft power threshold of 16. These analyses all indicated high levels of
)20 preservation ($Z_{summary} > 10$) (70) for all detected modules for each pairwise group comparison
)21 (Fig. S3).

)22

)23 *MRI Data Acquisition and Analyses*

)24

)25 Imaging data were collected on a 1.5 Tesla General Electric MRI scanner during natural
)26 sleep at night; no sedation was used. Structural MRI data was collected with a T1-weighted IR-
)27 FSPGR sagittal protocol (TE = 2.8 ms, TR = 6.5 ms, flip angle = 12 degrees, bandwidth = 31.25
)28 kHz, FOV = 24 cm, slice thickness = 1.2 mm). Cortical surface reconstruction was performed
)29 using FreeSurfer v5.3 (<http://surfer.nmr.mgh.harvard.edu/>) (71–73), which uses routinely acquired
)30 T₁-weighted MRI volumes (74), includes tools for estimation of brain morphometry measures such
)31 as cortical thickness and surface area (75, 76), and enables inter-subject alignment via nonlinear,
)32 surface-based registration to an average brain, driven by cortical folding patterns (77). FreeSurfer
)33 has been validated for use in children (78) and used successfully in large pediatric studies (79, 80).
)34 Total cortical volume, surface area (SA) and mean cortical thickness (CT) were computed based
)35 on the Desikan-Killiany parcellation. Regional SA and CT values were computed from a 12-region
)36 parcellation reported by Chen and colleagues (26, 27) based on genetic similarity in monozygotic
)37 twins. This parcellation scheme, known as GCLUST, is highly relevant for our purposes here,
)38 since the parcellations are based on genetic patterning. Thus, GCLUST should help increase
)39 statistical power while also minimizing multiple comparisons. The GCLUST parcellation is also
)40 important as it can be used to leverage information about genetic similarity gradients (e.g., rank
)41 ordering of regions by fuzzy clustering) in further analyses. The 2-cluster anterior-posterior (A-P)
)42 or dorsal-ventral (D-V) partitions discovered by Chen and colleagues (26, 27) are also relevant in
)43 further analyses for A-P and D-V gradient questions. For all 12 regions of the SA and CT GCLUST
)44 parcellation, global effects were controlled for by dividing SA values by the mean SA, and for CT
)45 we subtracted mean CT from each region, as was done in prior papers using this parcellation
)46 scheme (26, 27).

)47

)48

MRI-Gene Expression Association Analysis

)49

)50

)51

)52

)53

)54

)55

)56

)57

)58

)59

)60

)61

)62

)63

)64

)65

)66

)67

)68

)69

)70

)71

)72

)73

)74

)75

)76

)77

)78

)79

)80

)81

)82

)83

)84

)85

)86

)87

)88

)89

Gene Set Enrichment Analyses

)90

)91

)92

We analyze enrichment between genes from PLS non-zero and zero modules and a host of other gene lists defined by a variety of criteria (see below for details). For these gene set enrichment

l93 analyses, we utilized custom R code written by MVL
l94 (<https://github.com/mvlombardo/utis/blob/master/genelistOverlap.R>) that computes
l95 hypergeometric p-values and enrichment odds ratios. The background pool for these enrichment
l96 tests was always set to 14,426. After all enrichment tests were computed, results are interpreted
l97 only if the enrichment was statistically significant after FDR correction for multiple comparisons
l98 at a threshold of FDR $q < 0.01$.

l99

l100 *Prenatal Gene Expression Gradients and Cell Types*

l101

l102 To assess gradients in prenatal gene expression we utilized RNA-seq data from the
l103 Development PsychENCODE dataset (<http://development.psychencode.org>) (14). The data
l104 utilized was already preprocessed as described by Li and colleagues (14) (e.g., normalized, batch
l105 effects removed) and summarized to RPKM. Sample data from all 12 available cortical regions
l106 from 12-22 weeks post-conception were utilized in order to capture the midgestational window of
l107 interest. Before running the analysis we removed low expressing genes with $\log_2(\text{RPKM})$ below
l108 2. The primary analysis to identify expression gradients was an adjustment-for-confounds
l109 principal components analysis (AC-PCA) (37) which allowed for adjustment due to repeat
l110 measurements from the same donor across sampled brain regions. Rank ordering of regions by A-
l111 P and D-V axes were utilized to statistically confirm that PC1 and PC2 components follow A-P
l112 and D-V gradients. Subsets of the most important genes for the top two principal components were
l113 identified with a sparse AC-PCA analysis, whereby the sparsity parameter, c_2 , was selected based
l114 on a grid search with 10-fold cross validation. These PC1 and PC2 gene sets were used in
l115 enrichment tests with PLS non-zero or zero modules.

l116

l117 We also examined enrichments between PLS non-zero and zero modules and prenatal cell
l118 types identified from single cell RNA-seq on midgestational prenatal brain tissue (42). These cell
l119 types included several classes of progenitor cells (ventricular radial glia, vRG; outer radial glia,
l120 oRG; cycling progenitors (S phase), PgS; cycling progenitors (G2/M phase), PgG2M; intermediate
l121 progenitors, IP), excitatory neurons (migrating excitatory, ExN; maturing excitatory, ExM;
l122 maturing excitatory upper enriched, ExM-U; excitatory deep layer 1, ExDp1; excitatory deep layer
l123 2, ExDp2), inhibitory neurons (interneuron CGE, InCGE; interneuron MGE, InMGE), and other
l124 non-neuronal cell types (oligodendrocyte precursors, OPC; pericytes, Per; endothelial cells, End;
l125 microglia, Mic).

l126

l127 *Tissue-specific enrichments*

l128

l129 To better understand how genes expressed in blood leukocytes could be brain-relevant we
l130 annotated gene co-expression modules based on enrichments in genes known from expression
l131 across multiple tissues to be either broadly expressed or brain-specific. Both of these categories
l132 contain genes that are expressed in cortical tissue, but differ in the pattern of expression across
l133 other non-neuronal tissues. To define these lists we downloaded transcript per million (TPM)
l134 normalized gene expression from 10,259 samples across 26 tissues from the GTEx dataset
l135 (<https://www.gtexportal.org>) (83, 84). In addition to brain and nerve tissue, the dataset included
l136 transcriptome data from 24 non-neuronal tissues, including: Adipose, Adrenal Gland, Blood
l137 Vessel, Breast, Blood, Skin, Colon, Esophagus, Heart, Liver, Lung, Salivary Gland, Muscle,
l138 Ovary, Pancreas, Pituitary, Prostate, Small Intestine, Spleen, Stomach, Testis, Thyroid, Uterus,

l39 and Vagina (Table S2). We next defined a gene expressed in a tissue if it met two criteria. First,
l40 the gene TPM expression level was ≥ 3 in at least half of the samples from the tissue. Second, the
l41 median expression of the gene was equal or larger than its 25-percentile expression in GTEx cortex
l42 samples. The second criterion was included to account for the differences in the base expression
l43 level of the genes and their dosage dependent translation and function. Broadly-expressed genes
l44 were defined as genes that were expressed in $\geq 50\%$ of non-neuronal tissues (i.e., tissues other than
l45 brain and nerve). The broadly-expressed and brain-specific genes included genes that were
l46 expressed in the adult cortex based on GTEx dataset.

l47

l48 *Vocal learning enrichments*

l49

l50 To test for enrichment between PLS non-zero modules and gene sets of functional
l51 relevance for language processes, we examined genes that are differentially expressed in a song
l52 bird vocal learning model. Song birds are often used as animal models relevant for the vocal
l53 learning component of language (44, 85, 86). We investigated enrichments with differentially
l54 expressed genes taken from a microarray dataset of Area X of song birds (44). To identify
l55 differentially expressed (DE) genes between singing versus non-singing birds, we re-analyzed this
l56 dataset (GEO Accession ID: GSE34819) using limma (87), and DE genes were identified if they
l57 passed Storey FDR $q < 0.05$ (88). These DE genes were also used for enrichment tests in our prior
l58 work examining gene expression relationships with language-relevant functional neural
l59 phenotypes measured with fMRI (6).

l60

l61 *Human-specific enrichments*

l62

l63 Given the uniquely human nature of language, we also tested hypotheses regarding
l64 enrichments with genes that are transcriptionally different in the cortical tissue between humans
l65 and other non-human primates across prenatal, early postnatal and adult periods (47). In addition,
l66 we also examined enrichments with genes linked to human accelerated regions (HAR), human-
l67 gained enhancers (HGE) in prenatal and adult tissue, and human-located enhancers (HLE) (48).

l68

l69 *Autism-associated enrichments*

l70

l71 Ample evidence suggests that prenatal periods are critical for ASD (4, 9, 89–91). To test
l72 enrichment with prenatal ASD-associated co-expression modules, we utilized co-expression
l73 modules from a study that analyzed the Allen Institute BrainSpan dataset (92). Parikshak and
l74 colleagues analyzed only cortical regions from BrainSpan and identified M2 and M3 as prenatally
l75 active and enriched for rare protein truncating variants with high penetrance for ASD (89). We
l76 also tested enrichments with gene lists known to be associated with ASD, either from genetic
l77 evidence or evidence from cortical transcriptomic dysregulation. In particular, we examined a list
l78 of 102 rare de novo protein-truncating variants (dnPTV) associated with ASD (49), genes listed as
l79 ASD-associated in SFARI Gene (<https://gene.sfari.org>) in categories S, 1, 2, and 3 (downloaded
l80 on July 16, 2020) (50), and DE genes and cortical co-expression modules measured from ASD
l81 post-mortem frontal and temporal cortex tissue (19, 51). To contrast ASD DE genes to genes that
l82 are DE in other psychiatric diagnoses that are genetically correlated with autism, we also use DE
l83 genes in schizophrenia (SCZ DE) and bipolar disorder (BD DE) from the same study that identified
l84 ASD DE genes (51). To go beyond DE genes identified in bulk tissue samples, we also examined

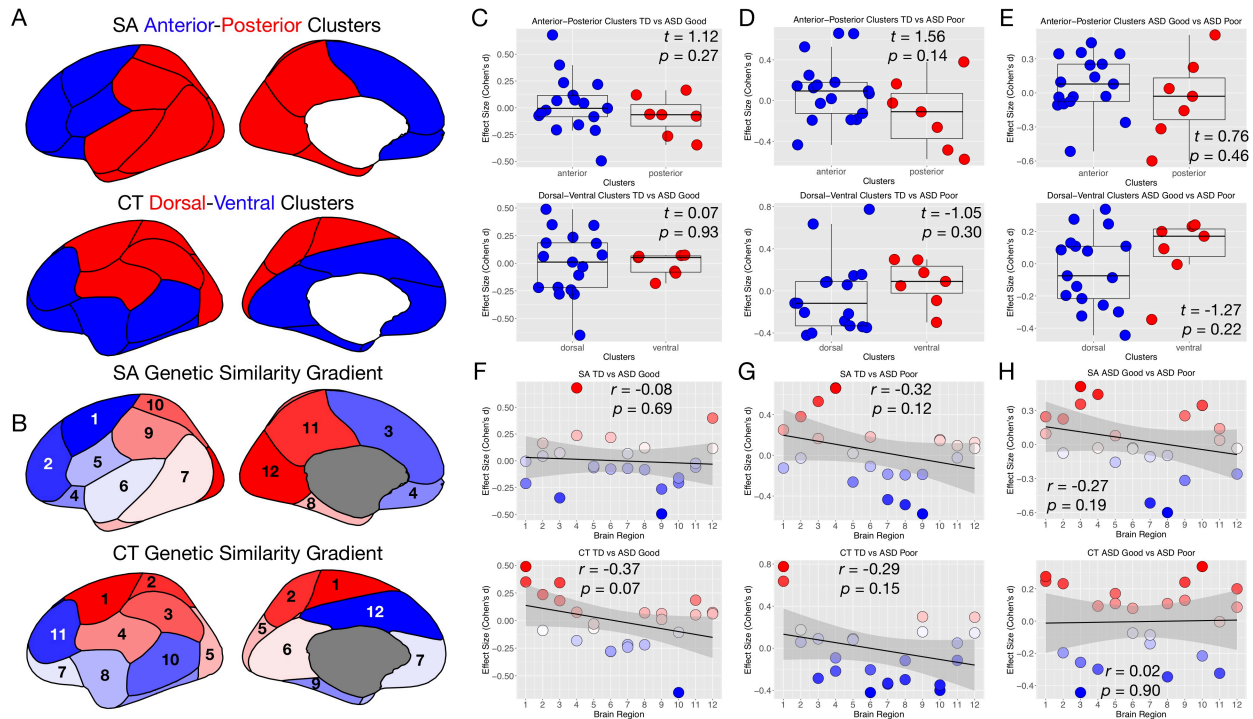
185 ASD DE genes identified in specific cell types - particularly, excitatory (ASD Excitatory) and
186 inhibitory (ASD Inhibitory) neurons, microglia (ASD Microglia), astrocytes (ASD Astrocyte),
187 oligodendrocytes (Oligodendrocyte), and endothelial (ASD Endothelial) cells (52). Finally, we
188 also tested for enrichments with known downstream targets of highly penetrant mutations known
189 to be associated with ASD – FMRP and CHD8. For each, we had lists of downstream targets for
190 two independent studies (93–96), where the overlap for FMRP targets was 3.71% and 27.61% for
191 CHD8 targets.

192
193

[94

[95

Figures S1-S3



[96

[97

[98

[99

[100

[101

[102

[103

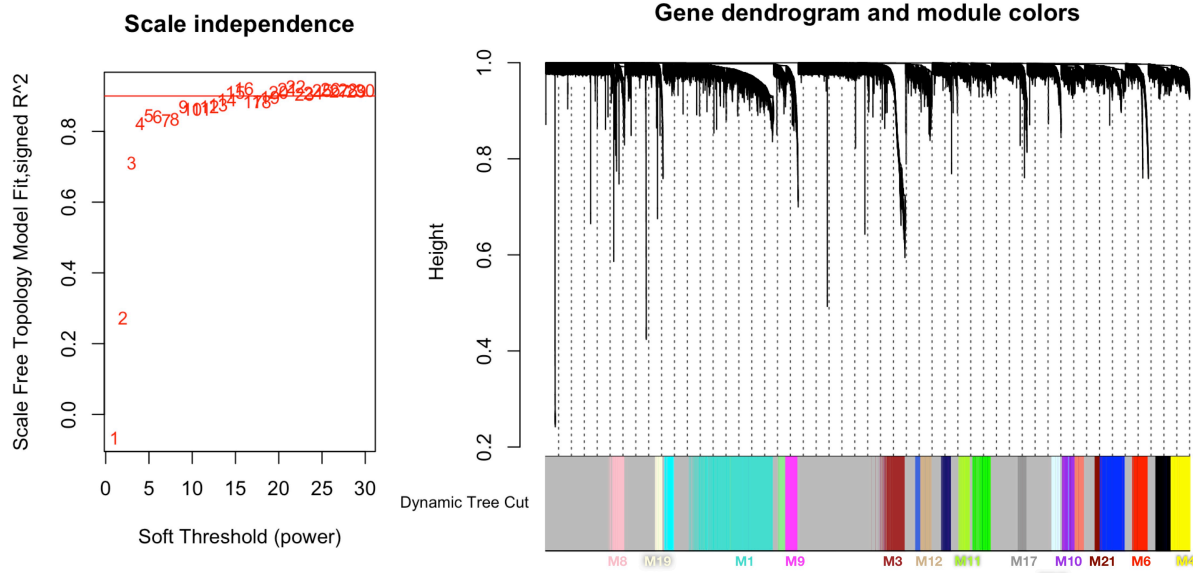
[104

[105

[106

[107

Fig. S1: Lack of A-P and D-V gradients in effect size of group comparisons. Panel A shows a depiction of the A-P and D-V partitions defined by Chen and colleagues (26, 27). Panel B shows how the genetic similarity gradient defined by Chen and colleagues (26, 27) manifests via numbered ordering of brain regions along that gradient. Panels C-E show standardized effect size (Cohen's d) for group comparisons of TD vs ASD Good (C), TD vs ASD Poor (D), and ASD Good vs ASD Poor (E). The top row of panels C-E are for the A-P partition, while the bottom row in each panel is for the D-V partition. Panels F-H show scatterplots of standardized effect size by genetic similarity gradient for each group comparison of TD vs ASD Good (F), TD vs ASD Poor (G), and ASD Good vs ASD Poor (H). Comparisons for SA are shown in the top of each panel in F-H, while CT is shown at the bottom.



208

209

210

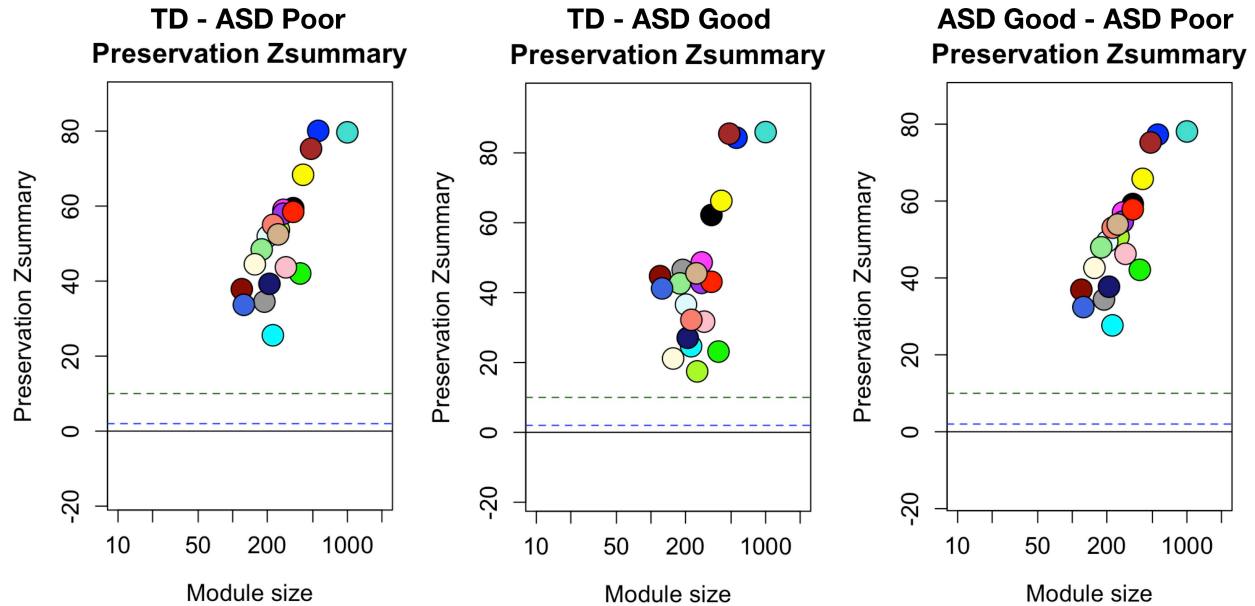
211

212

213

214

Fig. S2: Soft power and TOM dendrogram from WGCNA analysis. On the left of this figure we show the soft power plot for the main WGCNA analysis including data from all groups. A horizontal red line depicts soft power topology model fit R^2 of 0.9, where the chosen soft power of 16 is located. On the right of this figure is the TOM dendrogram with modules labeled at the bottom.



215

216

217

218

Fig. S3: Module preservation when WGCNA analysis is run separately on each group. This figure shows the module preservation Zsummary statistic for WGCNA analyses run separately on each group in order to show that networks are highly preserved ($Zsummary > 10$) across groups.

119

120 **Tables S1-S4**

121

122 *Table S1: Table annotating overlap of each gene with gene sets used in enrichment analyses.*

123

124 *Table S2: Summary of clinical and demographic variables.*

125

126 *Table S3: ANOVA stats from CellCODE deconvolution of leukocyte cell types.*

127

128 *Table S4: WGCNA module assignments for each gene and module membership scores.*

129