Title: Atypical genomic cortical patterning in autism with poor early language outcome

Authors: Michael V. Lombardo[1,2]*, Lisa Eyler[3,4], Tiziano Pramparo[5], Vahid H. Gazestani[5], Donald J. Hagler Jr.[6,7], Chi-Hua Chen[3], Anders M. Dale[6,7,8], Jakob Seidlitz[9,10], Richard A. I. Bethlehem[2,11], Natasha Bertelsen[1,12], Cynthia Carter Barnes[5], Linda Lopez[5], Kathleen Campbell[5,13], Nathan E. Lewis[14,15,16], Karen Pierce[5], & Eric Courchesne[5]

Affiliations:

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

[2] Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

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

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

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

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

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

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

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

[10] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA USA

[11] Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

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

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

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

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

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

*Correspondence to: michael.lombardo@iit.it

Abstract

Cortical regionalization develops via genomic patterning along anterior-posterior (A-P) and dorsal-ventral (D-V) gradients. Here we find that normative A-P and D-V genomic patterning of cortical surface area (SA) and thickness (CT) present in typically developing and autistic toddlers with good early language outcome, is absent in autistic toddlers with poor early language outcome. Autistic toddlers with poor early language outcome are instead specifically characterized by secondary and independent genomic patterning effect on CT. Genes involved in these effects can be traced back to midgestational A-P and D-V gene expression gradients and different prenatal cell (e.g., progenitor cells, excitatory neurons), are functionally important for vocal learning and human-specific evolution, and are prominent in prenatal co-expression networks enriched for high-penetrance autism risk genes. Autism with poor early language outcome may be explained by atypical genomic cortical patterning starting in prenatal development, which may detrimentally impact later regional functional specialization and circuit formation.

Keywords: autism, heterogeneity, language, outcome, prenatal, cortical patterning, gradients, gene expression, surface area, cortical thickness

It is widely accepted that autism spectrum disorder (ASD) is etiologically and clinically highly heterogeneous (*1*). This heterogeneity is theorized to manifest as a complex multiscale cascade from a diverse genetic architecture that converges onto a common set of downstream hierarchical mechanisms linked to the domains of early social-communication and restricted repetitive behaviors (*2–4*). Within the context of precision medicine (*5*), the major priorities for the field are to best understand how this complex multiscale cascade takes place for individuals with differing clinical outcomes and to isolate what could the common set of downstream mechanisms be for such individuals (*1*, *4*, *6*).

Regarding different clinical outcomes in ASD, perhaps the most understudied, yet most important distinction is between those with relatively intact and good language levels versus those who are minimally verbal or have very poor early language outcome (*7–9*). ASD individuals with poor language represent a sizeable proportion of the early diagnosed population and are the most in need of intervention to facilitate better outcomes (*9*). However, since language level is a key ingredient in helping to facilitate better outcomes, available early interventions may be least effective for these types of individuals, (*10–13*). A better understanding of the underlying biology behind this good versus poor language distinction may be key to developing new individualized interventions that may be more effective at facilitating better outcomes. Thus, a key question looms about whether good versus poor language in ASD represents a biologically distinct subtype with different multiscale biological cascades from genomics, up to neural phenotypes, and through to behavior. If good versus poor language signals a biologically distinct subtype, what is the common downstream explanation for how diverse genetic mechanisms lead to altered brain and behavioral phenotypic development?

At the nexus of this puzzle, our prior functional imaging (fMRI) work showed that the neural systems responsible for language respond differently between good (i.e. ASD Good) and poor (i.e. ASD Poor) early language outcome subtypes (*8*). Linked to this functional abnormality are different large-scale patterns of activity in blood leukocyte gene co-expression modules (*14*). This work suggests that early cortical functional specialization for language is lacking in the ASD Poor subtype and that large-scale functional genomic signal may explain this type of pathology. This large-scale functional genomic signal can be characterized as an omnigenic (*15*) array of genes that are typically broadly expressed across many organs and tissues, including the brain, and are highly active during prenatal periods of development (*14*). This prenatal enrichment is key, since a large proportion of broadly expressed ASD-risk genes remarkably show peak expression during early prenatal periods when processes such as cell proliferation, differentiation, neurogenesis, and migration are highly prominent (*16*, *17*). If these genes affect proliferation, differentiation, neurogenesis, and migration, it follows then that macroscale structural features of the developing cerebral cortex that are predicated on such processes, such as surface area (SA) and cortical thickness (CT) (*18–22*), may also be substantially altered in ASD Poor versus Good language outcome subtypes.

The prenatal actions of broadly expressed genes may also be important for telling us about some emergent consequences of such perturbations on how the cortex is genomically patterned, and thus regionally and functionally differentiated. 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 (*18*, *22–27*). 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 (*18*, *23*, *25*, *26*) and thus aid in later regional functional specialization. Indeed, in the adult brain this genomic gradient patterning is still evident and correlates with large-scale gradient patterning of structural and functional features (*19*, *20*, *28–30*). Cortical arealization or patterning may also be atypical in ASD. Functional connectome gradient organization is altered in ASD (*31*). Case-control comparisons of gene expression in post-mortem cortical tissue has found dysregulation of cortical patterning genes and attenuation of gene expression differences in frontal versus temporal cortex (*32–34*). WNT-signaling is also known to affect cortical patterning (*18*, *25*, *26*, *35*) and WNT-signaling abnormalities are also identified in ASD (*33*, *34*, *36–38*), particularly within broadly expressed ASD-risk genes (*17*). Therefore, if broadly expressed genes prenatally impact proliferation, differentiation, neurogenesis, and migration processes differently in the ASD Good versus Poor subtypes, could this explain the lack of functional specialization seen in prior work (e.g., (*8*, *14*))?

Here we investigated these questions by examining how early variability in morphometric measures of the cerebral cortex such as cortical thickness (CT) and surface area (SA) are patterned by large-scale variability in gene expression measured in blood leukocytes. We find that CT and SA associations with large-scale gene expression patterns are different in ASD Poor versus Good early language outcome subtypes. This difference can be described as the absence of normative genomic patterning of CT and SA in the ASD Poor subtype along anterior-posterior (A-P) and dorsal-ventral (D-V) gradients, and the establishment of a second unique type of patterning of CT specific to the ASD Poor subtype. These A-P and D-V genomic patterning effects on CT and SA comprise many of the same genes involved in actual prenatal A-P and D-V gene expression gradients and prenatal cell-types predicted to be involved in SA and CT (*18*, *22*, *39*). Consequently, these atypical genomic cortical patterning effects have important functional consequences, with enrichments in genes important for vocal learning, human evolution, and known ASD-associated genomic mechanisms.

Results

Enlargements of cortical volume and surface area in the ASD Poor language subtype

In this study we examined a cohort of n=123 toddlers (mean age in months = 27.82, SD = 9.32) with and without ASD (ASD n = 76, TD n = 47) with both a T1-weighted structural MRI scan and a blood sample that was used to examine gene expression in blood leukocyte cells (see Methods for sample description and Table S1 for characterization of the sample). The n=76 ASD toddlers were split into ASD Poor (n=38) and ASD Good (n=38) early language outcome subtypes using 1 standard deviation cutoffs on Mullen expressive and receptive language T-scores as in previous studies (see Methods) (*8*, *14*).

One of the most robust findings on early structural brain development in ASD is the on-average effect of early brain overgrowth in the first years of life (*6*, *40–42*). Thus, we started by examining whether there are subtype differences on global measures such as total cortical volume (CV), SA and mean CT. Statistical models controlling for age and sex identified a group effect for

total CV and total SA, but no effect of group for mean CT (Table S2). The group effects for CV and SA were explained by enlargements in ASD Poor versus TD (Fig. 1). Upon examining regional level SA or CT effects while adjusting for global differences, we find no evidence of SA or CT group differences for any region within the GCLUST parcellation (Table S3). These results generally indicate that ASD Poor explains the on-average effect of early brain overgrowth in ASD. These effects are restricted to CV and SA and are not apparent in regional measures after global differences are accounted.

Normative associations between gene expression and surface area or cortical thickness are preserved in TD and ASD Good, but are absent in ASD Poor language subtypes

To identify large-scale associations between gene expression and regional SA or CT we used weighted gene co-expression network analysis (WGCNA) to reduce expression data of 14,426 genes highly expressed in blood leukocytes to 21 co-expression modules (Table S4). Co-expression modules were summarized by the module eigengene and were input into a partial least squares (PLS) analysis to test for large-scale multivariate associations with SA or CT phenotypes from the GCLUST parcellation, which is sensitive to genomic effects on SA and CT A-P and D-V gradients (*19*, *20*). This analysis allowed us to identify statistically significant multivariate relationships between gene co-expression modules and SA or CT and then allowed for examination of how the relationship manifests across brain regions and co-expression modules, and also how these relationships manifest in each group (see Methods for more details).

For SA, PLS identified one statistically significant latent variable (LV) pair (SA LV1: d = 3.99, p = 0.0001, split-half p_{ucorr} = 0.01, p_{vcorr} = 0.06), which explains 36% of the covariance between SA and gene expression. A highly similar result was obtained with a PLS on vertex-wise data (Fig. S1). However, a PLS model using vertex-wise data explained far less percentage of covariance than the GCLUST parcellated PLS model (Fig. S2). This indicates that the PLS model on GCLUST parcellated features is more sensitive for highlighting associations between gene expression and SA. To decompose how this multivariate relationship manifests across co-expression modules and groups, in Fig. 2A we show which co-expression modules have 'non-zero' relationships in each group. These 'non-zero modules' are the co-expression modules of highest importance, as they have 95% confidence intervals (CIs) estimated by bootstrapping that do not include a correlation of 0 and are thus indicative of stable co-expression modules highly contributing to the SA LV1 relationship. In contrast, co-expression modules that we dub as 'zero modules' are those whereby the 95% CIs include a correlation of 0 and thus do not reliably contribute to the overall SA LV1 relationship.

Non-zero modules for SA LV1 account for a good majority of all genes analyzed (68%) and were highly enriched for broadly expressed genes (enrichment odds ratio (OR) =3.48, p = 1.90e-71; Table S5). These two observations are compatible with predictions from the omnigenic theory of complex traits, in which variance in complex traits like ASD are exerted en-masse by a large majority of genes that impact the primary tissue of relevance and by genes that are broadly expressed across many organs and tissues, but which also have impact on the brain (*15*). These effects are also in line with similar results observed for large-scale gene expression associations with functional imaging phenotypes in ASD early language outcome subtypes (*14*).

Fig. 2A also shows that non-zero modules are highly similar for ASD Good and TD groups, whereas hardly any non-zero modules are present for ASD Poor. This similarity between ASD Good and TD can be quantified as a significant positive correlation in the PLS correlations values for these groups (Fig. 2C) ($r = 0.55$, $p = 0.008$). This result indicates that the SA LV1 relationship manifests similarly in TD and ASD Good groups. In contrast, there is a lack of correlation between ASD Poor and the other groups (ASD Poor-ASD Good: $r = 0.27$, $p = 0.23$; ASD Poor-TD: $r = -0.41$, $p = 0.06$) (Fig. 2C). Therefore, SA LV1 can be described as a large-scale SA-gene expression relationship that likely reflects a normative phenomenon present in TD and which is also preserved in the ASD Good subtype. However, this normative SA-gene expression relationship is absent in the ASD Poor subtype.

PLS analysis applied to GCLUST CT data isolated 2 statistically significant LV pairs (CT LV1: $d = 4.30$, $p = 0.0001$, split-half $p_{ucorr} = 0.04$, $p_{vcorr} = 0.01$; CT LV2: $d = 3.09$, $p = 0.0001$, split-half $p_{ucorr} = 0.02$, $p_{vcorr} = 0.05$), explaining 37% and 19% of the covariance between CT and gene expression respectively. PLS analysis on vertex-wise data produced similar results (Fig. S1), but again was not as good as the PLS on GCLUST parcellated features, as indicated by percentage of covariance explained (Fig. S2). Similar to SA LV1, non-zero modules for CT LV1 comprise a large majority of all genes examined (65%), are enriched for broadly expressed genes (OR= 2.96, p = 4.43e-43; Table S5), and thus is compatible with predictions about omnigenic effects exerted by broadly expressed genes. The relationships are also highly similar for ASD Good and TD, but not ASD Poor (Fig. 2D, F), which indicates that CT LV1 mostly pertains to a normative relationship preserved across TD and ASD Good, but which is absent in ASD Poor.

Atypical association between gene expression and cortical thickness, specific to ASD Poor language subtype

In contrast to CT LV1, the non-zero modules for CT LV2 are almost exclusively relevant for the ASD Poor subtype, comprise about 48% of all genes examined, do not show specific enrichment for broadly expressed genes (Table S5), and do not show strong correlations between groups (Fig. 3A-B). These results indicate that CT LV2 captures a relationship that is specific to ASD Poor. Furthermore, because PLS LVs are orthogonal to each other, CT LV2 is an independent relationship capturing effects that appear primarily in the ASD Poor subtype. While CT LV2's non-zero modules appear to be somewhat overlapping to CT LV1, the way these modules can affect CT are sometimes opposite to the directionality shown for CT LV1. For example, superior parietal cortex in CT LV1 has negative BSR values (Fig. 2E), while in CT LV2, the BSR values are strong and positive (Fig. 3C). This reversal in BSR values indicates different directionality of the gene expression-CT relationship. In other brain regions such as language-sensitive left hemisphere perisylvian, middle temporal, and inferior parietal cortex, CT LV1 shows brain bootstrap ratios (BSRs) that are close to 0 in CT LV1 (Fig. 2E) indicating little to no importance of these regions for LV1. However, the BSRs in CT LV2 for these regions are very strong (either blue or red colored BSRs in Fig. 3C), indicating that these regions are of strong importance for the CT LV2 relationship. These observations further indicate how CT LV2 captures a specific and independent type of genomic association with CT that is present primarily in ASD Poor.

Anterior-posterior and dorsal-ventral gradient patterning of surface area and cortical thickness are atypical in the ASD Poor language subtype

We next investigated how large-scale genomic variability patterns SA and CT cortical phenotypes. The patterning of PLS BSR values (Fig. 4A) can be used to answer this question. Brain BSRs can be interpreted as pseudo Z-statistics computed for each brain region and indicate the importance of each brain region for the LV in question and also indicating the directionality through which gene expression is associated with SA and CT. It is visually evident from Fig. 4A 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 (*19*, *20*) to be the genetically parcellated A-P and D-V axes of SA and CT (Fig. 4B), we confirm that BSRs highly differ along these A-P and D-V clusters (Fig. 4D). 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.

Perhaps even more striking than these differences between binary A-P and D-V partitions is that BSRs also covary along continuous A-P and D-V genetic similarity gradients. After ordering regions by genetic similarity gradients discovered by Chen and colleagues (*19*, *20*) (Fig. 4C) we find that BSRs are highly correlated with the ordering along this axis of genetic similarity between regions (Fig. 4E). This indicates that large-scale blood leukocyte gene co-expression relationships with SA and CT reveal how the cortex is genomically patterned to promote the development of cortical regionalization and areal identity (*18*). Because SA LV1 and CT LV1 are normative effects primarily relevant for TD and ASD Good, but not ASD Poor, these results indicate that normative genomic patterning of the cortex does not occur in the ASD Poor subtype. Conversely, CT in ASD Poor subtype may be patterned in a different way given that CT LV2 was primarily relevant to this subtype and given that the BSR patterning is reversed for CT LV2 compared to CT LV1 (Fig. 3E and G versus Fig. 3F and H). Given evidence of focal laminar patches throughout the cortex in ASD (*43*), it will be important for future work to investigate further how such phenomena may be relevant to atypical CT patterning, particularly in the ASD Poor subtype.

The use of the GCLUST parcellation does not appear to bias the emergence of these A-P, D-V, or genetic similarity gradients. In a vertex-wise PLS we find that the patterning of brain BSRs follow the same types of gradients for SA LV1, CT LV1, and CT LV2 (Fig. S3). Unlike the genomic patterning effects, we also found that patterning of the group differences in effect size for CT and SA do not follow similar A-P and D-V gradients (Fig. S4). This result suggests that these cortical patterning effects are not simply effects that can be seen as on-average group differences in SA or CT, or biases due to the parcellation scheme, and point more towards the specific importance of how the underlying genomic mechanisms act to pattern SA and CT across the cortex.

Genes involved in gradient patterning of surface area and cortical thickness follow similar gradients of gene expression during prenatal development

Because cortical regionalization begins in early prenatal periods from A-P and D-V gradient patterning of gene expression (*18*, *22*, *24*, *27*), we next assessed whether genes from SA and CT non-zero modules encompass many of the same genes that play important prenatal roles in the genomic gradient patterning of the cortex. Using the prenatal RNA-seq data from the Development PsychENCODE dataset, we used sparse PCA (*44*) to identify A-P (PC1) and D-V (PC2) gene expression gradients and the most important genes contributing to those gradients from

12 regions of prenatal cortical tissue sampled from 12-24 weeks post-conception (e.g., midgestation) (Fig. 5A-C). Remarkably, SA LV1 and CT LV1 non-zero modules are highly enriched for genes that comprise the prenatal A-P and D-V gradients (Fig. 5D). Enrichments were also seen for CT LV2, but unlike SA LV1 and CT LV1, the enrichments were apparent for both zero and non-zero modules (Fig. 5D). These results suggest that the genes responsible for the normative SA LV1 and CT LV1 relationships are also genes in prenatal periods that act to initialize the regionalization and patterning of cortex along A-P and D-V axes. Since SA LV1 and CT LV1 relationships are largely absent in the ASD Poor subtype, this result suggests that the atypical genomic patterning of SA and CT in this subtype could stem from perturbations in earlier prenatal development.

Genes expressed in prenatal progenitor cell types explain surface area associations, while genes expressed in later differentiated excitatory neurons explain cortical thickness associations

The evidence that SA and CT non-zero modules are enriched for genes that are important for midgestational A-P and D-V expression gradients leaves open the question of what prenatal cell types might explain such effects. The radial unit hypothesis (*22*) suggests that symmetric cell division in progenitor cell types (e.g., radial glia) in the ventricular zone leads to a substantial proliferation of radial units that then each become their own cortical columns and thus, leads to substantial expansion of SA. Variation in this proliferative process in different parts of the ventricular zone protomap regulates regional differences in SA (*18*, *22*, *45*). Therefore, progenitor cells are the primary cell types expected to explain SA effects. Programmed cell death could also be another mechanism regulating SA (*18*) and could implicate microglia involvement in SA. In contrast, CT is likely regulated by asymmetric cell division leading to more neurons within particular cortical columns (*22*) as well as intermediate progenitor cell types (*18*). CT is also heavily influenced by dendritic arborization (*39*). While arborization changes over development due to a variety of factors such as experience-dependent pruning, CT and the trajectory it follows over development is also known to be heavily influenced by genetic factors even in middle-aged adults, suggesting that individual differences in CT have a genetic and neurodevelopmental origin (*21*, *46*). These ideas would support the prediction that relatively later differentiated cell types (compared to progenitor cells), such as excitatory neurons, could explain CT effects.

Given that cell type markers from midgestational periods are available (*47*), we next asked if specific prenatal cell type markers are enriched for genes from SA and CT non-zero modules. In striking agreement with the prediction that progenitor cells explain SA effects (*18*, *22*), we find that SA LV1 non-zero modules show enrichments for all progenitor cells types - ventricular and outer radial glia (vRG, oRG), cycling progenitors in S and G2M phases of cell cycle (PgS, PgG2M), and intermediate progenitors (IP). Several non-neuronal cells also show SA LV1 enrichments, including oligodendrocyte precursors (OPC), endothelial cells (End), and microglia (Mic) (Fig. 5E; Table S6). In contrast to these cell type enrichments, there is little evidence of enrichment of genes specific to later differentiated excitatory (ExM, ExN, ExM-U, ExDp1, ExDp2) and inhibitory (InCGE, InMGE) neurons.

For CT LV1 and LV2 we identify enrichments for vRG and IP progenitor cell types that are compatible with hypothesized effects of IP cells on CT (*18*). However, CT LV1 and LV2 primarily show a different cell type enrichment profile from SA LV1, through the marked presence

of enrichments with several types of excitatory neurons (Fig. 5F-G; Table S6). This result indicates a striking contrast between the SA LV1 enrichment profile of primarily progenitor cell types and are compatible with the radial unit and protomap hypotheses (*22*), differential SA and CT GWAS enrichments (*21*), and other viewpoints regarding contributors to CT (*39*). These results also highlight effects of non-neuronal cell types such as microglia cells. Microglia enrichments are present and particularly strong for SA LV1 and CT LV1 non-zero modules. This effect may have implications for programmed cell death and pruning explanations (*48*) and which may be relevant to ideas behind ASD-relevant broadly expressed genes and their particularly strong effects on non-neuronal cell types such as microglia (*17*).

Non-zero modules are enriched for genes involved in vocal learning

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 (*14*) were highly enriched for differentially expressed genes in Area X from a songbird model of vocal learning (*49*). 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 (*49*). 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 = 1.90, p = 9.61e-4), but not zero modules (p >0.08) (Fig. 6A-C; Table S6). For CT LV2, enrichments were present at FDR q<0.05 (but not FDR q<0.01) for both non-zero (OR = 1.62, p = 0.006) and zero modules (OR = 1.61, p = 0.017). These effects suggest that many genes responsible for vocal learning in songbirds are conserved and highly represented specifically within SA and CT non-zero modules that are relevant for groups with relatively intact language (e.g., TD and ASD Good).

Genes within surface area non-zero modules are specifically enriched in human-specific genes

Language is a uniquely human ability and there is some evidence that genes implicated in human-specific evolution are also relevant for autism (*50–53*). In prior work we found that PLS non-zero modules associated with speech-related fMRI response (*14*) were enriched for differentially expressed genes in the cortex of humans versus non-human primates (i.e. 'human-specific' genes). Given that cortical SA is a phenotype that is dramatically expanded in human evolution, and much more so than CT, we investigated the hypothesis of whether SA non-zero modules would be specifically enriched for human-specific genes. Using 3 lists of human differentially expressed genes in prenatal, early postnatal, and adulthood periods (*52*), we find that SA LV1 non-zero modules are specifically enriched for prenatal and adulthood human-specific genes (prenatal OR = 1.86, p = 1.93e-3; adulthood OR = 1.97, p = 1.02e-5) (Fig. 6D; Table S6). In contrast, no such enrichments are found with genes relevant to CT LV1 or LV2 (Fig. 6E-F). In addition to differentially expressed genes we also examined genes that are targets of human-accelerated regions (HAR) or human-gained (HGE) or lossed enhancer (HLE) regions (*53*). However, no enrichments for SA or CT were identified for HAR, HGE and HLE genes (Fig. 6D-F). These results expand on the notion that human-specific genes are of relevance to ASD by

showing that the normative genomic mechanisms associated with SA are also genes of importance for human-specific evolution. Given that the SA LV1 relationship is absent in ASD Poor, this suggests that the loss of such normative associations may allow for early SA expansion and possibly early brain overgrowth for ASD Poor.

Non-zero modules are enriched for ASD-associated genes that affect prenatal development

Next, we asked whether SA and CT non-zero modules were relevant for known autism-associated genomic mechanisms. SA LV1 and CT LV1 non-zero or zero modules are not enriched for rare de novo protein truncating variants (*54*) or other genes that are annotated as autism-associated in SFARI Gene (*55*). However, CT LV2 non-zero modules were enriched for SFARI ASD genes (Table S6). Thus, at the level of ASD-risk gene mutations, CT LV2 was the only feature showing enrichments with non-zero modules. This could be compatible with the nature of CT LV2 being mostly specific to the ASD Poor subtype.

At the level of genes with evidence of ASD-dysregulated expression from post-mortem cortical tissue, we find that both CT LV1 and LV2 non-zero modules were enriched for ASD upregulated genes (*56*). In contrast, genes from cortically downregulated co-expression modules (*33*) were highly enriched with genes from SA LV1 non-zero modules (Table S6). This result shows an interesting contrast between CT and genes that show upregulated expression versus SA and genes that show downregulated expression in ASD.

Non-zero modules from SA LV1, CT LV1, and CT LV2 are also enriched for co-expression modules that are highly transcriptionally active during prenatal periods and which contain many high-penetrance ASD-related mutations (Fig. 7A-C; Table S6). This is compatible with the idea that broadly expressed genes can interact and impact key ASD-risk genes, particularly in prenatal periods (*16*, *17*). Downstream targets of highly penetrant genes like *FMR1* and *CHD8* were also enriched in non-zero modules from SA LV1, CT LV1, and CT LV2. However, not all of these enrichments are specific to autism-associated genes. Genes differentially expressed in schizophrenia (*56*) were also significantly enriched in non-zero modules across SA LV1, CT LV1, and CT LV2.

Finally, we examined enrichments with cell type specific differentially expressed genes in autism (*57*). Here we found that only SA LV1 non-zero modules are enriched for differentially expressed genes in microglia cells (Fig. 7D). No other comparisons for DE cell types were statistically significant. See Fig. 7 and Table S6 for a summary of autism-associated enrichments. The fact that non-zero modules are devoid of enrichments in most DE genes from specific cell types is compatible with the notion that these genes are of primary relevance for early prenatal periods and will not be a highly discoverable DE signal in post-mortem ASD tissue.

To aid future work examining specific genes of interest, we focused on identifying high-confidence ASD-risk genes (annotated as the 'high-confidence' category 1 list in SFARI Gene) that are also SA-and prenatally-relevant progenitor and A-P patterning genes (i.e. the intersection of SFARI ASD, SA non-zero modules, PC1 A-P genes, prenatal progenitor cell types, and ASD prenatal co-expression modules). *SON* and *BAZ2B* were identified and these genes play roles in splicing, cell cycle, transcriptional regulation, and chromatin remodeling. For CT LV1 genes, we

next searched for high-confidence ASD-risk genes that were also prenatally-relevant excitatory and D-V patterning genes (i.e. the intersection of SFARI ASD, CT LV1 non-zero modules, PC2 D-V genes, prenatal excitatory cell types, and ASD prenatal co-expression modules). Here we find ASD 'high-confidence' genes of *ATRX*, *AUTS2*, and *BCL11A*. In a similar search within CT LV2 non-zero modules of prenatal relevance to excitatory neurons and D-V patterning, we identified *ATRX*, *AUTS2*, *BCL11A*, *CACNA1E*, and *MEIS2* as high-confidence ASD-risk genes. A common theme of all these CT-relevant genes is their role in chromatin modification and remodeling (with the exception of *CACNA1E*) and their links to syndromes causing intellectual disability. Additionally, with the exceptions of *BCL11A* and *CACNA1E*, all SA- and CT-relevant high-confidence genes listed here fall into the broadly expressed gene list, highlighting the importance of these high-impact genes in ASD biology (*16*).

Discussion

These findings represent a significant enhancement to the mechanistic and clinical precision of our understanding of the prenatal brain basis behind different types of ASD. The evidence here solidifies the idea that the ASD Poor subtype is biologically distinct (*8*, *14*) by revealing how large-scale functional prenatal genomic signal is differentially associated with structural cortical phenotypes such as CT and SA.

Going beyond the idea of whether these subtypes are biologically distinct, we also need an explanation for how diverse genetic mechanisms affecting ASD individuals may converge onto a common atypical downstream biological process and to understand when this process manifests as different during development. The current work gives the first insights into how to answer this question for the critical ASD Poor subtype. The prenatal genomic patterning of the cerebral cortex is the key explanation behind how diverse genetic mechanisms impacts brain development for the ASD Poor subtype. Normative genomic patterning gradients emerge in the first and second trimesters of prenatal development along A-P and D-V axes and allow cortical areas to develop distinct cellular, functional and circuit-level identities (*18*, *22–27*). Prenatal cortical arealization processes are critical for circuit-formation, maturation of distinct types of neurophysiological response, and the development of regional and network-level functional specialization that occurs with postnatal experience (*18*, *58*). These processes are explained by genetic variation and manifest after birth as patterned effects in CT and SA phenotypes measured with structural MRI data (*19*, *20*). This normative prenatal genomic cortical patterning effect is absent in the ASD Poor subtype. In addition to the lack of normative genomic cortical patterning of SA and CT, CT in ASD Poor is patterned in a unique manner (e.g., CT LV2) and potentially indicative of a separate route through which genomic pathology affects CT and penetrates up through to the poor language clinical phenotype.

There are some caveats and limitations that are necessary to address in order to interpret the present findings. First, the sample size of this study is moderate to above average for what is typical in most toddlerhood brain imaging and gene expression studies (*59*, *60*) and this is the first to relate MRI phenotypes such as SA and CT in toddlerhood to large-scale gene expression activity. Thus, future work replicating these findings with larger samples is needed. Within the context of gene expression studies of brain tissue, sample sizes are typically much smaller than the

current work and deal with RNA quality that is much lower than what is typical in studies using blood samples. Additionally, postmortem brain tissue studies typically have much larger age ranges spanning toddlerhood to adulthood, suggesting that there is much larger age-related heterogeneity in studies of brain tissue compared to blood. Thus, a relative strength of the current work compared to postmortem brain tissue studies is the restricted age range to the early toddler years, which helps enhance sensitivity for very early developmental effects. Combining these two caveats with the rare ability to make stratifications, the current study is ahead of the norms typical for the context of gene expression studies in ASD patients. Second, it is notable that we did not compare ASD Poor to a non-ASD comparison group with language and/or developmental delay (LD/DD). In prior work, we have shown how the developmental clinical trajectories for a non-ASD LD/DD group are in fact different from the ASD Poor group, suggesting that ASD Poor is developmentally and behaviorally distinguished from LD/DD (*8*). We also showed in prior work that speech-related fMRI response in ASD Poor was distinctly different from a non-ASD LD/DD comparison group (*8*), which again supports the idea that ASD Poor is not simply just a reflection of LD or DD. In the current study we did not possess enough concurrent blood samples and MRI data from enough LD/DD subjects for a sufficient comparison group. Future work should attempt to collect this data as a further comparison to ASD Poor to better understand if the atypical genomic cortical patterning effects are indeed specific to ASD Poor. Third, it is important to clarify that while there is some utility in using blood gene expression to relate to neurodevelopmental mechanisms in autism, there are limitations in how far it can go in highlighting mechanisms that can only be identified in brain tissue. For example, brain-specific genes cannot be adequately assessed in blood and thus, the findings here do not represent the contributions of such important genomic mechanisms and how they might affect neural phenotypes like cortical patterning. Furthermore, using blood will not be able to capture tissue-specific effects regarding different isoforms, splicing, and/or epigenetic mechanisms. Despite these limitations, complex traits are theorized to be largely underpinned by omnigenic effects and include genes that are broadly expressed across several tissues other than the tissue of relevance (*15*). Applied to autism, it is known that the genomic landscape of autism includes many genes that are broadly expressed across many tissues and have strong regulatory impact (*16*). Given the inaccessibility of brain tissue in living patients, blood may be a key in-vivo window into how some of these types of broadly expressed and regulatory genomic mechanisms affect complex cortical phenotypes in an omnigenic fashion (*14*).

In conclusion, in the face of large heterogeneity in the ASD population, the current work indicates that individuals with poor versus good early language outcome are explained by distinct genomic mechanisms that cascade to shape cortical phenotypes and later clinical outcomes. A common downstream impact of the diverse genomic mechanisms discovered in this work is the emergent effect of atypical genomic patterning of the cerebral cortex in the ASD Poor subtype. This atypical genomic cortical patterning effect points to early prenatal periods and the importance of an omnigenic signal driven by broadly expressed genes. The functional consequences of atypical genomic patterning of the cortex may be the curtailed development of molecular cortical arealization processes that prohibit canonical circuit-formation and later regional functional specialization that is likely necessary for facilitating better outcomes in such individuals.

Materials and Methods

Participants

This study was approved by the Institutional Review Board at University of California, San Diego. Parents provided written informed consent according to the Declaration of Helsinki and were paid for their participation. Identical to the approach used in our earlier studies (*8*, *14*, *36*, *61–66*) toddlers were recruited through two mechanisms: community referrals (e.g., website) or a general population-based screening method called *Get SET Early* (*67*) that allowed for the prospective study of ASD beginning at 12 months based on a toddler's failure of the CSBS-DP Infant-Toddler Checklist (*68*, *69*). All toddlers were tracked from an intake assessment around 12 months and followed roughly every 12 months until 3–4 years of age. All toddlers, including normal control subjects, participated in a series of tests collected longitudinally across all visits, including the Autism Diagnostic Observation Schedule (ADOS; Module T, 1, or 2) (*70*), the Mullen Scales of Early Learning (*71*), and the Vineland Adaptive Behavior Scales (*72*). All testing occurred at the University of California, San Diego Autism Center of Excellence (ACE).

Stratification of ASD Poor versus ASD Good was made on the basis of Mullen EL and RL T-scores. An ASD toddler was classified as ASD Poor if both Mullen EL and RL T-scores at the final outcome assessment was below 1 standard deviation of the T-score norm of 50 (i.e. $T<40$). ASD Good labels were made if the toddler had either Mullen EL or RL T-scores within 1 standard deviation or above the normative T-score of 50 (i.e. $T \geq 40$). A total of n=123 toddlers had T1 structural MRI and gene expression data available. From these 123 toddlers, n=76 ASD individuals were examined and were split into the 2 language outcome subtypes - ASD Poor n=38 (32 male, 6 female; mean age at MRI scan = 29.01 months, SD at MRI scan = 7.22, range = 12-50 months), ASD Good n=38 (28 male, 10 female; mean age at MRI scan = 29.02 months, SD at MRI scan = 9.55, range = 14-46 months) and TD n=47 (25 male, 22 female; mean age at MRI scan = 25.91 months, SD at MRI scan = 10.44, range = 13-46 months). ASD subtypes and TD did not statistically differ in age at the time of scanning ($F(2,120) = 1.62$, $p = 0.20$). For more demographic and phenotypic information, please see Table S1.

Blood sample collection, leukocyte capture, RNA extraction, quality control, and samples preparation

Four to six milliliters of blood was collected into EDTA-coated tubes from toddlers on visits when they had no fever, cold, flu, infections or other illnesses, or use of medications for illnesses 72 hours prior blood draw. Blood samples were passed over a LeukoLOCK™ filter (Ambion, Austin, TX, USA) to capture and stabilize leukocytes and immediately placed in a -20°C freezer. Given the role of the immune system in autism (*73*, *74*) as well as interactions between the brain and the immune system (*75*), immune cells in blood like leukocytes were specifically examined. This choice also allows for constraint on the cell types for which RNA might arise from in blood, since whole blood is a bulk sample and RNA could potentially come from many different cell types (e.g., leukocytes, platelets). Total RNA was extracted following standard procedures and manufacturer's instructions (Ambion, Austin, TX, USA). LeukoLOCK disks (Ambion Cat #1933) were freed from RNA-later and Tri-reagent (Ambion Cat #9738) was used to flush out the captured lymphocyte and lyse the cells. RNA was subsequently precipitated with ethanol and purified

though washing and cartridge-based steps. The quality of mRNA samples was quantified by the RNA Integrity Number (RIN), values of 7.0 or greater were considered acceptable (*76*), and all processed RNA samples passed RIN quality control. Quantification of RNA was performed using Nanodrop (Thermo Scientific, Wilmington, DE, USA). Samples were prepped in 96-well plates at the concentration of 25 ng/µl.

Gene expression and data processing

RNA was assayed at Scripps Genomic Medicine (La Jolla, CA, USA) for labeling, hybridization, and scanning using the Illumina BeadChips pipeline (Illumina, San Diego, CA, USA) per the manufacturer's instruction. All arrays were scanned with the Illumina BeadArray Reader and read into Illumina GenomeStudio software (version 1.1.1). Raw data was exported from Illumina GenomeStudio, and data pre-processing was performed using the lumi package (*77*) for R (http://www.R-project.org) and Bioconductor (https://www.bioconductor.org) (*78*). Raw and normalized data are part of larger sets deposited in the Gene Expression Omnibus database (GSE42133; GSE111175).

A larger primary dataset of blood leukocyte gene expression was available from 383 samples from 314 toddlers with the age range of 1-to-4 years old. The samples were assayed using the Illumina microarray platform on three batches. The datasets were combined by matching the Illumina Probe ID and probe nucleotide sequences. The final set included a total of 20,194 gene probes. Quality control analysis was performed to identify and remove 23 outlier samples from the dataset. Samples were marked as outlier if they showed low signal intensity (average signal two standard deviations lower than the overall mean), deviant pairwise correlations, deviant cumulative distributions, deviant multi-dimensional scaling plots, or poor hierarchical clustering, as described elsewhere (*62*). The high-quality dataset included 360 samples from 299 toddlers. High reproducibility was observed across technical replicates (mean Spearman correlation of 0.97 and median of 0.98). Thus, we randomly removed one of each of two technical replicates from the primary dataset. From the subjects in the larger primary dataset, n=123 also had MRI data and thus a total of n=105 from the Illumina HT12 platform along with n=18 from the Illumina WG6 platform were used in this study. Batch was not asymmetrically distributed across one subgroup more than another, as chi-square analyses on the contingency table between subgroup and batch show no effect ($\chi^2(4) = 0.84$, $p = 0.93$). ASD subtypes and TD toddlers also did not statistically differ in age at the time of blood sampling ($F(2,120) = 1.27$, $p = 0.28$). The 20,194 probes were then collapsed to 14,426 genes based on picking the probe with maximal mean expression across samples. Data were quantile normalized and then adjusted for batch effects, sex, and RIN. This batch, sex, and RIN adjusted data were utilized in all further downstream analyses. We also checked for differences in proportion estimates of different leukocyte cell types (i.e. neutrophils, B cells, T cells, NK cells, and monocytes) using the CellCODE deconvolution method (*79*), but found no evidence of differences across groups for any cell type (see Table S7).

Weighted Gene Co-Expression Network Analysis

We reduced the number of features in the gene expression dataset from 14,426 genes down to 21 modules of tightly co-expressed genes. This data reduction step was achieved using weighted gene co-expression network analysis (WGCNA), implemented within the WGCNA library in R

(*80*). Correlation matrices estimated with the robust correlation measure of biweight midcorrelation were computed and then converted into adjacency matrices that retain the sign of the correlation. These adjacency matrices were then raised to a soft power of 16 (Fig. S5). This soft power was chosen by finding the first soft power where a measure of R^2 scale-free topology model fit saturates. The soft power thresholded adjacency matrix was then converted into a topological overlap matrix (TOM) and then a TOM dissimilarity matrix (e.g., 1-TOM). The TOM dissimilarity matrix was then input into agglomerative hierarchical clustering using the average linkage method. Gene modules were defined from the resulting clustering tree, and branches were cut using a hybrid dynamic tree cutting algorithm (deepSplit parameter = 4) (Fig. S5). Modules were merged at a cut height of 0.2, and the minimum module size was set to 100. Only genes with a module membership of $r > 0.2$ were retained within modules. For each gene module, a summary measure called the module eigengene (ME) was computed as the first principal component of the scaled (standardized) module expression profiles. We also computed module membership for each gene and module. Module membership indicates the correlation between each gene and the module eigengene (see Table S4). Genes that could not be clustered into any specific module are left within the M0 module, and this module was not considered in any further analyses. Further WGCNA analyses were run separately within each group in order to check for preservation of detected modules across groups at a soft power threshold of 16. These analyses all indicated high levels of preservation (Zsummary>10) (*81*) for all detected modules for each pairwise group comparison (Fig. S6).

MRI Data Acquisition and Analyses

Imaging data were collected on a 1.5 Tesla General Electric MRI scanner during natural sleep at night; no sedation was used. Structural MRI data was collected with a T1-weighted IR-FSPGR sagittal protocol (TE = 2.8 ms, TR = 6.5 ms, flip angle = 12 degrees, bandwidth = 31.25 kHz, FOV = 24 cm, slice thickness = 1.2 mm). Cortical surface reconstruction was performed using FreeSurfer v5.3 (http://surfer.nmr.mgh.harvard.edu/) (*82–84*), which uses routinely acquired T_1-weighted MRI volumes (*85*), includes tools for estimation of brain morphometry measures such as cortical thickness and surface area (*86*, *87*), and enables inter-subject alignment via nonlinear, surface-based registration to an average brain, driven by cortical folding patterns (*88*). FreeSurfer has been validated for use in children (*89*) and used successfully in large pediatric studies (*90*, *91*). Total cortical volume, surface area (SA) and mean cortical thickness (CT) were computed based on the Desikan-Killiany parcellation. Regional SA and CT values were computed from a 12-region parcellation reported by Chen and colleagues (*19*, *20*) based on genetic similarity in monozygotic twins. This parcellation scheme, known as GCLUST, is highly relevant for our purposes here, since the parcellations are based on genetic patterning and has also been effectively utilized in developmental samples (*46*). Thus, GCLUST should help increase statistical power while also minimizing multiple comparisons. The GCLUST parcellation is also important as it can be used to leverage information about genetic similarity gradients (e.g., rank ordering of regions by fuzzy clustering) in further analyses. The 2-cluster anterior-posterior (A-P) or dorsal-ventral (D-V) partitions discovered by Chen and colleagues (*19*, *20*) are also relevant in further analyses for A-P and D-V gradient questions. For all 12 regions of the SA and CT GCLUST parcellation, global effects were controlled for by dividing SA values by the mean SA, and for CT we subtracted mean CT from each region, as was done in prior papers using this parcellation scheme (*19*, *20*).

MRI-Gene Expression Association Analysis

To assess multivariate MRI-gene expression relationships we used partial least squares (PLS) analysis (*92*, *93*). PLS is widely used in the neuroimaging literature, particularly when explaining multivariate neural responses in terms of multivariate behavioral patterns of variation or a design matrix. Given that the current dataset is massively multivariate both in terms of MRI and gene expression datasets, we used PLS to elucidate how variation in SA or CT covaries with gene expression as measured by module eigengene values of co-expression modules. PLS allows for identifying such relationships by finding latent MRI-gene expression variable pairs (LV) that maximally explain covariation in the dataset and which are uncorrelated with other MRI-gene expression LV pairs. The strength of such covariation is denoted by the singular value (d) for each brain-gene expression LV, and hypothesis tests are made via using permutation tests on the singular values. Furthermore, identifying brain regions that most strongly contribute to each LV pair is achieved via bootstrapping, whereby a brain bootstrap ratio (BSR) is created for each region, and represents the reliability of that region for contributing strongly to the LV pattern identified. The brain BSR is roughly equivalent to a Z-statistic and can be used to threshold data to find voxels that reliably contribute to an LV pair.

The PLS analyses reported here were implemented within the plsgui MATLAB toolbox (www.rotman-baycrest.on.ca/pls/). Here we ran 2 separate PLS analyses - one on SA and another on CT. Neuroimaging data entered into the PLS analyses come from the 12 region GCLUST parcellations for SA and CT. Because the TD group differed in the proportion and males versus females compared to the ASD groups, we used a linear model to remove the effect of sex from the SA and CT data. This SA and CT data with the sex effect removed was input into the PLS analysis. For gene expression data, we input module eigengene values for all 21 co-expression modules. For statistical inference on identified MRI-gene expression LV pairs, a permutation test was run with 10,000 permutations. To identify reliably contributing regions for MRI-gene expression LVs and to compute 95% confidence intervals (CIs) on MRI-gene expression correlations, bootstrapping was used with 10,000 resamples. Gene co-expression modules whereby 95% CIs do not encompass 0 are denoted as 'non-zero' association modules. All other modules where 95% CIs include 0 are denoted as 'zero' modules. Additionally, we ran 10,000 split-half resamples whereby the correlation between brain or gene expression saliences (Ucorr, Dcorr) were computed between the two split-halves. These correlations between split-half saliences were then compared to the null distribution from 10,000 permutations to compute p-values (p_{ucorr}, p_{dcorr}), which statistically test the reliability of salience patterns in split-half resamples (*94*).

From the PLS results we tested whether groups show similar correlation patterns across modules. To test this question, we computed Pearson correlations on the PLS correlation values for all pairwise group comparisons. Groups with similar PLS correlations will show statistically significant correlations. We also used the brain bootstrap ratios (BSR) from the PLS analysis to identify whether BSRs covary along the genetic similarity gradients and A-P and D-V partitions discovered by Chen and colleagues (*19*, *20*). Pearson correlations were used to identify correlations with genetic similarity gradients, while independent-samples t-tests were used to compare A-P and D-V partitions.

All PLS analyses were computed on vertex-wise data as well as GCLUST parcellated data. This analysis used the same parameters (10,000 permutations, 10,000 bootstrap resamples) as the GCLUST analysis and was computed on the sex and mean SA or CT adjusted vertex-wise data. A-P and D-V distinctions were assessed on vertex-wise brain BSR data with violin-boxplots (Fig. S3) separated by A-P or D-V partitions. Genetic similarity gradient organization was assessed by first computing the median BSR for each GCLUST parcel and then computing the correlation between median BSR and the genetic similarity gradient rank ordering from GCLUST. Similarity in gene co-expression module PLS correlations between GCLUST and this vertex-wise analysis were also computed as Pearson correlations from the PLS correlation values for each module and group. Finally, to assess which model (GCLUST or vertex-wise) was a better model, we assessed which model had the highest percentage covariance explained. Comparison between GCLUST and vertex-wise PLS results can be seen in Fig. S1-S3.

Gene Set Enrichment Analyses

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

Prenatal Gene Expression Gradients and Cell Types

To assess gradients in prenatal gene expression we utilized RNA-seq data from the Development PsychENCODE dataset (http://development.psychencode.org) (*24*). The data utilized was already preprocessed as described by Li and colleagues (*24*) (e.g., normalized, batch effects removed) and summarized to RPKM. Sample data from all 12 available cortical regions from 12-22 weeks post-conception were utilized in order to capture the midgestational window of interest. Before running the analysis we removed low expressing genes with log2(RPKM) below 2. The primary analysis to identify expression gradients was an adjustment-for-confounds principal components analysis (AC-PCA) (*44*) which allowed for adjustment due to repeat measurements from the same donor across sampled brain regions. Rank ordering of regions by A-P and D-V axes were utilized to statistically confirm that PC1 and PC2 components follow A-P and D-V gradients. Subsets of the most important genes for the top two principal components were identified with a sparse AC-PCA analysis, whereby the sparsity parameter, c_2, was selected based on a grid search with 10-fold cross validation. These PC1 and PC2 gene sets were used in enrichment tests with PLS non-zero or zero modules.

We also examined enrichments between PLS non-zero and zero modules and prenatal cell types identified from single cell RNA-seq on midgestational prenatal brain tissue (*47*). These cell types included several classes of progenitor cells (ventricular radial glia, vRG; outer radial glia, oRG; cycling progenitors (S phase), PgS; cycling progenitors (G2/M phase), PgG2M; intermediate progenitors, IP), excitatory neurons (migrating excitatory, ExN; maturing excitatory, ExM;

maturing excitatory upper enriched, ExM-U; excitatory deep layer 1, ExDp1; excitatory deep layer 2, ExDp2), inhibitory neurons (interneuron CGE, InCGE; interneuron MGE, InMGE), and other non-neuronal cell types (oligodendrocyte precursors, OPC; pericytes, Per; endothelial cells, End; microglia, Mic).

Tissue-specific enrichments

To better understand how genes expressed in blood leukocytes could be brain-relevant we annotated gene co-expression modules based on enrichments in genes known from expression across multiple tissues to be either broadly expressed or brain-specific. Both of these categories contain genes that are expressed in cortical tissue, but differ in the pattern of expression across other non-neuronal tissues. To define these lists we downloaded transcript per million (TPM) normalized gene expression from 10,259 samples across 26 tissues from the GTEx dataset (https://www.gtexportal.org) (*95*, *96*). In addition to brain and nerve tissue, the dataset included transcriptome data from 24 non-neuronal tissues, including: Adipose, Adrenal Gland, Blood Vessel, Breast, Blood, Skin, Colon, Esophagus, Heart, Liver, Lung, Salivary Gland, Muscle, Ovary, Pancreas, Pituitary, Prostate, Small Intestine, Spleen, Stomach, Testis, Thyroid, Uterus, and Vagina. We next defined a gene expressed in a tissue if it met two criteria. First, the gene TPM expression level was ≥ 3 in at least half of the samples from the tissue. Second, the median expression of the gene was equal or larger than its 25-percentile expression in GTEx cortex samples. The second criterion was included to account for the differences in the base expression level of the genes and their dosage dependent translation and function. Broadly-expressed genes were defined as genes that were expressed in $\geq 50\%$ of non-neuronal tissues (i.e., tissues other than brain and nerve). The broadly-expressed and brain-specific genes included genes that were expressed in the adult cortex based on GTEx dataset.

Vocal learning enrichments

To test for enrichment between PLS non-zero modules and gene sets of functional relevance for language processes, we examined genes that are differentially expressed in a song bird vocal learning model. Song birds are often used as animal models relevant for the vocal learning component of language (*49*, *97*, *98*). We investigated enrichments with differentially expressed genes taken from a microarray dataset of Area X of song birds (*49*). To identify differentially expressed (DE) genes between singing versus non-singing birds, we re-analyzed this dataset (GEO Accession ID: GSE34819) using limma (*99*), and DE genes were identified if they passed Storey FDR $q<0.05$ (*100*). These DE genes were also used for enrichment tests in our prior work examining gene expression relationships with language-relevant functional neural phenotypes measured with fMRI (*14*).

Human-specific enrichments

Given the uniquely human nature of language, we also tested hypotheses regarding enrichments with genes that are transcriptionally different in the cortical tissue between humans and other non-human primates across prenatal, early postnatal and adult periods (*52*). In addition, we also examined enrichments with genes linked to human accelerated regions (HAR), human-gained enhancers (HGE) in prenatal and adult tissue, and human-lossed enhancers (HLE) (*53*).

Autism-associated enrichments

Ample evidence suggests that prenatal periods are critical for ASD (*6*, *16*, *101–103*). To test enrichment with prenatal ASD-associated co-expression modules, we utilized co-expression modules from a study that analyzed the Allen Institute BrainSpan dataset (*104*). Parikshak and colleagues analyzed only cortical regions from BrainSpan and identified M2 and M3 as prenatally active and enriched for rare protein truncating variants with high penetrance for ASD (*101*). We also tested enrichments with gene lists known to be associated with ASD, either from genetic evidence or evidence from cortical transcriptomic dysregulation. In particular, we examined a list of 102 rare de novo protein-truncating variants (dnPTV) associated with ASD (*54*), genes listed as ASD-associated in SFARI Gene (https://gene.sfari.org) in categories S, 1, 2, and 3 (downloaded on July 16, 2020) (*55*), and DE genes and cortical co-expression modules measured from ASD post-mortem frontal and temporal cortex tissue (*33*, *56*). To contrast ASD DE genes to genes that are DE in other psychiatric diagnoses that are genetically correlated with autism, we also use DE genes in schizophrenia (SCZ DE) and bipolar disorder (BD DE) from the same study that identified ASD DE genes (*56*). To go beyond DE genes identified in bulk tissue samples, we also examined ASD DE genes identified in specific cell types - particularly, excitatory (ASD Excitatory) and inhibitory (ASD Inhibitory) neurons, microglia (ASD Microglia), astrocytes (ASD Astrocyte), oligodendrocytes (Oligodendrocyte), and endothelial (ASD Endothelial) cells (*57*). Finally, we also tested for enrichments with known downstream targets of highly penetrant mutations known to be associated with ASD – FMRP and CHD8. For each, we had lists of downstream targets for two independent studies (*105–108*), where the overlap for FMRP targets was 3.71% and 27.61% for CHD8 targets.

Data and code availability

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

Acknowledgments

We thank all participants and their families for participating in this study. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No 755816 (ERC Starting Grant to MVL). This work was also supported by the following grants to EC, KP, LE, and NEL - NIMH R01-MH080134 (KP), NIMH R01-MH104446 (KP), NFAR grant (KP), NIMH Autism Center of Excellence grant P50-MH081755 (EC, KP), NIMH R01-MH036840 (EC), NIMH R01-MH110558 (EC, NEL), NIMH U01-MH108898 (EC), NIDCD R01-DC016385 (EC, KP, LE, MVL), CDMRP AR130409 (EC), and the Simons Foundation 176540 (EC). KC was supported by the Utah Stimulating Access to Research in Residency Transition Scholar (StARRTS) under Award Number 1R38HL143605-01.

Author contributions

Conceptualization: MVL, EC, KP, LE TP. Methodology: MVL, TP, VHG, C-HC, NEL, EC, KP, LE. Software: MVL, DJH, AMD, VHG. Formal analysis: MVL, VHG, DJH, TP, JS, RAIB, NB. Investigation: LE, KC, CCB, LL, KP, EC. Data curation: EC, KP, LE, TP, VHG, MVL. Writing - original draft preparation: MVL, EC. Writing - review and editing: MVL, EC, KP, LE, TP, JS RAIB, NB, KC, NEL, VHG, DJH, C-HC, AMD. Visualization: MVL. Supervision: EC, KP, LE, NEL. Project administration: EC, KP, LE, NEL. Funding acquisition: MVL, EC, KP, LE, NEL.

Declaration of interests

None of the authors have any biomedical financial interests or potential conflicts of interest to report.

References

1. 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).

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

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

4. 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).

5. F. S. Collins, H. Varmus, A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).

6. 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).

7. A. Jack, K. Pelphrey, Annual Research Review: Understudied populations within the autism spectrum - current trends and future directions in neuroimaging research. *J Child Psychol Psychiatry*. **58**, 411–435 (2017).

8. 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).

9. H. Tager-Flusberg, C. Kasari, Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. *Autism Res*. **6**, 468–478 (2013).

10. R. J. Landa, Efficacy of early interventions for infants and young children with, and at risk for, autism spectrum disorders. *Int Rev Psychiatry*. **30**, 25–39 (2018).

11. M. V. Lombardo, E. M. Busuoli, L. Schreibman, A. C. Stahmer, T. Pramparo, I. Landi, V. Mandelli, N. Bertelsen, C. C. Barnes, V. Gazestani, L. Lopez, E. C. Bacon, E. Courchesne, K. Pierce, "Pre-treatment clinical behavioral and blood leukocyte gene expression patterns predict rate of change in response to early intervention in autism" (preprint, Psychiatry and Clinical Psychology, 2020), , doi:10.1101/2020.12.21.20248674.

12. D. Zachor, E. Ben-Itzchak, Variables Affecting Outcome of Early Intervention in Autism Spectrum Disorder. *J Pediatr Neurol*. **15**, 129–133 (2017).

13. L. Zwaigenbaum, M. L. Bauman, R. Choueiri, C. Kasari, A. Carter, D. Granpeesheh, Z. Mailloux, S. Smith Roley, S. Wagner, D. Fein, K. Pierce, T. Buie, P. A. Davis, C. Newschaffer, D. Robins, A. Wetherby, W. L. Stone, N. Yirmiya, A. Estes, R. L. Hansen,

879 J. C. McPartland, M. R. Natowicz, Early Intervention for Children With Autism Spectrum
880 Disorder Under 3 Years of Age: Recommendations for Practice and Research. *Pediatrics*.
881 **136 Suppl 1**, S60-81 (2015).

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

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

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

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

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

895 19. C.-H. Chen, E. D. Gutierrez, W. Thompson, M. S. Panizzon, T. L. Jernigan, L. T. Eyler, C.
896 Fennema-Notestine, A. J. Jak, M. C. Neale, C. E. Franz, M. J. Lyons, M. D. Grant, B.
897 Fischl, L. J. Seidman, M. T. Tsuang, W. S. Kremen, A. M. Dale, Hierarchical genetic
898 organization of human cortical surface area. *Science*. **335**, 1634–1636 (2012).

899 20. C.-H. Chen, M. Fiecas, E. D. Gutiérrez, M. S. Panizzon, L. T. Eyler, E. Vuoksimaa, W. K.
900 Thompson, C. Fennema-Notestine, D. J. Hagler, T. L. Jernigan, M. C. Neale, C. E. Franz,
901 M. J. Lyons, B. Fischl, M. T. Tsuang, A. M. Dale, W. S. Kremen, Genetic topography of
902 brain morphology. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17089–17094 (2013).

903 21. K. L. Grasby, N. Jahanshad, J. N. Painter, L. Colodro-Conde, J. Bralten, D. P. Hibar, P. A.
904 Lind, F. Pizzagalli, C. R. K. Ching, M. A. B. McMahon, N. Shatokhina, L. C. P. Zsembik,
905 S. I. Thomopoulos, A. H. Zhu, L. T. Strike, I. Agartz, S. Alhusaini, M. A. A. Almeida, D.
906 Alnæs, I. K. Amlien, M. Andersson, T. Ard, N. J. Armstrong, A. Ashley-Koch, J. R.
907 Atkins, M. Bernard, R. M. Brouwer, E. E. L. Buimer, R. Bülow, C. Bürger, D. M.
908 Cannon, M. Chakravarty, Q. Chen, J. W. Cheung, B. Couvy-Duchesne, A. M. Dale, S.
909 Dalvie, T. K. de Araujo, G. I. de Zubicaray, S. M. C. de Zwarte, A. den Braber, N. T.
910 Doan, K. Dohm, S. Ehrlich, H.-R. Engelbrecht, S. Erk, C. C. Fan, I. O. Fedko, S. F. Foley,
911 J. M. Ford, M. Fukunaga, M. E. Garrett, T. Ge, S. Giddaluru, A. L. Goldman, M. J. Green,
912 N. A. Groenewold, D. Grotegerd, T. P. Gurholt, B. A. Gutman, N. K. Hansell, M. A.
913 Harris, M. B. Harrison, C. C. Haswell, M. Hauser, S. Herms, D. J. Heslenfeld, N. F. Ho,
914 D. Hoehn, P. Hoffmann, L. Holleran, M. Hoogman, J.-J. Hottenga, M. Ikeda, D. Janowitz,
915 I. E. Jansen, T. Jia, C. Jockwitz, R. Kanai, S. Karama, D. Kasperaviciute, T. Kaufmann, S.
916 Kelly, M. Kikuchi, M. Klein, M. Knapp, A. R. Knodt, B. Krämer, M. Lam, T. M.
917 Lancaster, P. H. Lee, T. A. Lett, L. B. Lewis, I. Lopes-Cendes, M. Luciano, F. Macciardi,
918 A. F. Marquand, S. R. Mathias, T. R. Melzer, Y. Milaneschi, N. Mirza-Schreiber, J. C. V.

Moreira, T. W. Mühleisen, B. Müller-Myhsok, P. Najt, S. Nakahara, K. Nho, L. M. Olde Loohuis, D. P. Orfanos, J. F. Pearson, T. L. Pitcher, B. Pütz, Y. Quidé, A. Ragothaman, F. M. Rashid, W. R. Reay, R. Redlich, C. S. Reinbold, J. Repple, G. Richard, B. C. Riedel, S. L. Risacher, C. S. Rocha, N. R. Mota, L. Salminen, A. Saremi, A. J. Saykin, F. Schlag, L. Schmaal, P. R. Schofield, R. Secolin, C. Y. Shapland, L. Shen, J. Shin, E. Shumskaya, I. E. Sønderby, E. Sprooten, K. E. Tansey, A. Teumer, A. Thalamuthu, D. Tordesillas-Gutiérrez, J. A. Turner, A. Uhlmann, C. L. Vallerga, D. van der Meer, M. M. J. van Donkelaar, L. van Eijk, T. G. M. van Erp, N. E. M. van Haren, D. van Rooij, M.-J. van Tol, J. H. Veldink, E. Verhoef, E. Walton, M. Wang, Y. Wang, J. M. Wardlaw, W. Wen, L. T. Westlye, C. D. Whelan, S. H. Witt, K. Wittfeld, C. Wolf, T. Wolfers, J. Q. Wu, C. L. Yasuda, D. Zaremba, Z. Zhang, M. P. Zwiers, E. Artiges, A. A. Assareh, R. Ayesa-Arriola, A. Belger, C. L. Brandt, G. G. Brown, S. Cichon, J. E. Curran, G. E. Davies, F. Degenhardt, M. F. Dennis, B. Dietsche, S. Djurovic, C. P. Doherty, R. Espiritu, D. Garijo, Y. Gil, P. A. Gowland, R. C. Green, A. N. Häusler, W. Heindel, B.-C. Ho, W. U. Hoffmann, F. Holsboer, G. Homuth, N. Hosten, C. R. Jack, M. Jang, A. Jansen, N. A. Kimbrel, K. Kolskår, S. Koops, A. Krug, K. O. Lim, J. J. Luykx, D. H. Mathalon, K. A. Mather, V. S. Mattay, S. Matthews, J. Mayoral Van Son, S. C. McEwen, I. Melle, D. W. Morris, B. A. Mueller, M. Nauck, J. E. Nordvik, M. M. Nöthen, D. S. O'Leary, N. Opel, M.-L. P. Martinot, G. B. Pike, A. Preda, E. B. Quinlan, P. E. Rasser, V. Ratnakar, S. Reppermund, V. M. Steen, P. A. Tooney, F. R. Torres, D. J. Veltman, J. T. Voyvodic, R. Whelan, T. White, H. Yamamori, H. H. H. Adams, J. C. Bis, S. Debette, C. Decarli, M. Fornage, V. Gudnason, E. Hofer, M. A. Ikram, L. Launer, W. T. Longstreth, O. L. Lopez, B. Mazoyer, T. H. Mosley, G. V. Roshchupkin, C. L. Satizabal, R. Schmidt, S. Seshadri, Q. Yang, Alzheimer's Disease Neuroimaging Initiative, CHARGE Consortium, EPIGEN Consortium, IMAGEN Consortium, SYS Consortium, Parkinson's Progression Markers Initiative, M. K. M. Alvim, D. Ames, T. J. Anderson, O. A. Andreassen, A. Arias-Vasquez, M. E. Bastin, B. T. Baune, J. C. Beckham, J. Blangero, D. I. Boomsma, H. Brodaty, H. G. Brunner, R. L. Buckner, J. K. Buitelaar, J. R. Bustillo, W. Cahn, M. J. Cairns, V. Calhoun, V. J. Carr, X. Caseras, S. Caspers, G. L. Cavalleri, F. Cendes, A. Corvin, B. Crespo-Facorro, J. C. Dalrymple-Alford, U. Dannlowski, E. J. C. de Geus, I. J. Deary, N. Delanty, C. Depondt, S. Desrivières, G. Donohoe, T. Espeseth, G. Fernández, S. E. Fisher, H. Flor, A. J. Forstner, C. Francks, B. Franke, D. C. Glahn, R. L. Gollub, H. J. Grabe, O. Gruber, A. K. Håberg, A. R. Hariri, C. A. Hartman, R. Hashimoto, A. Heinz, F. A. Henskens, M. H. J. Hillegers, P. J. Hoekstra, A. J. Holmes, L. E. Hong, W. D. Hopkins, H. E. Hulshoff Pol, T. L. Jernigan, E. G. Jönsson, R. S. Kahn, M. A. Kennedy, T. T. J. Kircher, P. Kochunov, J. B. J. Kwok, S. Le Hellard, C. M. Loughland, N. G. Martin, J.-L. Martinot, C. McDonald, K. L. McMahon, A. Meyer-Lindenberg, P. T. Michie, R. A. Morey, B. Mowry, L. Nyberg, J. Oosterlaan, R. A. Ophoff, C. Pantelis, T. Paus, Z. Pausova, B. W. J. H. Penninx, T. J. C. Polderman, D. Posthuma, M. Rietschel, J. L. Roffman, L. M. Rowland, P. S. Sachdev, P. G. Sämann, U. Schall, G. Schumann, R. J. Scott, K. Sim, S. M. Sisodiya, J. W. Smoller, I. E. Sommer, B. St Pourcain, D. J. Stein, A. W. Toga, J. N. Trollor, N. J. A. Van der Wee, D. van 't Ent, H. Völzke, H. Walter, B. Weber, D. R. Weinberger, M. J. Wright, J. Zhou, J. L. Stein, P. M. Thompson, S. E. Medland, Enhancing NeuroImaging Genetics through Meta-Analysis Consortium (ENIGMA)—Genetics working group, The genetic architecture of the human cerebral cortex. *Science*. **367** (2020), doi:10.1126/science.aay6690.

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

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

24. 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-Rodríguez, H. Won, C. A. de Leeuw, A. F. Pardiñas, BrainSpan Consortium, PsychENCODE Consortium, PsychENCODE Developmental Subgroup, M. Hu, F. Jin, Y. Li, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, D. Posthuma, M. A. Reimers, P. Levitt, D. R. Weinberger, T. M. Hyde, J. E. Kleinman, D. H. Geschwind, M. J. Hawrylycz, M. W. State, S. J. Sanders, P. F. Sullivan, M. B. Gerstein, E. S. Lein, J. A. Knowles, N. Sestan, Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science*. **362** (2018), doi:10.1126/science.aat7615.

25. M. Sur, J. L. R. Rubenstein, Patterning and plasticity of the cerebral cortex. *Science*. **310**, 805–810 (2005).

26. E. S. Monuki, C. A. Walsh, Mechanisms of cerebral cortical patterning in mice and humans. *Nat. Neurosci.* **4 Suppl**, 1199–1206 (2001).

27. M. B. Johnson, Y. I. Kawasawa, C. E. Mason, Z. Krsnik, G. Coppola, D. Bogdanović, D. H. Geschwind, S. M. Mane, M. W. State, N. Sestan, Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron*. **62**, 494–509 (2009).

28. J. B. Burt, M. Demirtaş, W. J. Eckner, N. M. Navejar, J. L. Ji, W. J. Martin, A. Bernacchia, A. Anticevic, J. D. Murray, Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat Neurosci*. **21**, 1251–1259 (2018).

29. R. Gao, R. L. van den Brink, T. Pfeffer, B. Voytek, Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *Elife*. **9** (2020), doi:10.7554/eLife.61277.

30. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. D. Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Vawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. N. Grant, A. R. Jones, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. **489**, 391–399 (2012).

1005 31. S.-J. Hong, R. Vos de Wael, R. A. I. Bethlehem, S. Lariviere, C. Paquola, S. L. Valk, M. P.
1006 Milham, A. Di Martino, D. S. Margulies, J. Smallwood, B. C. Bernhardt, Atypical
1007 functional connectome hierarchy in autism. *Nat Commun*. **10**, 1022 (2019).

1008 32. I. Voineagu, X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, J. Mill, R. M. Cantor,
1009 B. J. Blencowe, D. H. Geschwind, Transcriptomic analysis of autistic brain reveals
1010 convergent molecular pathology. *Nature*. **474**, 380–384 (2011).

1011 33. N. N. Parikshak, V. Swarup, T. G. Belgard, M. Irimia, G. Ramaswami, M. J. Gandal, C.
1012 Hartl, V. Leppa, L. de la T. Ubieta, J. Huang, J. K. Lowe, B. J. Blencowe, S. Horvath, D.
1013 H. Geschwind, Genome-wide changes in lncRNA, splicing, and regional gene expression
1014 patterns in autism. *Nature*. **540**, 423–427 (2016).

1015 34. M. L. Chow, T. Pramparo, M. E. Winn, C. C. Barnes, H.-R. Li, L. Weiss, J.-B. Fan, S.
1016 Murray, C. April, H. Belinson, X.-D. Fu, A. Wynshaw-Boris, N. J. Schork, E. Courchesne,
1017 Age-dependent brain gene expression and copy number anomalies in autism suggest
1018 distinct pathological processes at young versus mature ages. *PLoS Genet.* **8**, e1002592
1019 (2012).

1020 35. J. L. Freese, D. Pino, S. J. Pleasure, Wnt signaling in development and disease. *Neurobiol.*
1021 *Dis.* **38**, 148–153 (2010).

1022 36. V. H. Gazestani, T. Pramparo, S. Nalabolu, B. P. Kellman, S. Murray, L. Lopez, K. Pierce,
1023 E. Courchesne, N. E. Lewis, A perturbed gene network containing PI3K–AKT, RAS–
1024 ERK and WNT–β-catenin pathways in leukocytes is linked to ASD genetics and symptom
1025 severity. *Nat Neurosci*. **22**, 1624–1634 (2019).

1026 37. M. C. Marchetto, H. Belinson, Y. Tian, B. C. Freitas, C. Fu, K. Vadodaria, P. Beltrao-Braga,
1027 C. A. Trujillo, A. P. D. Mendes, K. Padmanabhan, Y. Nunez, J. Ou, H. Ghosh, R. Wright,
1028 K. Brennand, K. Pierce, L. Eichenfield, T. Pramparo, L. Eyler, C. C. Barnes, E.
1029 Courchesne, D. H. Geschwind, F. H. Gage, A. Wynshaw-Boris, A. R. Muotri, Altered
1030 proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol.*
1031 *Psychiatry*. **22**, 820–835 (2017).

1032 38. T. N. Turner, K. Sharma, E. C. Oh, Y. P. Liu, R. L. Collins, M. X. Sosa, D. R. Auer, H.
1033 Brand, S. J. Sanders, D. Moreno-De-Luca, V. Pihur, T. Plona, K. Pike, D. R. Soppet, M.
1034 W. Smith, S. W. Cheung, C. L. Martin, M. W. State, M. E. Talkowski, E. Cook, R.
1035 Huganir, N. Katsanis, A. Chakravarti, Loss of δ-catenin function in severe autism. *Nature*.
1036 **520**, 51–56 (2015).

1037 39. K. Wagstyl, J. P. Lerch, in *Brain Morphometry*, G. Spalletta, F. Piras, T. Gili, Eds. (Springer
1038 New York, New York, NY, 2018; http://link.springer.com/10.1007/978-1-4939-7647-
1039 8_3), vol. 136 of *Neuromethods*, pp. 35–49.

1040 40. E. Courchesne, R. Carper, N. Akshoomoff, Evidence of brain overgrowth in the first year of
1041 life in autism. *JAMA*. **290**, 337–344 (2003).

41. E. Courchesne, P. R. Mouton, M. E. Calhoun, K. Semendeferi, C. Ahrens-Barbeau, M. J. Hallet, C. C. Barnes, K. Pierce, Neuron number and size in prefrontal cortex of children with autism. *JAMA*. **306**, 2001–2010 (2011).

42. R. Sacco, S. Gabriele, A. M. Persico, Head circumference and brain size in autism spectrum disorder: A systematic review and meta-analysis. *Psychiatry Res*. **234**, 239–251 (2015).

43. R. Stoner, M. L. Chow, M. P. Boyle, S. M. Sunkin, P. R. Mouton, S. Roy, A. Wynshaw-Boris, S. A. Colamarino, E. S. Lein, E. Courchesne, Patches of disorganization in the neocortex of children with autism. *N. Engl. J. Med.* **370**, 1209–1219 (2014).

44. Z. Lin, C. Yang, Y. Zhu, J. Duchi, Y. Fu, Y. Wang, B. Jiang, M. Zamanighomi, X. Xu, M. Li, N. Sestan, H. Zhao, W. H. Wong, Simultaneous dimension reduction and adjustment for confounding variation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14662–14667 (2016).

45. T. J. Nowakowski, A. Bhaduri, A. A. Pollen, B. Alvarado, M. A. Mostajo-Radji, E. Di Lullo, M. Haeussler, C. Sandoval-Espinosa, S. J. Liu, D. Velmeshev, J. R. Ounadjela, J. Shuga, X. Wang, D. A. Lim, J. A. West, A. A. Leyrat, W. J. Kent, A. R. Kriegstein, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*. **358**, 1318–1323 (2017).

46. A. M. Fjell, H. Grydeland, S. K. Krogsrud, I. Amlien, D. A. Rohani, L. Ferschmann, A. B. Storsve, C. K. Tamnes, R. Sala-Llonch, P. Due-Tønnessen, A. Bjørnerud, A. E. Sølsnes, A. K. Håberg, J. Skranes, H. Bartsch, C.-H. Chen, W. K. Thompson, M. S. Panizzon, W. S. Kremen, A. M. Dale, K. B. Walhovd, Development and aging of cortical thickness correspond to genetic organization patterns. *Proc Natl Acad Sci USA*. **112**, 15462–15467 (2015).

47. D. Polioudakis, L. de la Torre-Ubieta, J. Langerman, A. G. Elkins, X. Shi, J. L. Stein, C. K. Vuong, S. Nichterwitz, M. Gevorgian, C. K. Opland, D. Lu, W. Connell, E. K. Ruzzo, J. K. Lowe, T. Hadzic, F. I. Hinz, S. Sabri, W. E. Lowry, M. B. Gerstein, K. Plath, D. H. Geschwind, A Single-Cell Transcriptomic Atlas of Human Neocortical Development during Mid-gestation. *Neuron*. **103**, 785-801.e8 (2019).

48. G. Tang, K. Gudsnuk, S.-H. Kuo, M. L. Cotrina, G. Rosoklija, A. Sosunov, M. S. Sonders, E. Kanter, C. Castagna, A. Yamamoto, Z. Yue, O. Arancio, B. S. Peterson, F. Champagne, A. J. Dwork, J. Goldman, D. Sulzer, Loss of mTOR-Dependent Macroautophagy Causes Autistic-like Synaptic Pruning Deficits. *Neuron*. **83**, 1131–1143 (2014).

49. A. T. Hilliard, J. E. Miller, E. R. Fraley, S. Horvath, S. A. White, Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron*. **73**, 537–552 (2012).

50. G. Konopka, T. Friedrich, J. Davis-Turak, K. Winden, M. C. Oldham, F. Gao, L. Chen, G.-Z. Wang, R. Luo, T. M. Preuss, D. H. Geschwind, Human-specific transcriptional networks in the brain. *Neuron*. **75**, 601–617 (2012).

51. X. Liu, D. Han, M. Somel, X. Jiang, H. Hu, P. Guijarro, N. Zhang, A. Mitchell, T. Halene, J. J. Ely, C. C. Sherwood, P. R. Hof, Z. Qiu, S. Pääbo, S. Akbarian, P. Khaitovich, Disruption of an Evolutionarily Novel Synaptic Expression Pattern in Autism. *PLoS Biol.* **14**, e1002558 (2016).

52. Y. Zhu, A. M. M. Sousa, T. Gao, M. Skarica, M. Li, G. Santpere, P. Esteller-Cucala, D. Juan, L. Ferrández-Peral, F. O. Gulden, M. Yang, D. J. Miller, T. Marques-Bonet, Y. Imamura Kawasawa, H. Zhao, N. Sestan, Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science*. **362** (2018), doi:10.1126/science.aat8077.

53. H. Won, J. Huang, C. K. Opland, C. L. Hartl, D. H. Geschwind, Human evolved regulatory elements modulate genes involved in cortical expansion and neurodevelopmental disease susceptibility. *Nat Commun*. **10**, 2396 (2019).

54. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, B. Aleksic, R. Anney, M. Barbosa, S. Bishop, A. Brusco, J. Bybjerg-Grauholm, A. Carracedo, M. C. Y. Chan, A. G. Chiocchetti, B. H. Y. Chung, H. Coon, M. L. Cuccaro, A. Curró, B. Dalla Bernardina, R. Doan, E. Domenici, S. Dong, C. Fallerini, M. Fernández-Prieto, G. B. Ferrero, C. M. Freitag, M. Fromer, J. J. Gargus, D. Geschwind, E. Giorgio, J. González-Peñas, S. Guter, D. Halpern, E. Hansen-Kiss, X. He, G. E. Herman, I. Hertz-Picciotto, D. M. Hougaard, C. M. Hultman, I. Ionita-Laza, S. Jacob, J. Jamison, A. Jugessur, M. Kaartinen, G. P. Knudsen, A. Kolevzon, I. Kushima, S. L. Lee, T. Lehtimäki, E. T. Lim, C. Lintas, W. I. Lipkin, D. Lopergolo, F. Lopes, Y. Ludena, P. Maciel, P. Magnus, B. Mahjani, N. Maltman, D. S. Manoach, G. Meiri, I. Menashe, J. Miller, N. Minshew, E. M. S. Montenegro, D. Moreira, E. M. Morrow, O. Mors, P. B. Mortensen, M. Mosconi, P. Muglia, B. M. Neale, M. Nordentoft, N. Ozaki, A. Palotie, M. Parellada, M. R. Passos-Bueno, M. Pericak-Vance, A. M. Persico, I. Pessah, K. Puura, A. Reichenberg, A. Renieri, E. Riberi, E. B. Robinson, K. E. Samocha, S. Sandin, S. L. Santangelo, G. Schellenberg, S. W. Scherer, S. Schlitt, R. Schmidt, L. Schmitt, I. M. W. Silva, T. Singh, P. M. Siper, M. Smith, G. Soares, C. Stoltenberg, P. Suren, E. Susser, J. Sweeney, P. Szatmari, L. Tang, F. Tassone, K. Teufel, E. Trabetti, M. del P. Trelles, C. A. Walsh, L. A. Weiss, T. Werge, D. M. Werling, E. M. Wigdor, E. Wilkinson, A. J. Willsey, T. W. Yu, M. H. C. Yu, R. Yuen, E. Zachi, E. Agerbo, T. D. Als, V. Appadurai, M. Bækvad-Hansen, R. Belliveau, A. Buil, C. E. Carey, F. Cerrato, K. Chambert, C. Churchhouse, S. Dalsgaard, D. Demontis, A. Dumont, J. Goldstein, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, D. P. Howrigan, H. Huang, J. Maller, A. R. Martin, J. Martin, M. Mattheisen, J. Moran, J. Pallesen, D. S. Palmer, C. B. Pedersen, M. G. Pedersen, T. Poterba, J. B. Poulsen, S. Ripke, A. J. Schork, W. K. Thompson, P. Turley, R. K. Walters, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* (2020), doi:10.1016/j.cell.2019.12.036.

55. B. S. Abrahams, D. E. Arking, D. B. Campbell, H. C. Mefford, E. M. Morrow, L. A. Weiss, I. Menashe, T. Wadkins, S. Banerjee-Basu, A. Packer, SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism*. **4**, 36 (2013).

56. M. J. Gandal, P. Zhang, E. Hadjimichael, R. L. Walker, C. Chen, S. Liu, H. Won, H. van Bakel, M. Varghese, Y. Wang, A. W. Shieh, J. Haney, S. Parhami, J. Belmont, M. Kim, P. Moran Losada, Z. Khan, J. Mleczko, Y. Xia, R. Dai, D. Wang, Y. T. Yang, M. Xu, K. Fish, P. R. Hof, J. Warrell, D. Fitzgerald, K. White, A. E. Jaffe, PsychENCODE Consortium, M. A. Peters, M. Gerstein, C. Liu, L. M. Iakoucheva, D. Pinto, D. H. Geschwind, Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science*. **362**, eaat8127 (2018).

57. D. Velmeshev, L. Schirmer, D. Jung, M. Haeussler, Y. Perez, S. Mayer, A. Bhaduri, N. Goyal, D. H. Rowitch, A. R. Kriegstein, Single-cell genomics identifies cell type–specific molecular changes in autism. *Science*. **364**, 685–689 (2019).

58. S. Mayer, J. Chen, D. Velmeshev, A. Mayer, U. C. Eze, A. Bhaduri, C. E. Cunha, D. Jung, A. Arjun, E. Li, B. Alvarado, S. Wang, N. Lovegren, M. L. Gonzales, L. Szpankowski, A. Leyrat, J. A. A. West, G. Panagiotakos, A. Alvarez-Buylla, M. F. Paredes, T. J. Nowakowski, A. A. Pollen, A. R. Kriegstein, Multimodal Single-Cell Analysis Reveals Physiological Maturation in the Developing Human Neocortex. *Neuron*. **102**, 143-158.e7 (2019).

59. A. Ansel, J. P. Rosenzweig, P. D. Zisman, M. Melamed, B. Gesundheit, Variation in Gene Expression in Autism Spectrum Disorders: An Extensive Review of Transcriptomic Studies. *Front Neurosci*. **10**, 601 (2016).

60. E. Courchesne, K. Campbell, S. Solso, Brain growth across the life span in autism: age-specific changes in anatomical pathology. *Brain Res*. **1380**, 138–145 (2011).

61. T. Pramparo, K. Pierce, M. V. Lombardo, C. C. Barnes, S. Marinero, C. Ahrens-Barbeau, S. S. Murray, L. Lopez, R. Xu, E. Courchesne, Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices. *JAMA Psychiatry*. **72**, 386–394 (2015).

62. T. Pramparo, M. V. Lombardo, K. Campbell, C. C. Barnes, S. Marinero, S. Solso, J. Young, M. Mayo, A. Dale, C. Ahrens-Barbeau, S. S. Murray, L. Lopez, N. Lewis, K. Pierce, E. Courchesne, Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol. Syst. Biol.* **11**, 841 (2015).

63. M. V. Lombardo, L. Eyler, A. Moore, M. Datko, C. Carter Barnes, D. Cha, E. Courchesne, K. Pierce, Default mode-visual network hypoconnectivity in an autism subtype with pronounced social visual engagement difficulties. *Elife*. **8** (2019), doi:10.7554/eLife.47427.

64. K. Pierce, D. Conant, R. Hazin, R. Stoner, J. Desmond, Preference for geometric patterns early in life as a risk factor for autism. *Arch. Gen. Psychiatry*. **68**, 101–109 (2011).

65. K. Pierce, S. Marinero, R. Hazin, B. McKenna, C. C. Barnes, A. Malige, Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated With Increased Symptom Severity. *Biol. Psychiatry*. **79**, 657–666 (2016).

66. K. Pierce, V. H. Gazestani, E. Bacon, C. C. Barnes, D. Cha, S. Nalabolu, L. Lopez, A. Moore, S. Pence-Stophaeros, E. Courchesne, Evaluation of the Diagnostic Stability of the Early Autism Spectrum Disorder Phenotype in the General Population Starting at 12 Months. *JAMA Pediatr*. **173**, 578–587 (2019).

67. K. Pierce, V. Gazestani, E. Bacon, E. Courchesne, A. Cheng, C. C. Barnes, S. Nalabolu, D. Cha, S. Arias, L. Lopez, C. Pham, K. Gaines, G. Gyurjyan, T. Cook-Clark, K. Karins, Get SET Early to Identify and Treatment Refer Autism Spectrum Disorder at 1 Year and Discover Factors That Influence Early Diagnosis. *J Pediatr* (2021), doi:10.1016/j.jpeds.2021.04.041.

68. A. Wetherby, B. Prizant, *Communication and Symbolic Behavior Scales Developmental Profile, First Normed Edition* (Paul H. Brookes, Baltimore, 2002).

69. A. M. Wetherby, S. Brosnan-Maddox, V. Peace, L. Newton, Validation of the Infant-Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism*. **12**, 487–511 (2008).

70. C. Lord, S. Risi, L. Lambrecht, E. H. Cook, B. L. Leventhal, P. C. DiLavore, A. Pickles, M. Rutter, The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. **30**, 205–223 (2000).

71. E. M. Mullen, *Mullen scales of early learning.* (American Guidance Service, Inc, Circle Pine, MN, 1995).

72. S. Sparrow, D. Cicchetti, Balla, D., *Vineland-II scales of adaptive behavior: survey form manual.* (American Guidance Service Inc, Circle Pines, MN, 2005).

73. M. V. Lombardo, E. Courchesne, N. E. Lewis, T. Pramparo, Hierarchical cortical transcriptome disorganization in autism. *Molecular Autism*. **8**, 29 (2017).

74. C. Onore, M. Careaga, P. Ashwood, The role of immune dysfunction in the pathophysiology of autism. *Brain Behav Immun*. **26**, 383–392 (2012).

75. A. Louveau, I. Smirnov, T. J. Keyes, J. D. Eccles, S. J. Rouhani, J. D. Peske, N. C. Derecki, D. Castle, J. W. Mandell, K. S. Lee, T. H. Harris, J. Kipnis, Structural and functional features of central nervous system lymphatic vessels. *Nature*. **523**, 337–341 (2015).

76. A. Schroeder, O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, T. Ragg, The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).

77. P. Du, W. A. Kibbe, S. M. Lin, lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. **24**, 1547–1548 (2008).

78. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, J. Zhang, Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

79. M. Chikina, E. Zaslavsky, S. C. Sealfon, CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics*. **31**, 1584–1591 (2015).

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

81. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).

82. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*. **9**, 179–194 (1999).

83. B. Fischl, A. Liu, A. M. Dale, Automated manifold surgery: constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans Med Imaging*. **20**, 70–80 (2001).

84. F. Ségonne, J. Pacheco, B. Fischl, Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans Med Imaging*. **26**, 518–529 (2007).

85. B. Fischl, D. H. Salat, A. J. W. van der Kouwe, N. Makris, F. Ségonne, B. T. Quinn, A. M. Dale, Sequence-independent segmentation of magnetic resonance images. *Neuroimage*. **23 Suppl 1**, S69-84 (2004).

86. B. Fischl, A. M. Dale, Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11050–11055 (2000).

87. X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, P. Maguire, D. Rosas, N. Makris, A. Dale, B. Dickerson, B. Fischl, Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *Neuroimage*. **32**, 180–194 (2006).

88. B. Fischl, M. I. Sereno, A. M. Dale, Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*. **9**, 195–207 (1999).

89. S. S. Ghosh, S. Kakunoori, J. Augustinack, A. Nieto-Castanon, I. Kovelman, N. Gaab, J. A. Christodoulou, C. Triantafyllou, J. D. E. Gabrieli, B. Fischl, Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age. *Neuroimage*. **53**, 85–93 (2010).

90. T. L. Jernigan, T. T. Brown, D. J. Hagler, N. Akshoomoff, H. Bartsch, E. Newman, W. K. Thompson, C. S. Bloss, S. S. Murray, N. Schork, D. N. Kennedy, J. M. Kuperman, C. McCabe, Y. Chung, O. Libiger, M. Maddox, B. J. Casey, L. Chang, T. M. Ernst, J. A. Frazier, J. R. Gruen, E. R. Sowell, T. Kenet, W. E. Kaufmann, S. Mostofsky, D. G. Amaral, A. M. Dale, Pediatric Imaging, Neurocognition and Genetics Study, The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *Neuroimage*. **124**, 1149–1154 (2016).

91. J. Levman, P. MacDonald, A. R. Lim, C. Forgeron, E. Takahashi, A pediatric structural MRI analysis of healthy brain development from newborns to young adults. *Hum Brain Mapp*. **38**, 5931–5942 (2017).

92. A. Krishnan, L. J. Williams, A. R. McIntosh, H. Abdi, Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage*. **56**, 455–475 (2011).

93. A. R. McIntosh, N. J. Lobaugh, Partial least squares analysis of neuroimaging data: applications and advances. *Neuroimage*. **23 Suppl 1**, S250-263 (2004).

94. N. Kovacevic, H. Abdi, D. Beaton, A. R. McIntosh, in *New perspectives in partial least squares and related methods* (Springer, New York), pp. 159–170.

95. GTEx Consortium, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. **348**, 648–660 (2015).

96. GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts:, Laboratory, Data Analysis &Coordinating Center (LDACC):, NIH program management:, Biospecimen collection:, Pathology:, eQTL manuscript working group:, A. Battle, C. D. Brown, B. E. Engelhardt, S. B. Montgomery, Genetic effects on gene expression across human tissues. *Nature*. **550**, 204–213 (2017).

97. G. Konopka, T. F. Roberts, Insights into the Neural and Genetic Basis of Vocal Communication. *Cell*. **164**, 1269–1276 (2016).

98. A. R. Pfenning, E. Hara, O. Whitney, M. V. Rivas, R. Wang, P. L. Roulhac, J. T. Howard, M. Wirthlin, P. V. Lovell, G. Ganapathy, J. Mouncastle, M. A. Moseley, J. W. Thompson, E. J. Soderblom, A. Iriki, M. Kato, M. T. P. Gilbert, G. Zhang, T. Bakken, A. Bongaarts, A. Bernard, E. Lein, C. V. Mello, A. J. Hartemink, E. D. Jarvis, Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*. **346**, 1256846 (2014).

99. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

100. J. D. Storey, A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **64**, 479–498 (2002).

101. N. N. Parikshak, R. Luo, A. Zhang, H. Won, J. K. Lowe, V. Chandran, S. Horvath, D. H. Geschwind, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. **155**, 1008–1021 (2013).

102. A. J. Willsey, S. J. Sanders, M. Li, S. Dong, A. T. Tebbenkamp, R. A. Muhle, S. K. Reilly, L. Lin, S. Fertuzinhos, J. A. Miller, M. T. Murtha, C. Bichsel, W. Niu, J. Cotney, A. G. Ercan-Sencicek, J. Gockley, A. R. Gupta, W. Han, X. He, E. J. Hoffman, L. Klei, J. Lei, W. Liu, L. Liu, C. Lu, X. Xu, Y. Zhu, S. M. Mane, E. S. Lein, L. Wei, J. P. Noonan, K. Roeder, B. Devlin, N. Sestan, M. W. State, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. **155**, 997–1007 (2013).

103. E. Eising, A. Carrion-Castillo, A. Vino, E. A. Strand, K. J. Jakielski, T. S. Scerri, M. S. Hildebrand, R. Webster, A. Ma, B. Mazoyer, C. Francks, M. Bahlo, I. E. Scheffer, A. T. Morgan, L. D. Shriberg, S. E. Fisher, A set of regulatory genes co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol. Psychiatry*. **24**, 1065–1078 (2019).

104. J. A. Miller, S.-L. Ding, S. M. Sunkin, K. A. Smith, L. Ng, A. Szafer, A. Ebbert, Z. L. Riley, J. J. Royall, K. Aiona, J. M. Arnold, C. Bennet, D. Bertagnolli, K. Brouner, S. Butler, S. Caldejon, A. Carey, C. Cuhaciyan, R. A. Dalley, N. Dee, T. A. Dolbeare, B. A. C. Facer, D. Feng, T. P. Fliss, G. Gee, J. Goldy, L. Gourley, B. W. Gregor, G. Gu, R. E. Howard, J. M. Jochim, C. L. Kuan, C. Lau, C.-K. Lee, F. Lee, T. A. Lemon, P. Lesnar, B. McMurray, N. Mastan, N. Mosqueda, T. Naluai-Cecchini, N.-K. Ngo, J. Nyhus, A. Oldre, E. Olson, J. Parente, P. D. Parker, S. E. Parry, A. Stevens, M. Pletikos, M. Reding, K. Roll, D. Sandman, M. Sarreal, S. Shapouri, N. V. Shapovalova, E. H. Shen, N. Sjoquist, C. R. Slaughterbeck, M. Smith, A. J. Sodt, D. Williams, L. Zöllei, B. Fischl, M. B. Gerstein, D. H. Geschwind, I. A. Glass, M. J. Hawrylycz, R. F. Hevner, H. Huang, A. R. Jones, J. A. Knowles, P. Levitt, J. W. Phillips, N. Sestan, P. Wohnoutka, C. Dang, A. Bernard, J. G. Hohmann, E. S. Lein, Transcriptional landscape of the prenatal human brain. *Nature*. **508**, 199–206 (2014).

105. J. Cotney, R. A. Muhle, S. J. Sanders, L. Liu, A. J. Willsey, W. Niu, W. Liu, L. Klei, J. Lei, J. Yin, S. K. Reilly, A. T. Tebbenkamp, C. Bichsel, M. Pletikos, N. Sestan, K. Roeder, M. W. State, B. Devlin, J. P. Noonan, The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun*. **6**, 6404 (2015).

106. A. Sugathan, M. Biagioli, C. Golzio, S. Erdin, I. Blumenthal, P. Manavalan, A. Ragavendran, H. Brand, D. Lucente, J. Miles, S. D. Sheridan, A. Stortchevoi, M. Kellis, S.

J. Haggarty, N. Katsanis, J. F. Gusella, M. E. Talkowski, CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci USA*. **111**, E4468–E4477 (2014).

107. J. C. Darnell, S. J. Van Driesche, C. Zhang, K. Y. S. Hung, A. Mele, C. E. Fraser, E. F. Stone, C. Chen, J. J. Fak, S. W. Chi, D. D. Licatalosi, J. D. Richter, R. B. Darnell, FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell*. **146**, 247–261 (2011).

108. E. J. Greenblatt, A. C. Spradling, Fragile X mental retardation 1 gene enhances the translation of large autism-related proteins. *Science*. **361**, 709–712 (2018).

Figures

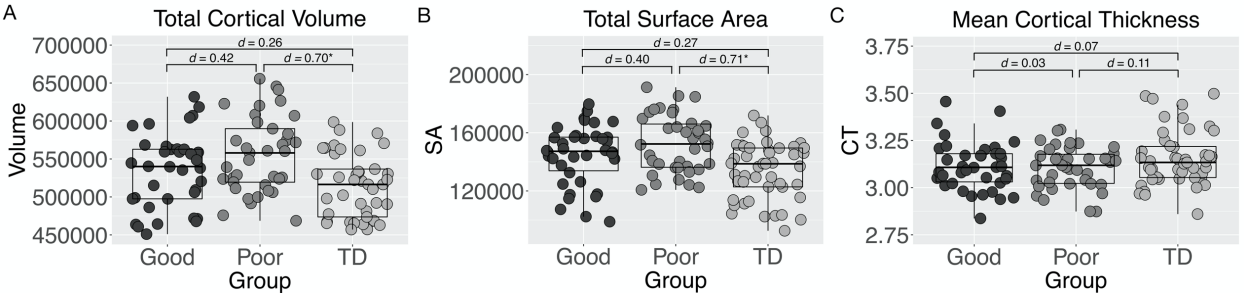

Fig. 1: Subtype differences in total cortical volume (A) and total surface area (B). *Standardized effect sizes (Cohen's d) are shown for each pairwise group comparison. The asterisk indicates statistically significant pairwise group comparisons that survive FDR q<0.05. Panel C shows the data for mean cortical thickness, which was not statistically significant between-groups. Abbreviations: SA, surface area; CT, cortical thickness; Good, ASD Good language subtype; Poor, ASD Poor language subtype, TD, typically-developing.*

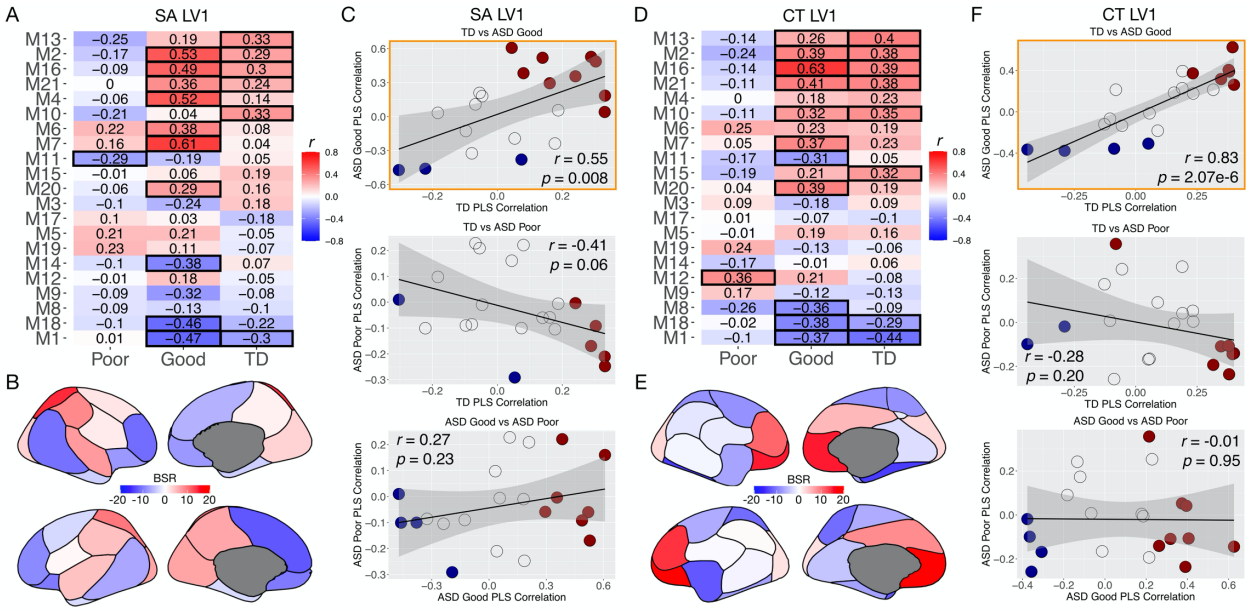

Fig. 2: Normative associations between gene expression and surface area or cortical thickness are absent in ASD Poor. *Panel A show PLS correlations for each gene co-expression module (rows) and each group (columns). Modules with a black outline are 'non-zero' modules where the correlation between gene expression and surface area (SA) 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 non-outlined cells are 'zero' modules that are not sufficiently correlated to SA in a non-zero way (e.g., 95% bootstrap confidence intervals include a correlation of 0. Panel B shows brain bootstrap ratios (BSRs) for each brain region in the GCLUST parcellation. Regions with red BSRs have correlations that manifest in the directionality shown in heatmaps in panel A. Brain regions with blue BSRs have correlations where the directionality is flipped relative to the heatmaps in panel A. Stronger BSRs indicate regions that are more important in driving the SA LV1 relationship. Panel C shows similarity in PLS correlations between groups. 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 (red for positive correlations, blue for negative correlations), while grey dots indicate zero modules. The scatterplots with the orange outline indicates that the PLS SA LV1 relationship manifests similarly for TD and ASD Good. Panels D-F are the same as panels A-C except they show the association between gene co-expression modules at cortical thickness (CT LV1).*

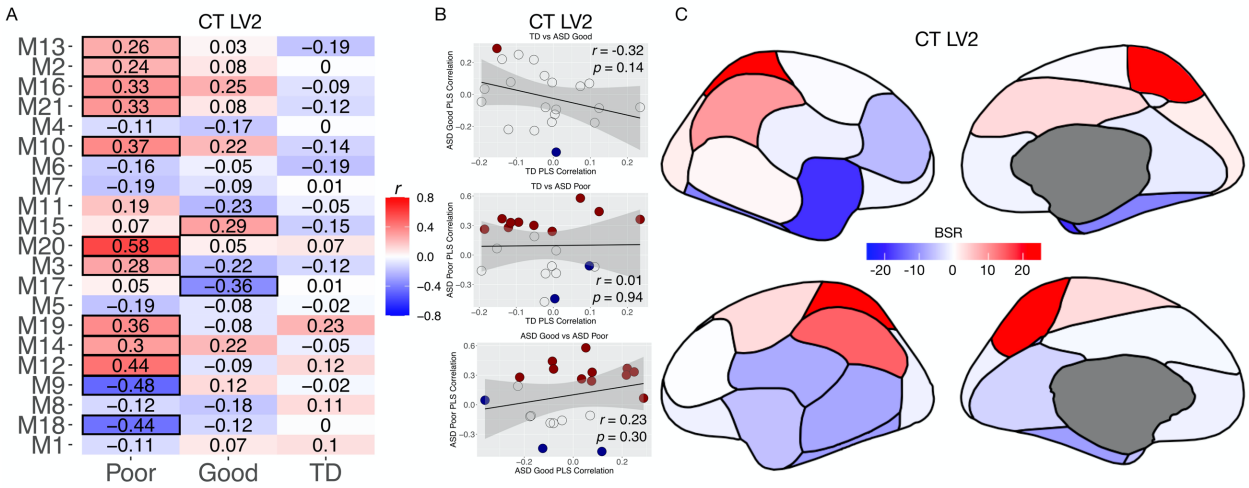

Fig. 3: Multivariate gene co-expression relationships with SA and CT. *Panels A-C show brain bootstrap ratios (BSR) for SA LV1 (A), CT LV1 (B), and CT LV2 (C) for all 12 regions from the GCLUST SA and CT parcellations. Regions that are increasingly colored red and blue are regions that most reliably 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.*

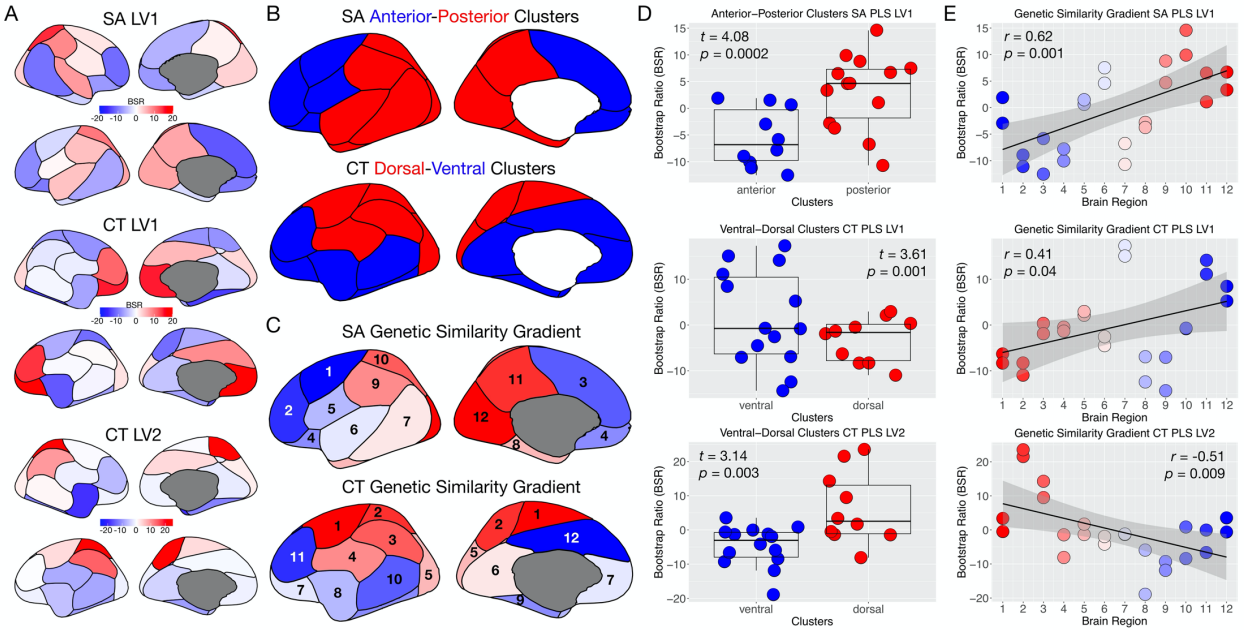

Fig. 4: Cortical patterning along genetic similarity gradients. *Panel A shows brain BSR maps for SA LV1 (top), CT LV1 (middle), and CT LV2 (bottom). Panel B shows the coarse 2-cluster anterior-posterior (A-P) and dorsal-ventral (D-V) genetic similarity partitions identified by Chen and colleagues (19, 20). Panel C shows the rank ordering of regions by hierarchical genetic similarity discovered by Chen and colleagues (19, 20). The rank ordering here defines a genetic similarity gradient for how SA or CT varies across brain regions. Areas rank numbered close together are more genetically similarity than regions numbered farther apart. Panel D shows how brain BSRs are differentiated along A-P and D-V axes (SA LV1, top; CT LV1, middle; CT LV2, bottom). Panel E shows how brain BSRs vary along genetic similarity gradients (SA LV1, top; CT LV1, middle; CT LV2, bottom).*

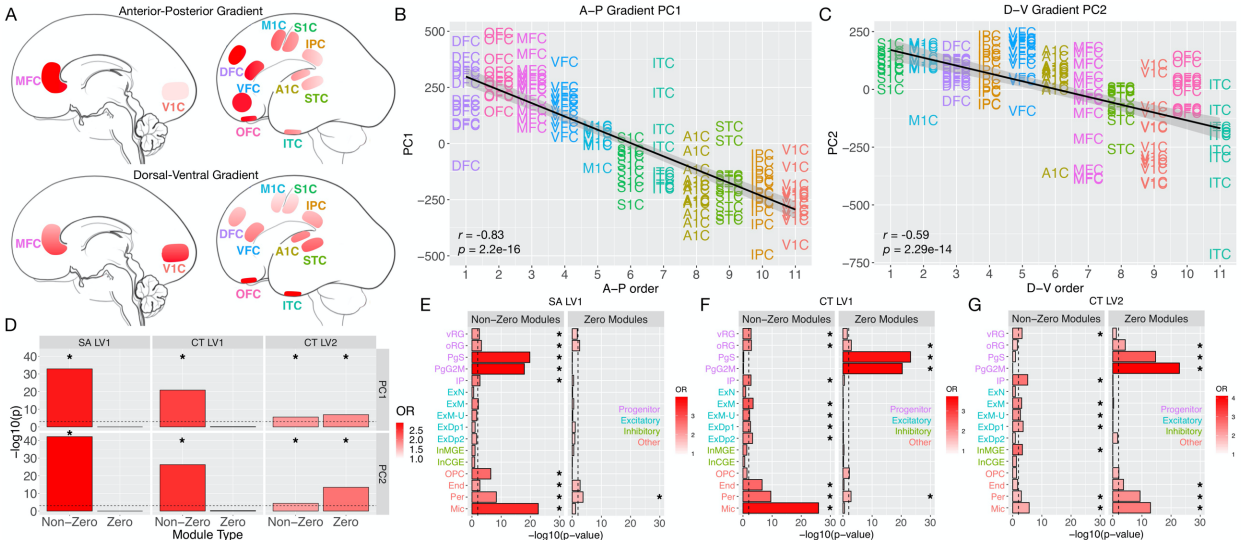

Fig. 5: Enrichment between PLS non-zero modules and genes involved in prenatal A-P and D-V expression gradients and prenatal cell types. *Panels A shows cortical brain areas sampled from 12-24 weeks post-conception from the Development PsychENCODE RNA-seq dataset from Li and colleagues (24). AC-PCA (44) was utilized to isolate anterior-posterior (A-P) (PC1, panel B) and dorsal-ventral (D-V) (PC2, panel C) expression gradients. Panel D shows -log10 p-values for enrichment tests of non-zero and zero modules for SA LV1, CT LV1, and CT LV2 for genes isolated from PC1 and PC2. Panels E-F show enrichments in prenatal cell types for SA LV1 (E), CT LV1 (F), and CT LV2 (G). Abbreviations: A-P, anterior-posterior; D-V dorsal-ventral; PC, principal component; OR, enrichment odds ratio; vRG, ventricular radial glia; oRG, outer radial glia; PgS, cycling progenitors (S phase); PgG2M, cycling progenitors (G2/M phase); IP, intermediate progenitors; ExM, maturing excitatory; ExN, migrating excitatory; ExM-U, maturing excitatory upper enriched; ExDp1, excitatory deep layer 1; ExDp2, excitatory deep layer 2; InCGE, interneuron caudal ganglion eminence; InMGE, interneuron medial ganglion eminence; OPC, oligodendrocyte precursor cells; End, endothelial cells; Per, pericytes; Mic, microglia.*

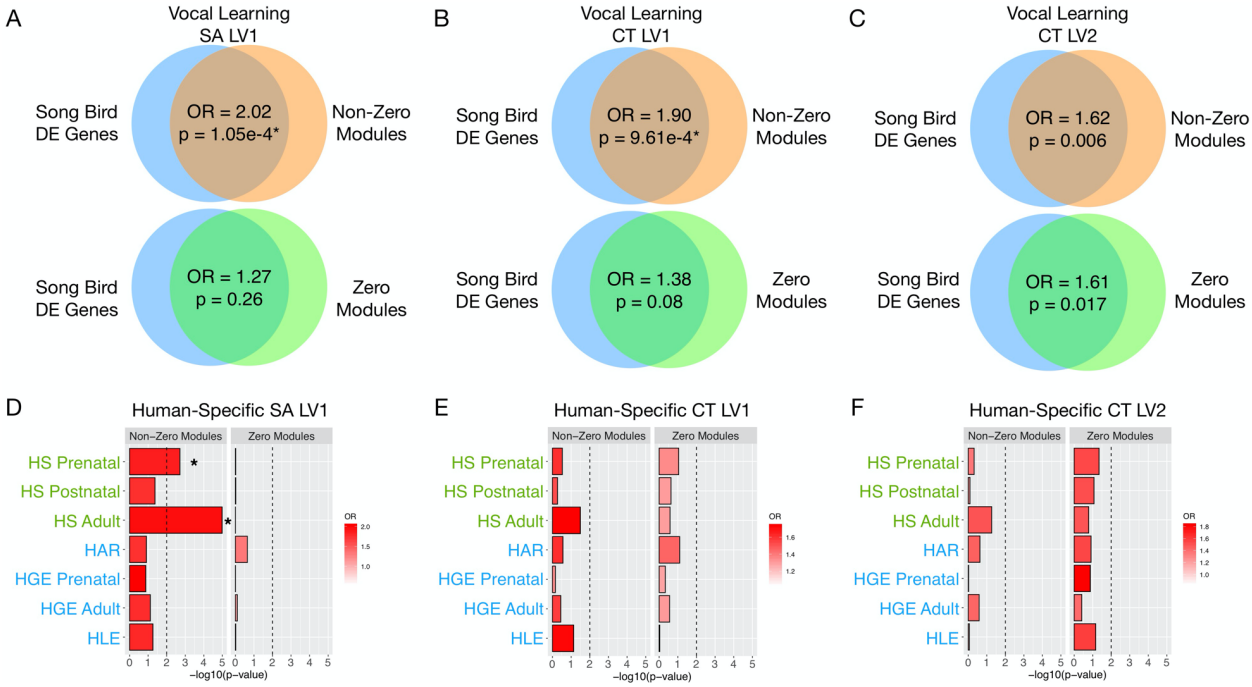

Fig. 6: Enrichments between PLS non-zero modules and songbird vocal learning or human-specific genes. *Panels A-C indicate enrichments between differentially expressed songbird vocal learning genes and non-zero and zero modules for SA LV1 (A), CT LV1 (B), AND CT LV2 (C). Panels D-F indicate enrichments between human-specific genes and non-zero and zero modules for SA LV1 (D), CT LV1 (E), and CT LV2 (F). Asterisks marks enrichments at FDR q<0.01. Abbreviations: DE, differentially expressed; OR, enrichment odds ratio; SA, surface area; CT, cortical thickness; LV, latent variable pair; HS, human-specific; HAR, human-accelerated region; HGE, human-gained enhancer; HLE, human-lossed enhancer.*

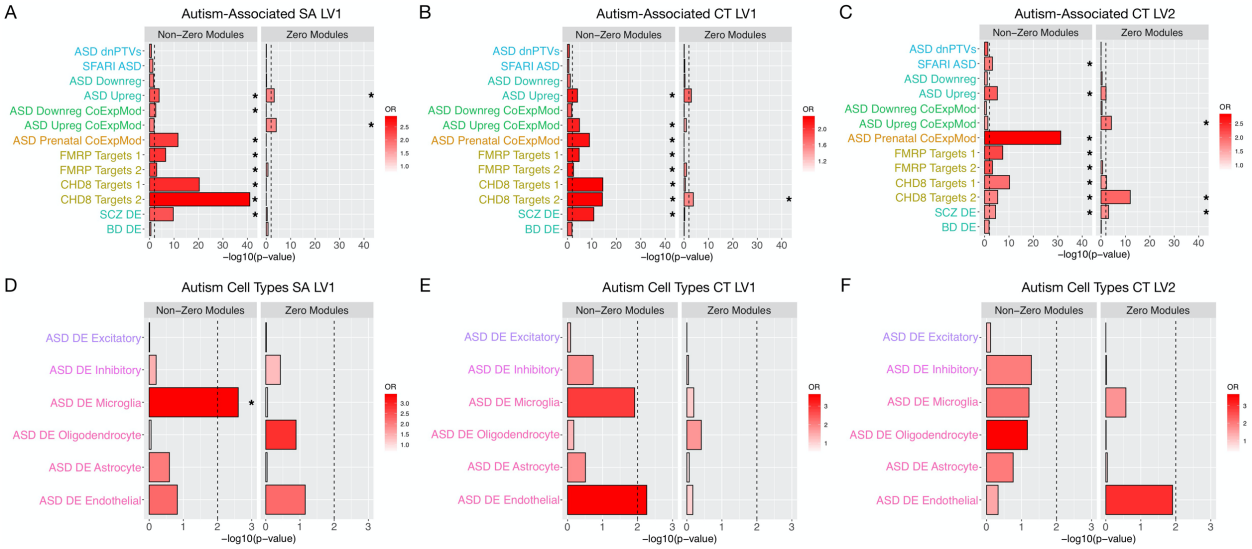

Fig. 7: Enrichment between PLS non-zero modules and autism-associated genes. *Panels A-C indicate enrichments between different autism-associated gene lists and non-zero and zero modules for SA LV1 (A), CT LV1 (B), AND CT LV2 (C). Panels D-F indicate enrichments between differentially expressed genes in specific cell types in autism and non-zero and zero modules for SA LV1 (D), CT LV1 (E), and CT LV2 (F). Asterisks marks enrichments at FDR q<0.01. Abbreviations: DE, differentially expressed; OR, enrichment odds ratio; SA, surface area; CT, cortical thickness; LV, latent variable pair; dnPTVs, de novo protein truncating variants.*

Supplementary Figures

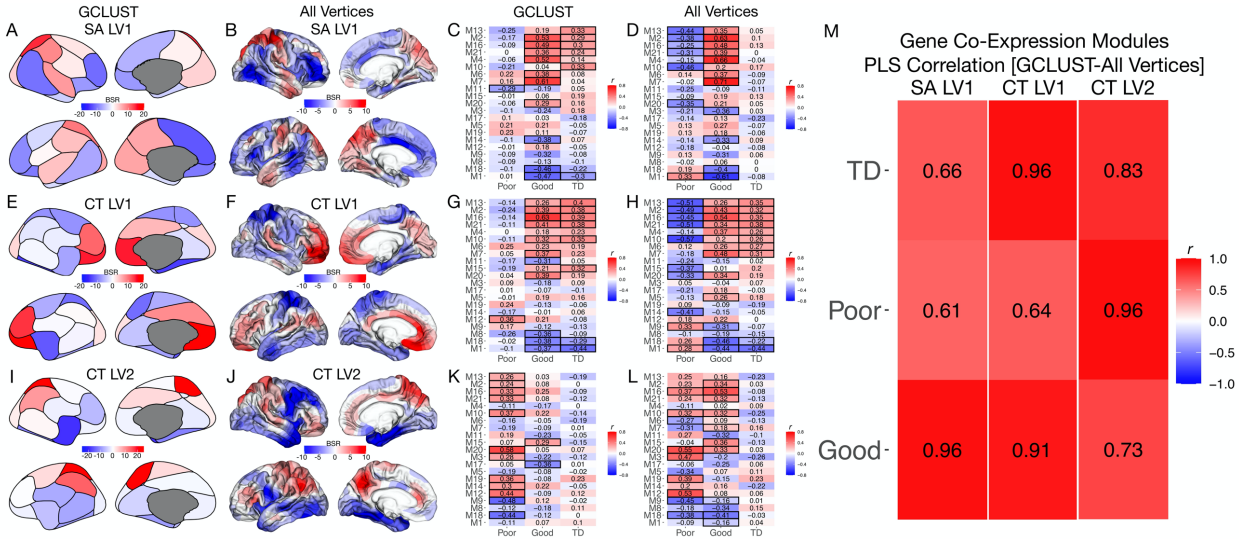

Fig. S1: PLS results on vertex-wise data and comparison to GCLUST. *Panels A-D show GCLUST and vertex-wise (All Vertices) PLS results for SA LV1. Panels E-H show results for CT LV1, while panels I-L show results for CT LV2. BSR maps are shown the two leftmost columns, while the heatmaps showing the correlations for each module and group are shown in the middle columns. Black outlines represent non-zero modules. Panel M shows the correlation between GCLUST and vertex-wise PLS correlations in each group.*

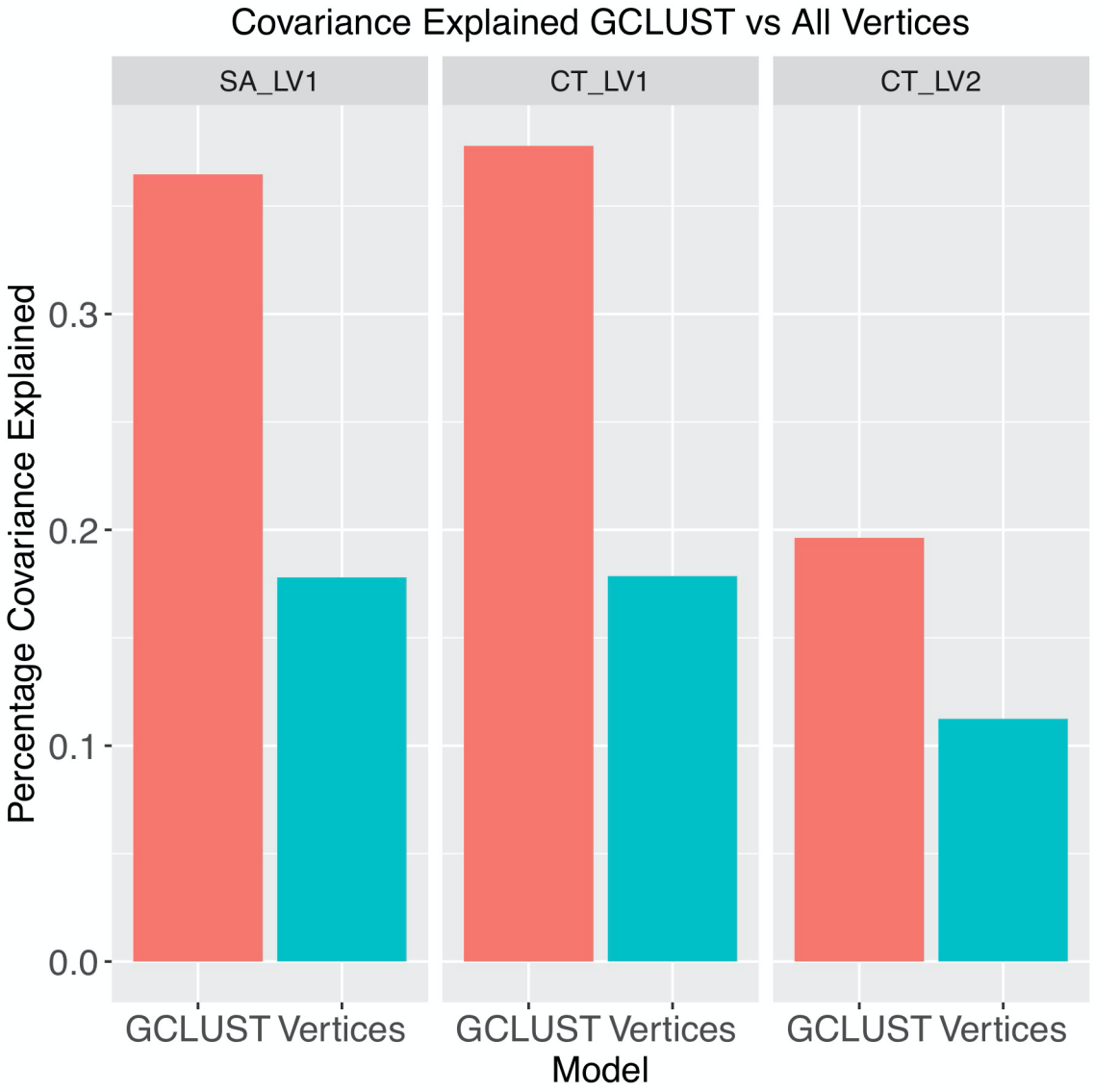

Fig. S2: Percentage of covariance explained in GCLUST and vertex-wise PLS models. *This plot shows the percentage of covariance explained from SA LV1, CT LV1, and CT LV2 in the GCLUST and vertex-wise PLS models.*

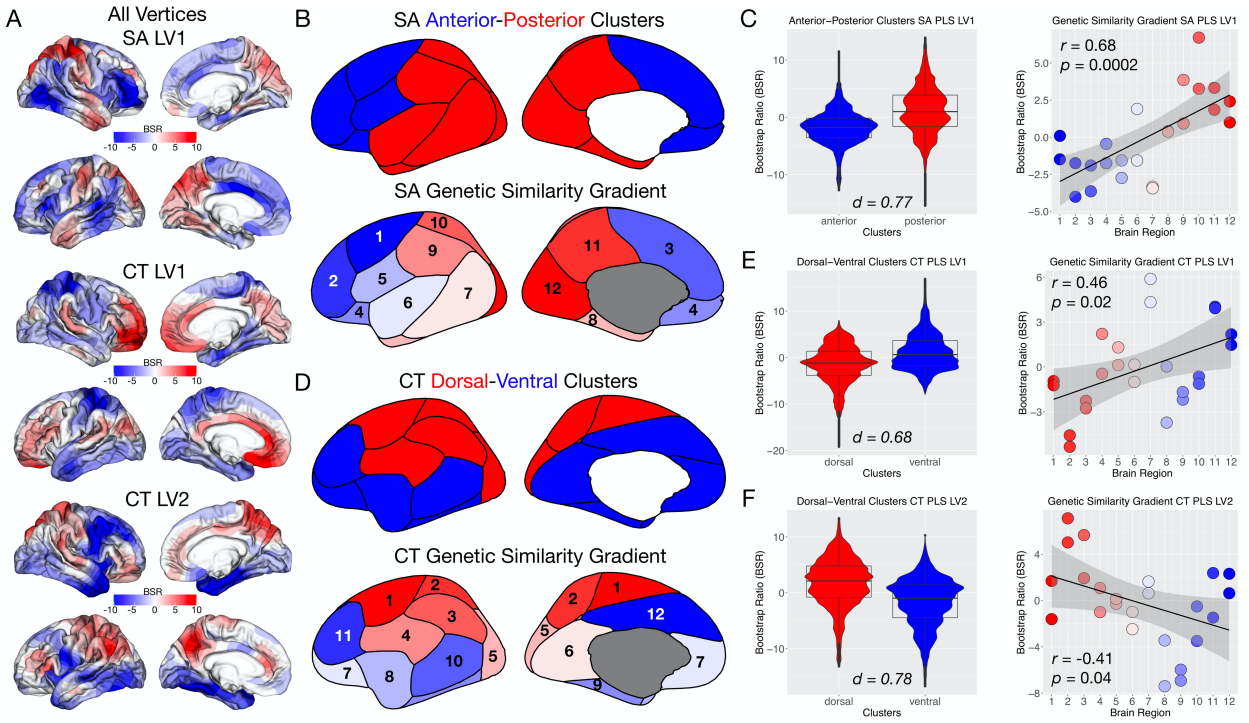

Fig. S3: A-P, D-V, and genetic similarity gradients for SA and CT in vertex-wise PLS model. *Panel A shows brain bootstrap ratios (BSRs) for vertex-wise PLS SA LV1 (top), CT LV1 (middle), and CT LV2 (bottom). Panel B shows the SA A-P and genetic similarity gradients from Chen et al., (19, 20) for the GCLUST parcellation. Panel C shows a violin-boxplot of the SA LV1 BSRs from all vertices within the anterior and posterior clusters (left), and a scatterplot of the median BSR from each GCLUST region, ordered by the genetic similarity rank ordering of brain regions (right). For the A-P difference, the Cohen's d effect size is cited to quantify how large the difference is. All p-values for this A-P distinction in SA LV1 and all other D-V comparisons CT LV1, CT LV2 have p<2.2e-16. Panel D shows the D-V clusters for CT as well as the CT genetic similarity gradient from the Chen et al., (19, 20) GCLUST parcellation. Panels E and F show the D-V (left) and genetic similarity gradient effect (right) for CT LV1 (top) and CT LV2 (bottom).*

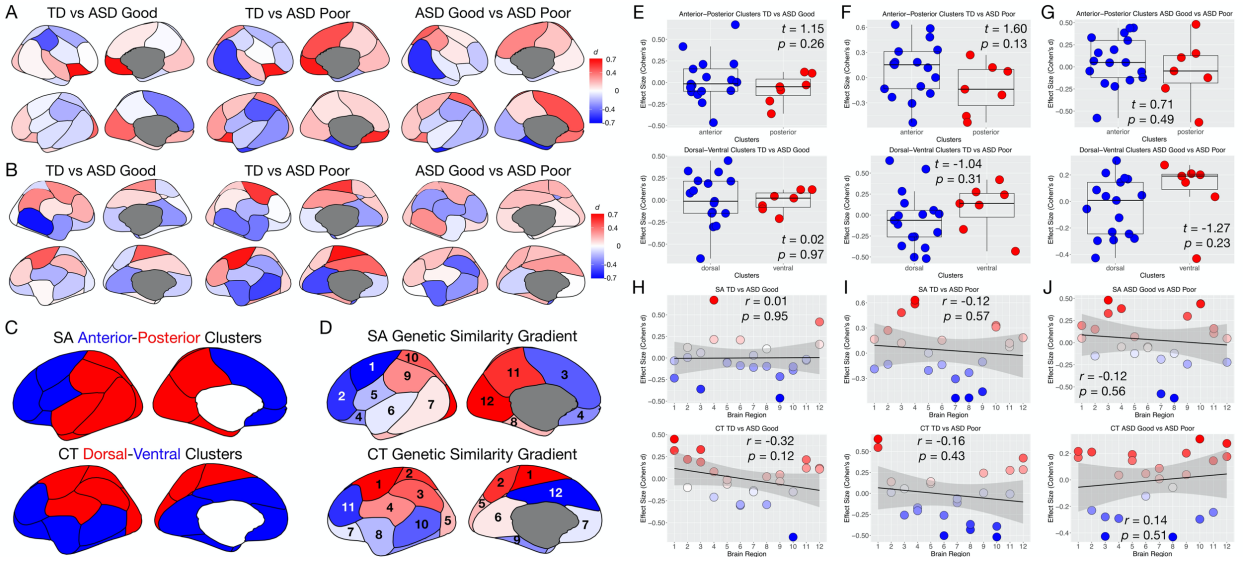

Fig. S4: Lack of A-P and D-V gradients in effect size of group comparisons. *Panel A shows each pairwise group comparison for differences in SA. Panel B shows each pairwise group comparison for differences in CT. Coloring reflects the standardized effect size for the group differene (Cohen's d). Panel C shows a depiction of the A-P and D-V partitions defined by Chen and colleagues (19, 20). Panel D shows how the genetic similarity gradient defined by Chen and colleagues (19, 20) manifests via numbered ordering of brain regions along that gradient. Panels E-G show scatter-boxplots for the GCLUST A-P (top) or D-V (bottom) clusters. Panels H-J show the correlations between Cohen's d effect size and the genetic similarity gradient rank ordering of brain regions for SA (top) and CT (bottom).*

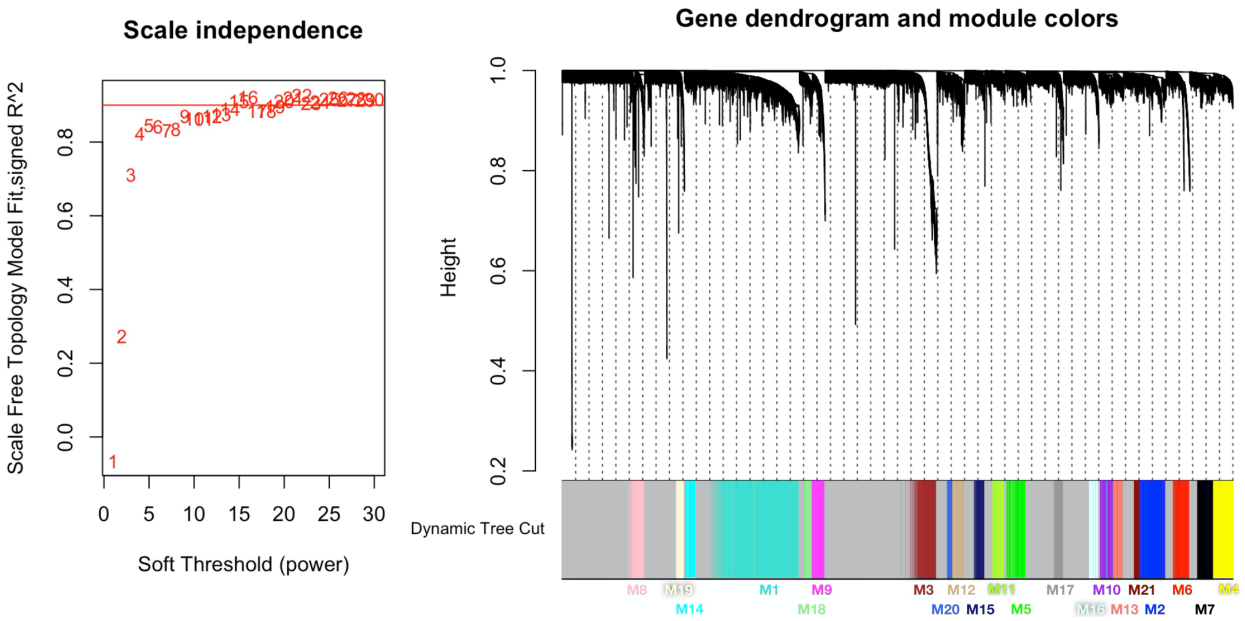

Fig. S5: Soft power and TOM dendrogram from WGCNA analysis. *On the left of this figure we show the soft power plot for the main WGCNA analysis including data from all groups. A horizontal red line depicts soft power topology model fit R^2 of 0.9, where the chosen soft power of 16 is located. On the right of this figure is the TOM dendrogram with modules labeled at the bottom.*

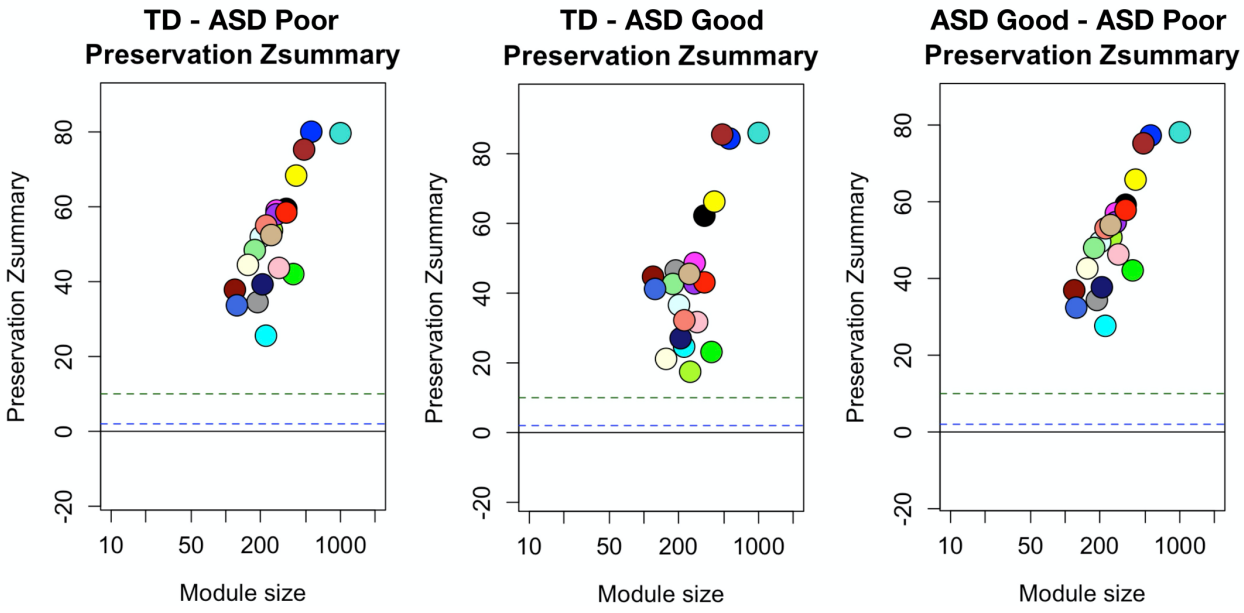

Fig. S6: Module preservation when WGCNA analysis is run separately on each group. *This figure shows the module preservation Zsummary statistic for WGCNA analyses run separately on each group in order to show that networks are highly preserved (Zsummary>10) across groups.*

Supplementary Tables S1-S5

Table S1: Summary of clinical and demographic variables.

Table S2: Statistics from hypothesis tests on group differences in cortical volume, surface area, and cortical thickness.

Table S3: Statistics from hypothesis tests on GCLUST regional group differences in surface area and cortical thickness.

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

Table S5: Enrichment odds ratios and p-values for tests of enrichment between zero and non-zero modules and broadly expressed or brain-specific gene sets.

Table S6: Table annotating overlap of each gene with gene sets used in enrichment analyses.

Table S7: ANOVA stats from CellCODE deconvolution of leukocyte cell types.