

## NEUROSCIENCE

# Atypical genomic cortical patterning in autism with poor early language outcome

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

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 a 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 types (e.g., progenitor cells and excitatory neurons), are functionally important for vocal learning and human-specific evolution, and are prominent in prenatal coexpression 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 affect later regional functional specialization and circuit formation.

## INTRODUCTION

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). Within the context of precision medicine (3), 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 these individuals (1, 2, 4).

Regarding different clinical outcomes in ASD, perhaps the most understudied yet most important distinction is between those with relatively intact and good language levels and those who are minimally verbal or have very poor early language outcome (5). 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 (5). However, because language

level is a key ingredient in helping to facilitate better outcomes, available early interventions may be least effective for these types of individuals (6). 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 magnetic resonance 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 (7). Linked to this functional abnormality are different large-scale patterns of activity in blood leukocyte gene coexpression modules (8). 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 (9) 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 (8). This prenatal enrichment is key because 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 (10, 11). 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 these processes, such as surface area (SA) and cortical thickness (CT) (12–15), may also be substantially altered in ASD Poor versus Good language outcome subtypes.

<sup>1</sup>Laboratory for Autism and Neurodevelopmental Disorders, Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy. <sup>2</sup>Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK. <sup>3</sup>Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. <sup>4</sup>VISN 22 Mental Illness Research, Education, and Clinical Center, VA San Diego Healthcare System, San Diego, CA, USA. <sup>5</sup>Autism Center of Excellence, Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. <sup>6</sup>Center for Multimodal Imaging and Genetics, University of California, San Diego, La Jolla, CA, USA. <sup>7</sup>Department of Radiology, University of California, San Diego, La Jolla, CA, USA. <sup>8</sup>Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA. <sup>9</sup>Department of Child and Adolescent Psychiatry and Behavioral Science, Children's Hospital of Philadelphia, Philadelphia, PA, USA. <sup>10</sup>Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA. <sup>11</sup>Department of Psychiatry, University of Cambridge, Cambridge, UK. <sup>12</sup>Center for Mind/Brain Sciences, University of Trento, Rovereto, Italy. <sup>13</sup>Department of Pediatrics, University of Utah, Salt Lake City, UT, USA. <sup>14</sup>Department of Pediatrics, University of California, San Diego, La Jolla, CA, USA. <sup>15</sup>Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA. <sup>16</sup>Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego, La Jolla, CA, USA.

\*Corresponding author. Email: michael.lombardo@iit.it

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 (12, 15–17). This prenatal genomic patterning is the beginning of cortical arealization processes that allow different cortical regions to develop their own cellular, functional, and circuit identities (12, 16) and thus aid in later regional functional specialization. In the adult brain, this genomic gradient patterning is still evident and correlates with large-scale gradient patterning of structural and functional features (13, 14, 18–20). Cortical arealization or patterning may also be atypical in ASD. Functional connectome gradient organization is altered in ASD (21). Case control comparisons of gene expression in postmortem cortical tissue have found dysregulation of cortical patterning genes and attenuation of gene expression differences in frontal versus temporal cortex (22–24). WNT signaling is also known to affect cortical patterning (12, 25), and WNT signaling abnormalities are also identified in ASD (23, 24, 26, 27), particularly within broadly expressed ASD risk genes (11). Therefore, if broadly expressed genes prenatally affect 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., (7, 8)]?

Here, we investigated these questions by examining how early variability in morphometric measures of the cerebral cortex such as CT and SA is 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 A-P and 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 (12, 15, 28). 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 SA 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 (typically developing),  $n = 47$ ] with both a T1-weighted structural MRI scan and a blood sample that was used to examine gene expression in blood leukocyte cells (see Materials and 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 (EL and RL)  $T$ -scores as in previous studies (see Materials and Methods) (7, 8).

One of the most robust findings on early structural brain development in ASD is the on-average effect of early brain overgrowth in

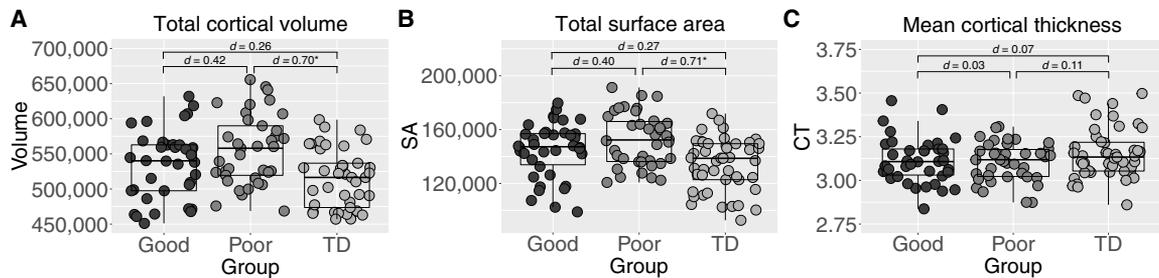
the first years of life (4). Thus, we started by examining whether there are subtype differences on global measures such as total cortical volume (CV), SA, and mean CT. Statistical models controlling for age and sex identified a group effect for total CV and total SA but no effect of group for mean CT (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 SA or CT 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 coexpression network analysis (WGCNA) to reduce expression data of 14,426 genes highly expressed in blood leukocytes to 21 coexpression modules (table S4). Coexpression 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 (13, 14). This analysis allowed us to identify statistically significant multivariate relationships between gene coexpression modules and SA or CT and then allowed for examination of how the relationship manifests across brain regions and coexpression modules and also how these relationships manifest in each group (see Materials and 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  $P_{\text{ucorr}} = 0.01$ , and  $P_{\text{vcorr}} = 0.06$ ), 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 (17%) 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 coexpression modules and groups, in Fig. 2A, we show which coexpression modules have “non-zero” relationships in each group. These “non-zero modules” are the coexpression 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 coexpression modules highly contributing to the SA LV1 relationship. In contrast, coexpression 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,  $P = 1.90 \times 10^{-71}$ ; table S5]. These two observations are compatible with predictions from the omnigenic theory of complex traits, in which variance in complex traits such as ASD is exerted en masse by a large majority of genes that affect 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 (9). These effects are also in line with



**Fig. 1. Subtype differences in total CV and total SA.** Standardized effect sizes (Cohen's  $d$ ) are shown for each pairwise group comparison for total CV (A), total SA (B), and mean CT (C). Asterisks indicate statistically significant pairwise group comparisons that survive false discovery rate (FDR)  $q < 0.05$ . Good, ASD Good language subtype; Poor, ASD Poor language subtype; TD, typically developing.

similar results observed for large-scale gene expression associations with functional imaging phenotypes in ASD early language outcome subtypes (8).

Figure 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 correlation values for these groups ( $r = 0.55$ ,  $P = 0.008$ ) (Fig. 2C). 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 that 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 two statistically significant LV pairs (CT LV1:  $d = 4.30$ ,  $P = 0.0001$ , split-half  $P_{\text{ucorr}} = 0.04$ , and  $P_{\text{vcorr}} = 0.01$ ; CT LV2:  $d = 3.09$ ,  $P = 0.0001$ , split-half  $P_{\text{ucorr}} = 0.02$ , and  $P_{\text{vcorr}} = 0.05$ ), 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, it was not as good as the PLS on GCLUST-parcellated features, as indicated by less percentage of covariance explained (CT LV1 = 17%; CT LV2 = 11%) (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 ( $OR = 2.96$ ,  $P = 4.43 \times 10^{-43}$ ; table S5), and thus are 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. 2, D and 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 CT, specific to ASD Poor language subtype

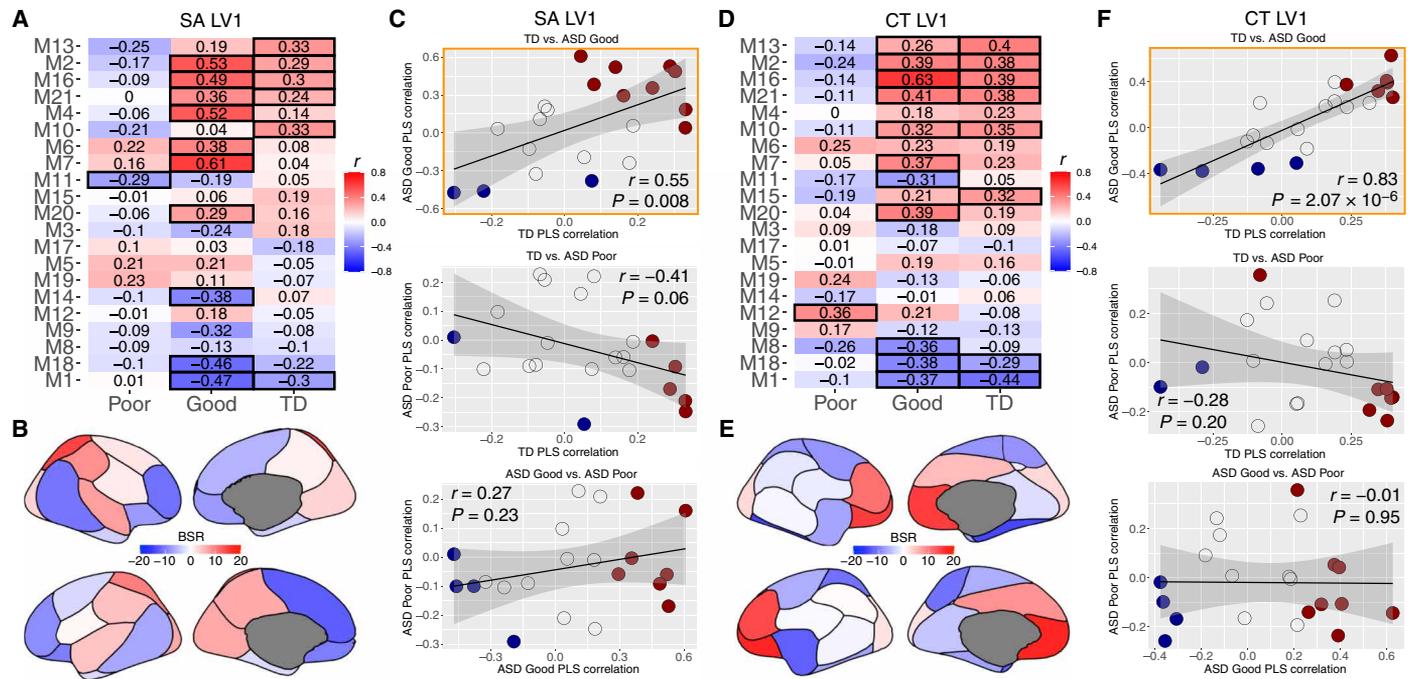
In contrast to CT LV1, the non-zero modules for CT LV2 are almost exclusively relevant for the ASD Poor subtype, comprise about 48% of all genes examined, do not show specific enrichment for broadly expressed genes (table S5), and do not show strong correlations between groups (Fig. 3, A and B). These 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 is sometimes opposite to the directionality shown for CT LV1. For example, superior parietal cortex in CT LV1 has negative brain bootstrap ratio (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 the language-sensitive left hemisphere perisylvian, middle temporal, and inferior parietal cortices, CT LV1 shows 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.

### A-P and D-V gradient patterning of SA and CT 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 not only the importance of each brain region for the LV in question but also 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 two-cluster solution previously identified by Chen and colleagues (13, 14) 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 notable 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 found by Chen and colleagues (13, 14) (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 coexpression relationships with SA and CT reveal how the cortex is genomically patterned to promote the development of cortical regionalization and areal identity (12). Because SA LV1 and CT LV1 are normative



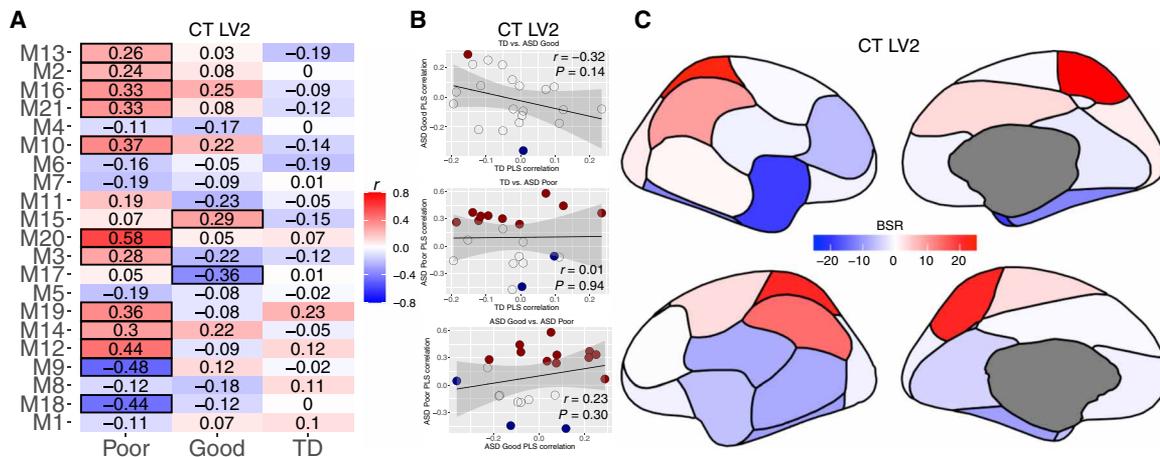
**Fig. 2. Normative associations between gene expression and SA or CT are absent in ASD Poor.** (A) PLS correlations for each gene coexpression module (rows) and each group (columns). Modules with a black outline are non-zero modules where the correlation between gene expression and SA is significantly non-zero, as indicated by 95% bootstrap CIs not encompassing a correlation of 0. These non-zero modules are the strongest contributors to the PLS relationship. All nonoutlined cells are zero modules that are not sufficiently correlated to SA in a non-zero way (e.g., 95% bootstrap CIs include a correlation of 0). (B) BSRs for each brain region in the GCLUST parcellation. Regions with red BSRs have correlations that manifest in the directionality shown in heatmaps in (A). Brain regions with blue BSRs have correlations where the directionality is flipped relative to the heatmaps in (A). Stronger BSRs indicate regions that are more important in driving the SA LV1 relationship. (C) Similarity in PLS correlations between groups. In these scatterplots, each dot is a coexpression 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 and blue for negative correlations), while gray dots indicate zero modules. The scatterplots with the orange outline indicate that the PLS SA LV1 relationship manifests similarly for TD and ASD Good. (D to F) The same as (A) to (C) except that they show the association between gene coexpression modules and CT (CT LV1).

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 the 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. 3). Given the evidence of focal laminar patches throughout the cortex in ASD (29), it will be important for future work to investigate further how these 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 does 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 toward the specific importance of how the underlying genomic mechanisms act to pattern SA and CT across the cortex.

### Genes involved in gradient patterning of SA and CT 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 (12, 15, 17), 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 sequencing (RNA-seq) data from the Development PsychENCODE dataset, we used sparse principal components analysis (PCA) (30) 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 to 24 weeks after conception (e.g., midgestation) (Fig. 5, A to C). 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. Because SA LV1 and CT LV1



**Fig. 3. Atypical association between gene expression and CT that is specific to the ASD Poor subtype.** Atypical association specific to ASD Poor between gene coexpression modules and CT, as described in CT LV2. **(A)** CT LV2 PLS correlations for each gene coexpression module (rows) and each group (columns). Modules with a black outline are non-zero modules where the correlation between gene expression and SA is significantly non-zero, as indicated by 95% bootstrap CIs not encompassing a correlation of 0. Nearly all of the non-zero modules for CT LV2 are present for the ASD Poor subtype. **(B)** Lack of similarity in CT LV2 PLS correlations between groups. In these scatterplots, each dot is a coexpression 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 and blue for negative correlations), while gray dots indicate zero modules. **(C)** CT LV2 BSRs for each brain region in the GCLUST parcellation. Regions with red BSRs have correlations that manifest in the directionality shown in heatmaps in (A). Brain regions with blue BSRs have correlations where the directionality is flipped relative to the heatmaps in (A). Stronger BSRs indicate regions that are more important in driving the CT LV2 relationship.

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.

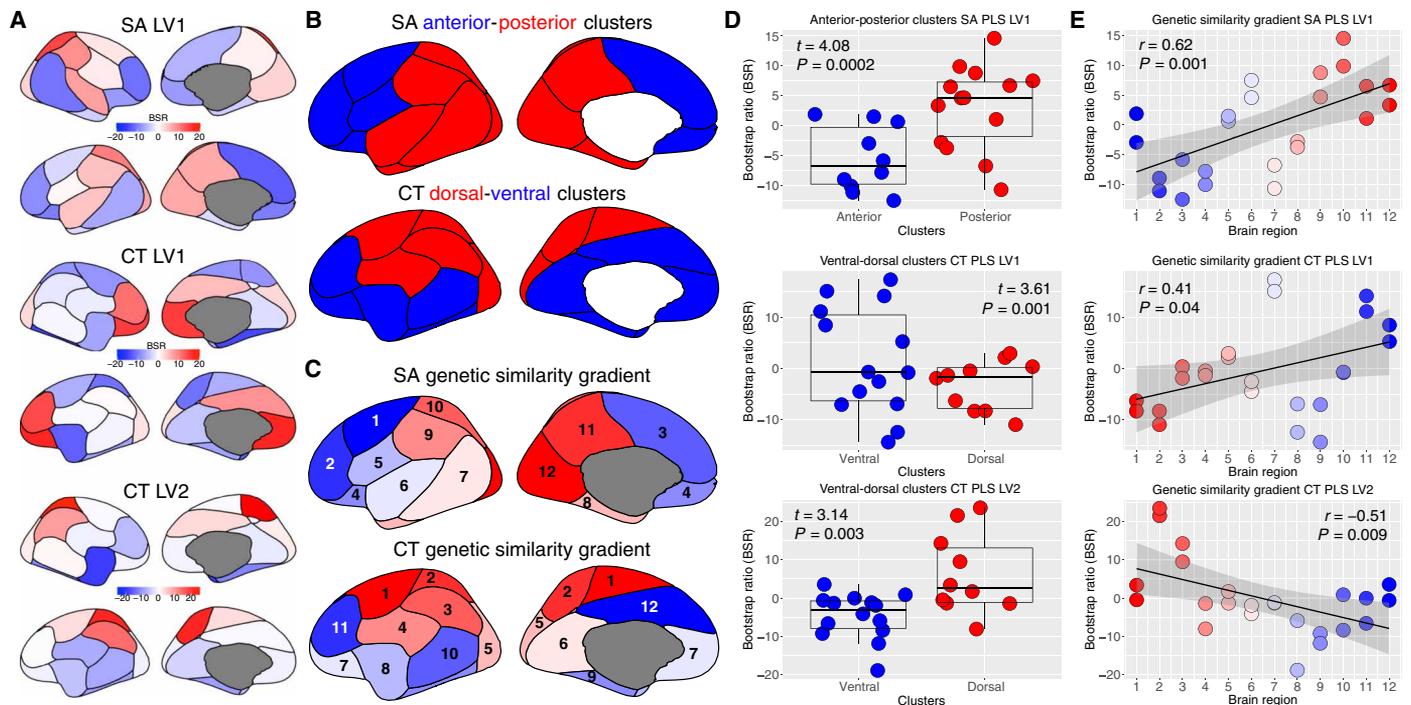
### Genes expressed in prenatal progenitor cell types explain SA associations, while genes expressed in later differentiated excitatory neurons explain CT associations

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 (15) 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 (12, 15). Therefore, progenitor cells are the primary cell types expected to explain SA effects. Programmed cell death could also be another mechanism regulating SA (12) 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 (15) and intermediate progenitor (IP) cell types (12). CT is also heavily influenced by dendritic arborization (28). While arborization changes over development due to a variety of factors such as experience-dependent pruning, CT and the trajectory it follows over development are 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 (31). 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 (32), we next asked whether specific prenatal cell

type markers are enriched for genes from SA and CT non-zero modules. In notable agreement with the prediction that progenitor cells explain SA effects (12, 15), we find that SA LV1 non-zero modules show enrichments for all progenitor cell types—ventricular and outer radial glia (vRG and oRG), cycling progenitors in S and G<sub>2</sub>M phases of cell cycle (PgS and PgG<sub>2</sub>M), and IPs. Several non-neuronal cells also show SA LV1 enrichments, including oligodendrocyte precursors (OPCs), endothelial cells (End), and microglia (Mic) (Fig. 5E and table S6). In contrast to these cell type enrichments, there is little evidence of enrichment of genes specific to later differentiated excitatory [maturing excitatory (ExM), migrating excitatory (ExN), maturing excitatory upper enriched (ExM-U), excitatory deep layer 1 (ExDp1), and excitatory deep layer 2 (ExDp2)] and inhibitory [interneuron caudal ganglion eminence (InCGE) and interneuron medial ganglion eminence (InMGE)] neurons.

For CT LV1 and LV2, we identify enrichments for vRG and IP progenitor cell types that are compatible with the hypothesized effects of IP cells on CT (12). 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. 5, F and G, and table S6). This result indicates a notable contrast between the SA LV1 enrichment profile of primarily progenitor cell types and is compatible with the radial unit and protomap hypotheses (15), differential SA and CT GWAS (genome wide association studies) enrichments (33), and other viewpoints regarding contributors to CT (28). These results also highlight the 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 (34) and may be relevant to ideas behind ASD-relevant broadly expressed genes and their particularly strong effects on non-neuronal cell types such as microglia (11).



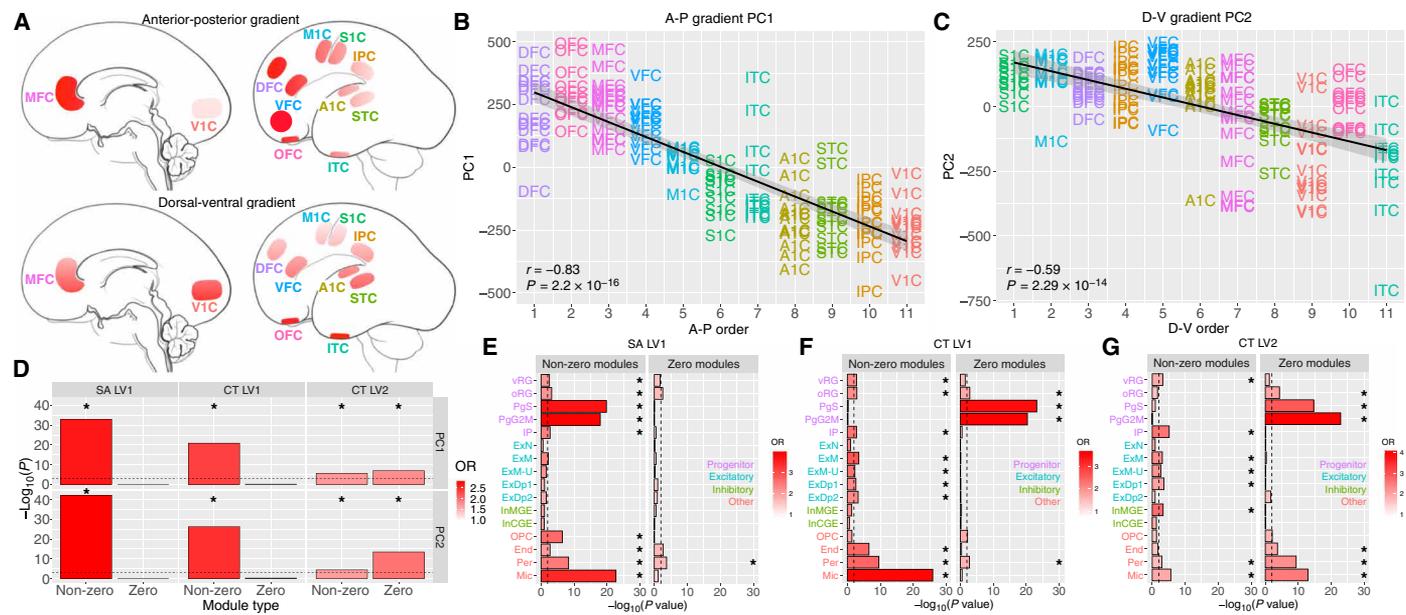
**Fig. 4. Cortical patterning along genetic similarity gradients.** (A) Brain BSR maps for SALV1 (top), CT LV1 (middle), and CT LV2 (bottom). (B) Coarse two-cluster A-P and D-V genetic similarity partitions identified by Chen and colleagues (13, 14). (C) Rank ordering of regions by hierarchical genetic similarity found by Chen and colleagues (13, 14). 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 similar than regions numbered farther apart. (D) BSRs are differentiated along A-P and D-V axes (SA LV1, top; CT LV1, middle; CT LV2, bottom). (E) BSRs co-vary with genetic similarity gradients (SA LV1, top; CT LV1, middle; CT LV2, bottom).

### 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 (8) were highly enriched for differentially expressed (DE) genes in Area X from a songbird model of vocal learning (35). To test whether similar enrichments held up for SA and CT non-zero modules, we ran enrichment tests with vocal learning DE genes from Hilliard and colleagues (35). We find similar types of enrichments between DE songbird vocal learning genes and PLS non-zero modules in SA LV1 ( $OR = 2.02$ ,  $P = 1.05 \times 10^{-4}$ ) and CT LV1 ( $OR = 1.90$ ,  $P = 9.61 \times 10^{-4}$ ) but not zero modules ( $P > 0.08$ ) (Fig. 6, A to C, and table S6). For CT LV2, enrichments were present at false discovery rate (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 SA 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 (36, 37). In prior work, we found that PLS non-zero modules associated with speech-related fMRI response (8) were enriched for DE genes in the cortex of humans versus nonhuman primates (i.e., “human-specific” genes). Given that cortical SA is a phenotype that is markedly 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 three lists of human DE genes in prenatal, early postnatal, and adulthood periods (36), we find that SA LV1 non-zero modules are specifically enriched for prenatal and adulthood human-specific genes (prenatal  $OR = 1.86$ ,  $P = 1.93 \times 10^{-3}$ ; adulthood  $OR = 1.97$ ,  $P = 1.02 \times 10^{-5}$ ) (Fig. 6D and table S6). In contrast, no such enrichments are found with genes relevant to CT LV1 or LV2 (Fig. 6, E and F). In addition to DE genes, we also examined genes that are targets of human-accelerated regions (HARs) or human-gained (HGE) or human-lost enhancer (HLE) regions (37). However, no enrichments for SA or CT were identified for HAR, HGE, and HLE genes (Fig. 6, D to 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



**Fig. 5. Enrichment between PLS non-zero modules and genes involved in prenatal A-P and D-V expression gradients and prenatal cell types.** (A) Cortical brain areas sampled from 12 to 24 weeks after conception from the Development PsychENCODE RNA-seq dataset from Li and colleagues (17). Adjustment for-confounds PCA (30) was used to isolate (B) A-P (PC1) and (C) D-V (PC2) expression gradients. (D)  $-\log_{10} P$  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. (E to G) Enrichments in prenatal cell types for SA LV1 (E), CT LV1 (F), and CT LV2 (G). Asterisks mark enrichments at FDR  $q < 0.01$ . MFC, medial prefrontal cortex; V1C, primary visual cortex; OFC, orbitofrontal cortex; DFC, dorsolateral refrontal cortex; M1C, primary motor cortex; VFC, ventrolateral prefrontal cortex; A1C, primary auditory cortex; ITC, inferior temporal cortex; STC, superior temporal cortex; IPC, posterior inferior parietal cortex; S1C, primary somatosensory cortex.

relationship is absent in ASD Poor, this suggests that the loss of these 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 (dnPTVs) (38) or other genes that are annotated as autism-associated in SFARI Gene (39). 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.

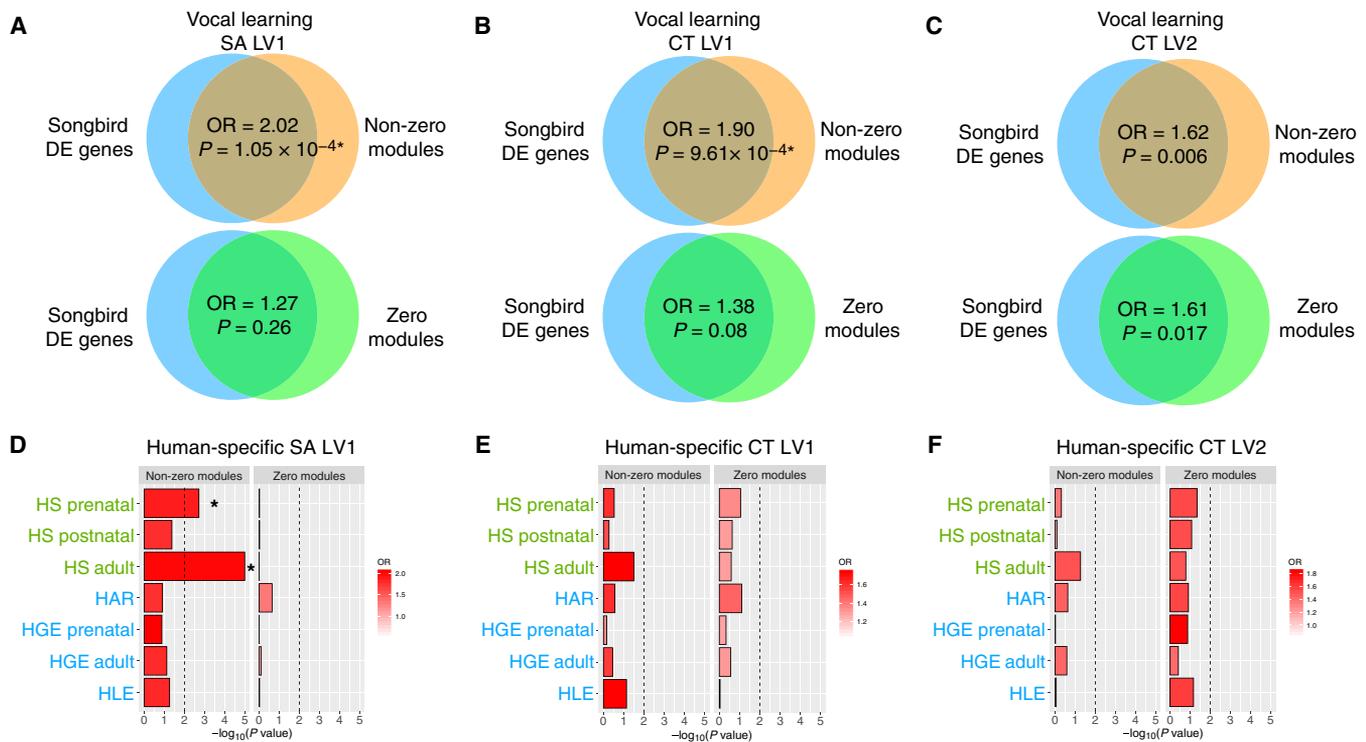
At the level of genes with evidence of ASD-dysregulated expression from postmortem cortical tissue, we find that both CT LV1 and LV2 non-zero modules were enriched for ASD up-regulated genes (40). In contrast, genes from cortically down-regulated coexpression modules (23) were highly enriched with genes from SA LV1 non-zero modules (table S6). This result shows an interesting contrast between CT and genes that show up-regulated expression versus SA and genes that show down-regulated expression in ASD.

Non-zero modules from SA LV1, CT LV1, and CT LV2 are also enriched for coexpression modules that are highly transcriptionally active during prenatal periods and that contain many high-penetrance ASD-related mutations (Fig. 7, A to C, and table S6). This is compatible with the idea that broadly expressed genes can interact and affect key ASD risk genes, particularly in prenatal periods (10, 11). Downstream targets of highly penetrant genes such as *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 (40) were also significantly enriched in non-zero modules across SA LV1, CT LV1, and CT LV2.

Last, we examined enrichments with cell type-specific DE genes in autism (41). Here, we found that only SA LV1 non-zero modules are enriched for DE 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 postmortem 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-relevant 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 coexpression modules). *SON* and *BAZZ2B* 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 coexpression 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



**Fig. 6. Enrichments between PLS non-zero modules and songbird vocal learning or human-specific genes.** (A to C) Enrichments between DE songbird vocal learning genes and non-zero and zero modules for SA LV1 (A), CT LV1 (B), and CT LV2 (C). (D to F) Enrichments between human-specific genes and non-zero and zero modules for SA LV1 (D), CT LV1 (E), and CT LV2 (F). Asterisks mark enrichments at  $FDR q < 0.01$ . HS, human-specific.

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. In addition, 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 (10).

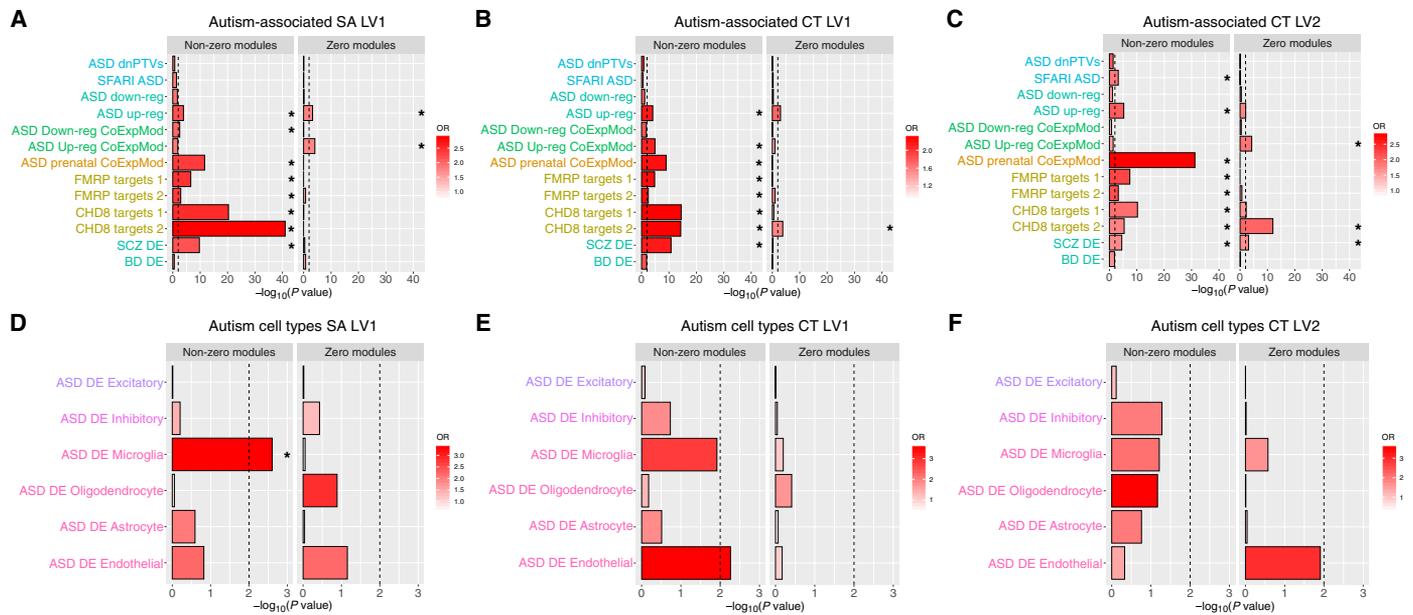
## DISCUSSION

These findings represent a substantial 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 (7, 8) 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 affect 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 (12, 15–17). 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 (12). These processes are explained by genetic variation and manifest after birth as patterned effects in CT and SA phenotypes measured with structural MRI data (13, 14). 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 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 (42, 43), 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. In addition, 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



**Fig. 7. Enrichment between PLS non-zero modules and autism-associated genes.** (A to C) Enrichments between different autism-associated gene lists and non-zero and zero modules for SA LV1 (A), CT LV1 (B), and CT LV2 (C). (D to F) Enrichments between DE 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 mark enrichments at FDR  $q < 0.01$ . SCZ DE, DE genes in schizophrenia; BD DE, DE genes in bipolar disorder.

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 patients with ASD. 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 (7). 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 (7), 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 have enough concurrent blood samples and MRI data from enough LD/DD participants for a sufficient comparison group. Future work should attempt to collect these data as a further comparison to ASD Poor to better understand whether 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 such as 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 (9). 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 (10). 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 (8).

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 found 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 these individuals.

## MATERIALS AND METHODS

### Participants

This study was approved by the Institutional Review Board at the 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 (7, 8, 26), toddlers were recruited through two mechanisms: community referrals (e.g., website) or a general population-based screening method called Get SET Early (44) that allowed for the prospective study of ASD beginning at 12 months based on a toddler's failure of the CSBS-DP (Communication and Symbolic Behavior Scales

Developmental Profile) Infant-Toddler Checklist (45, 46). All toddlers were tracked from an intake assessment around 12 months and followed roughly every 12 months until 3 to 4 years of age. All toddlers, including normal control participants, participated in a series of tests collected longitudinally across all visits, including the Autism Diagnostic Observation Schedule (Module T, 1, or 2) (47), the Mullen Scales of Early Learning (48), and the Vineland Adaptive Behavior Scales (49). All testing occurred at the University of California, San Diego Autism Center of Excellence.

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 SD 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 SD 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 two language outcome subtypes—ASD Poor,  $n = 38$  (32 males and 6 females; mean age at MRI scan = 29.01 months, SD at MRI scan = 7.22, range = 12 to 50 months); ASD Good,  $n = 38$  (28 males and 10 females; mean age at MRI scan = 29.02 months, SD at MRI scan = 9.55, range = 14 to 46 months); and TD,  $n = 47$  (25 males and 22 females; mean age at MRI scan = 25.91 months, SD at MRI scan = 10.44, range = 13 to 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 sample preparation

Four to six milliliters of blood were 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 before blood draw. Blood samples were passed over a LeukoLOCK filter (Ambion, Austin, TX, USA) to capture and stabilize leukocytes and immediately placed in a  $-20^{\circ}\text{C}$  freezer. Given the role of the immune system in autism (50) as well as interactions between the brain and the immune system (51), immune cells in blood such as leukocytes were specifically examined. This choice also allows for constraint on the cell types for which RNA might arise from in blood because whole blood is a bulk sample and RNA could potentially come from many different cell types (e.g., leukocytes and platelets). Total RNA was extracted following standard procedures and the manufacturer's instructions (Ambion, Austin, TX, USA). LeukoLOCK disks (Ambion, catalog no. 1933) were freed from RNAlater, and TRI Reagent (Ambion, catalog no. 9738) was used to flush out the captured lymphocyte and lyse the cells. RNA was subsequently precipitated with ethanol and purified through 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 (52), and all processed RNA samples passed RIN quality control. Quantification of RNA was performed using NanoDrop (Thermo Fisher Scientific, Wilmington, DE, USA). Samples were prepped in 96-well plates at the concentration of 25 ng/ $\mu\text{l}$ .

### Gene expression and data processing

RNA was assayed at Scripps Genomic Medicine (La Jolla, CA, USA) for labeling, hybridization, and scanning with 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 were exported from Illumina GenomeStudio, and data preprocessing was performed using the lumi package (53) for R (www.R-project.org) and Bioconductor (www.bioconductor.org) (54). Raw and normalized data are part of larger sets deposited in the Gene Expression Omnibus (GEO) 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 2 SD lower than the overall mean), deviant pairwise correlations, deviant cumulative distributions, deviant multidimensional scaling plots, or poor hierarchical clustering, as described elsewhere (55). 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 participants 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. These batch-, sex-, and RIN-adjusted data were used 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, natural killer cells, and monocytes) using the CellCODE deconvolution method (56) but found no evidence of differences across groups for any cell type (see table S7).

### Weighted gene coexpression network analysis

We reduced the number of features in the gene expression dataset from 14,426 genes down to 21 modules of tightly coexpressed genes. This data reduction step was achieved using WGCNA, implemented within the WGCNA library in R (57). 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 was computed as the first PC 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 to check for preservation of detected modules across groups at a soft power threshold of 16. These analyses all indicated high levels of preservation ( $Z_{\text{summary}} > 10$ ) (58) for all detected modules for each pairwise group comparison (fig. S6).

### MRI data acquisition and analyses

Imaging data were collected on a 1.5-T General Electric MRI scanner during natural sleep at night; no sedation was used. Structural MRI data were collected with a T1-weighted IR-FSPGR (inversion recovery fast-spoiled prepared gradient recalled) sagittal protocol [TE (echo time) = 2.8 ms, TR (repetition time) = 6.5 ms, flip angle = 12°, bandwidth = 31.25 kHz, field of view = 24 cm, and slice thickness = 1.2 mm]. Cortical surface reconstruction was performed using FreeSurfer v5.3 (<http://surfer.nmr.mgh.harvard.edu/>) (59–61), which uses routinely acquired T1-weighted MRI volumes (62), includes tools for estimation of brain morphometry measures such as CT and SA (63, 64), and enables interparticipant alignment via non-linear, surface-based registration to an average brain, driven by cortical folding patterns (65). FreeSurfer has been validated for use in children (66) and used successfully in large pediatric studies (67, 68). Total CV, SA, and mean CT were computed on the basis of the Desikan-Killiany parcellation. Regional SA and CT values were computed from a 12-region parcellation reported by Chen and colleagues (13, 14) based on genetic similarity in monozygotic twins. This parcellation scheme, known as GCLUST, is highly relevant for our purposes here because the parcellations are based on genetic patterning and has also been effectively used in developmental samples (31). 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 two-cluster A-P or D-V partitions found by Chen and colleagues (13, 14) 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 the mean CT from each region, as was done in prior papers using this parcellation scheme (13, 14).

### MRI-gene expression association analysis

To assess multivariate MRI-gene expression relationships, we used PLS analysis (69). 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 coexpression modules. PLS allows for identifying these relationships by finding latent MRI-gene expression variable pairs (LV) that maximally explain covariation in the

dataset and that 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 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/](http://www.rotman-baycrest.on.ca/pls/)). Here, we ran two 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. These SA and CT data with the sex effect removed were input into the PLS analysis. For gene expression data, we input module eigengene values for all 21 coexpression 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% CIs on MRI-gene expression correlations, bootstrapping was used with 10,000 resamples. Gene coexpression 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. In addition, we ran 10,000 split-half resamples whereby the correlation between brain and gene expression saliences ( $U_{\text{corr}}$  and  $D_{\text{corr}}$ ) was computed between the 2 split-halves. These correlations between split-half saliences were then compared to the null distribution from 10,000 permutations to compute  $P$  values ( $P_{U_{\text{corr}}}$  and  $P_{D_{\text{corr}}}$ ), which statistically test the reliability of salience patterns in split-half resamples (70).

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 BSRs from the PLS analysis to identify whether BSRs covary along the genetic similarity gradients and A-P and D-V partitions found by Chen and colleagues (13, 14). Pearson correlations were used to identify correlations with genetic similarity gradients, while independent-sample  $t$  tests were used to compare A-P and D-V partitions.

All PLS analyses were computed on vertex-wise data and GCLUST-parcellated data. This analysis used the same parameters (10,000 permutations and 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 coexpression module PLS correlations between GCLUST and this vertex-wise analysis was also computed as Pearson correlations from the PLS correlation values for each module and group. Last, 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 figs. S1 to 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 used a custom R code written by M.V.L. (<https://github.com/mvlombardo/utills/blob/master/genelistOverlap.R>) that computes hypergeometric *P* values and enrichment ORs. 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 used RNA-seq data from the Development PsychENCODE dataset (<http://development.psychencode.org>) (17). The data used were already preprocessed as described by Li and colleagues (17) (e.g., normalized and batch effects removed) and summarized to RPKM (reads per kilobase million). Sample data from all 12 available cortical regions from 12 to 22 weeks after conception were used to capture the midgestational window of interest. Before running the analysis, we removed low-expressing genes with  $\log_2(\text{RPKM})$  below 2. The primary analysis to identify expression gradients was an adjustment-for-confounds PCA (AC-PCA) (30), 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 was used 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 PCs were identified with a sparse AC-PCA analysis, whereby the sparsity parameter,  $c_2$ , was selected on the basis of 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 (32). These cell types included several classes of progenitor cells (vRG, oRG, PgS, PgG2M, and IPs), excitatory neurons (ExN, ExM, ExM-U, ExDp1, and ExDp2), inhibitory neurons (InCGE and InMGE), and other non-neuronal cell types (OPCs, Per, End, and Mic).

### Tissue-specific enrichments

To better understand how genes expressed in blood leukocytes could be brain relevant, we annotated gene coexpression 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 ([www.gtexportal.org](http://www.gtexportal.org)) (71). In addition to brain and nerve tissue, the dataset included transcriptome data from 24 non-neuronal tissues, including the following: 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  $\geq 3$  in at least half of the samples from the tissue. Second, the median expression of the gene was equal or larger than its 25th 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 the 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 songbird vocal learning model. Songbirds are often used as animal models relevant for the vocal learning component of language (72). We investigated enrichments with DE genes taken from a microarray dataset of Area X of songbirds (35). To identify DE genes between singing and nonsinging birds, we reanalyzed this dataset (GEO accession ID: GSE34819) using limma (73), and DE genes were identified if they passed Storey FDR  $q < 0.05$  (74). 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 (8).

### 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 nonhuman primates across prenatal, early postnatal, and adult periods (36). In addition, we also examined enrichments with genes linked to HARs, HGEs in prenatal and adult tissue, and HLEs (37).

### Autism-associated enrichments

Ample evidence suggests that prenatal periods are critical for ASD (4, 10). To test enrichment with prenatal ASD-associated coexpression modules, we used coexpression modules from a study that analyzed the Allen Institute BrainSpan dataset (75). Parikshak and colleagues (76) 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. 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 dnPTVs associated with ASD (38); genes listed as ASD-associated in SFARI Gene (<https://gene.sfari.org>) in categories S, 1, 2, and 3 (downloaded on 16 July 2020) (39); and DE genes and cortical coexpression modules measured from ASD postmortem frontal and temporal cortex tissue (23, 40). To contrast ASD DE genes to genes that are DE in other psychiatric diagnoses that are genetically correlated with autism, we also used DE genes in schizophrenia and bipolar disorder from the same study that identified ASD DE genes (40). To go beyond DE genes identified in bulk tissue samples, we also examined ASD DE genes identified in specific cell types, particularly excitatory (ASD DE Excitatory) and inhibitory (ASD DE Inhibitory) neurons, microglia (ASD DE Microglia), astrocytes (ASD DE Astrocyte), oligodendrocytes (ASD DE Oligodendrocyte), and endothelial (ASD DE Endothelial)

cells (41). Last, 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 (77–80), where the overlap for FMRP targets was 3.71 and 27.61% for CHD8 targets.

## SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abh1663>

[View/request a protocol for this paper from Bio-protocol.](#)

## REFERENCES AND NOTES

- 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).
- 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).
- F. S. Collins, H. Varmus, A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
- 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).
- 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).
- 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, J. C. McPartland, M. R. Natowicz, Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics* **136** (Suppl. 1), S60–S81 (2015).
- 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).
- M. V. Lombardo, T. Pramparo, V. Gazestani, V. Warrior, R. A. I. Bethlehem, C. Carter Barnes, L. Lopez, N. E. Lewis, L. Eyler, K. Pierce, E. Courchesne, Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. *Nat. Neurosci.* **21**, 1680–1688 (2018).
- E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
- E. Courchesne, V. H. Gazestani, N. E. Lewis, Prenatal origins of ASD: The when, what, and how of ASD development. *Trends Neurosci.* **43**, 326–342 (2020).
- V. H. Gazestani, A. W. T. Chiang, E. Courchesne, N. E. Lewis, Autism genetics perturb prenatal neurodevelopment through a hierarchy of broadly-expressed and brain-specific genes. *bioRxiv* 2020.05.23.112623 (2020).
- C. R. Cadwell, A. Bhaduri, M. A. Mostajo-Radji, M. G. Keefe, T. J. Nowakowski, Development and arealization of the cerebral cortex. *Neuron* **103**, 980–1004 (2019).
- C.-H. Chen, E. D. Gutierrez, W. Thompson, M. S. Panizzon, T. L. Jernigan, L. T. Eyler, C. Fennema-Notestine, A. J. Jak, M. C. Neale, C. E. Franz, M. J. Lyons, M. D. Grant, B. Fischl, L. J. Seidman, M. T. Tsuang, W. S. Kremen, A. M. Dale, Hierarchical genetic organization of human cortical surface area. *Science* **335**, 1634–1636 (2012).
- C.-H. Chen, M. Fiecas, E. D. Gutiérrez, M. S. Panizzon, L. T. Eyler, E. Vuoksimaa, W. K. Thompson, C. Fennema-Notestine, D. J. Hagler Jr., T. L. Jernigan, M. C. Neale, C. E. Franz, M. J. Lyons, B. Fischl, M. T. Tsuang, A. M. Dale, W. S. Kremen, Genetic topography of brain morphology. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17089–17094 (2013).
- P. Rakic, Specification of cerebral cortical areas. *Science* **241**, 170–176 (1988).
- D. D. M. O’Leary, S.-J. Chou, S. Sahara, Area patterning of the mammalian cortex. *Neuron* **56**, 252–269 (2007).
- M. Li, G. Santpere, Y. I. 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. Kitchin, 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**, eaat7615 (2018).
- 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).
- R. Gao, R. L. van den Brink, T. Pfeffer, B. Voytek, Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. *eLife* **9**, e61277 (2020).
- 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, J. M. 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).
- S.-J. Hong, R. Vos de Wael, R. A. I. Bethlehem, S. Larivière, C. Paquola, S. L. Valk, M. P. Millham, A. Di Martino, D. S. Margulies, J. Smallwood, B. C. Bernhardt, Atypical functional connectome hierarchy in autism. *Nat. Commun.* **10**, 1022 (2019).
- I. Voineagu, X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, J. Mill, R. M. Cantor, B. J. Blencowe, D. H. Geschwind, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
- N. N. Parikshak, V. Swarup, T. G. Belgard, M. Irimia, G. Ramaswami, M. J. Gandal, C. Hartl, V. Leppä, L. dela T. Ubieta, J. Huang, J. K. Lowe, B. J. Blencowe, S. Horvath, D. H. Geschwind, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
- M. L. Chow, T. Pramparo, M. E. Winn, C. C. Barnes, H.-R. Li, L. Weiss, J.-B. Fan, S. Murray, C. April, H. Belinson, X.-D. Fu, A. Wynshaw-Boris, N. J. Schork, E. Courchesne, Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLOS Genet.* **8**, e1002592 (2012).
- J. L. Freese, D. Pino, S. J. Pleasure, Wnt signaling in development and disease. *Neurobiol. Dis.* **38**, 148–153 (2010).
- V. H. Gazestani, T. Pramparo, S. Nalabolu, B. P. Kellman, S. Murray, L. Lopez, K. Pierce, E. Courchesne, N. E. Lewis, A perturbed gene network containing RAS–ERK and WNT– $\beta$ -catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nat. Neurosci.* **22**, 1624–1634 (2019).
- T. N. Turner, K. Sharma, E. C. Oh, Y. P. Liu, R. L. Collins, M. X. Sosa, D. R. Auer, H. Brand, S. J. Sanders, D. Moreno-De-Luca, V. Pihur, T. Plona, K. Pike, D. R. Soppet, M. W. Smith, S. W. Cheung, C. L. Martin, M. W. State, M. E. Talkowski, E. Cook, R. Hugarin, N. Katsanis, A. Chakravarti, Loss of  $\delta$ -catenin function in severe autism. *Nature* **520**, 51–56 (2015).
- K. Wagstyl, J. P. Lerch, in *Brain Morphometry*, G. Spalletta, F. Piras, T. Gili, Eds. (Springer New York, 2018), vol. 136 of *Neuroinformatics*, pp. 35–49.
- 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).
- 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).
- 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. Solsnes, 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. U.S.A.* **112**, 15462–15467 (2015).
- 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. Leung, T. Hadzic, F. I. Hinz, S. Sabri, W. E. Lowry, M. D. Gerstein, K. Plath, D. H. Geschwind, A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* **103**, 785–801.e8 (2019).
- K. L. Grasby, N. Jahanshad, J. N. Painter, L. Colodro-Conde, J. Bralten, D. P. Hibar, P. A. Lind, F. Pizzagalli, C. R. K. Ching, M. A. B. McMahon, N. Shatohkina, L. C. P. Zsembik, S. I. Thomopoulos, A. H. Zhu, L. T. Strike, I. Agartz, S. Alhusaini, M. A. A. Almeida, D. Alnæs, I. K. Amlien, M. Andersson, T. Ard, N. J. Armstrong, A. Ashley-Koch, J. R. Atkins, M. Bernard, R. M. Brouwer, E. E. L. Buimer, R. Bülow, C. Bürger, D. M. Cannon, M. Chakravarty, Q. Chen, J. W. Cheung, B. Couvy-Duchesne, A. M. Dale, S. Dalvie, T. K. de Araujo, G. I. de Zubicaray, S. M. C. de Zwarte, A. den Braber, N. T. Doan, K. Dohm, S. Ehrlich, H.-R. Engelbrecht, S. Erk, C. C. Fan, I. O. Fedko, S. F. Foley, J. M. Ford, M. Fukunaga, M. E. Garrett, T. Ge, S. Giddaluru, A. L. Goldman, M. J. Green, N. A. Groenewold, D. Grotegerd, T. P. Gurholt, B. A. Gutman, N. K. Hansell, M. A. Harris, M. B. Harrison, C. C. Haswell, M. Hauser, S. Herms, D. J. Heslenfeld, N. F. Ho, D. Hoehn, P. Hoffmann, L. Holleran, M. Hoogman, J.-J. Hottenga, M. Ikeda, D. Janowitz, I. E. Jansen, T. Jia, C. Jockwitz, R. Kanai, S. Karama, D. Kasperavičiute, T. Kaufman, S. Kelly, M. Kikuchi, M. Klein, M. Knapp, A. R. Knodt, B. Krämer, M. Lam, T. M. Lancaster, P. H. Lee, T. A. Lett, L. B. Lewis, I. Lopes-Cendes, M. Luciano, F. Maciardi, 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. O. 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. Sonderby, E. Sprooten, K. E. Tansey, A. Teumer, A. Thalathuthu, D. Tordesillas-Gutiérrez, J. A. Turner, A. Uhlmann, C. L. Valleria, 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. Garjjo, 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. Kimgrel, K. Kolskår, S. Koops, A. Krug, K. O. Lim, J. J. Luykx, D. H. Mathalon, K. A. Mather, V. S. Mattay, S. Matthews, J. M. 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. Desrivieres, G. Donohoe, T. Espeseth, G. Fernández, S. E. Fisher, H. Flor, A. J. Forstner, C. Francks, B. Franke, D. C. Glahn, R. L. Gollub, J. H. Grabe, O. Gruber, A. K. Häberg, 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. H. Pol, T. L. Jernigan, S. E. Jönsson, R. S. Kahn, M. A. Kennedy, 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. S. 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**, eaay6690 (2020).
34. G. Tang, K. Gudsruk, 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).
35. 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).
36. 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**, eaat8077 (2018).
37. 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).
38. 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. D. Bernardino, R. Doan, E. Domenici, S. Dong, C. Fallarini, 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. Loperogolo, F. Lopes, Y. Ludena, P. Maciel, P. Magnus, B. Mahjani, N. Maltman, D. S. Manocha, 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. D. 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. Borglum, 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* **180**, 568–584.e23 (2020).
39. 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).
40. M. J. Gandal, P. Zhang, E. Hadjimihael, 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. M. 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).
41. 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).
42. 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).
43. 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).
44. 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.* In Press (2021).
45. A. Wetherby, B. Prizant, *Communication and Symbolic Behavior Scales Developmental Profile, First Normed Edition* (Paul H. Brookes, 2002).
46. 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).
47. 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).
48. E. M. Mullen, *Mullen Scales of Early Learning* (American Guidance Service Inc, 1995).
49. S. Sparrow, D. Cicchetti, D. Balla, *Vineland-II Scales of Adaptive Behavior: Survey form Manual* (American Guidance Service Inc, 2005).
50. C. Onore, M. Careaga, P. Ashwood, The role of immune dysfunction in the pathophysiology of autism. *Brain Behav. Immun.* **26**, 383–392 (2012).
51. 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).
52. 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).
53. P. Du, W. A. Kibbe, S. M. Lin, lumi: A pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547–1548 (2008).
54. 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).
55. 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).
56. 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).

57. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
58. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
59. A. M. Dale, B. Fischl, M. I. Sereno, Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* **9**, 179–194 (1999).
60. 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).
61. F. Ségonne, J. Pacheco, B. Fischl, Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. *IEEE Trans. Med. Imaging* **26**, 518–529 (2007).
62. 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–S84 (2004).
63. 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).
64. 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).
65. 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).
66. 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).
67. 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).
68. 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).
69. 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).
70. N. Kovacevic, H. Abdi, D. Beaton, A. R. McIntosh, in *New Perspectives in Partial Least Squares and Related Methods* (Springer), pp. 159–170 (2013).
71. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature*. **550**, 204–213 (2017).
72. G. Konopka, T. F. Roberts, Insights into the neural and genetic basis of vocal communication. *Cell* **164**, 1269–1276 (2016).
73. 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).
74. J. D. Storey, A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat Methodol.* **64**, 479–498 (2002).
75. 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. Cuhacyan, 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. Nalua-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. Sodd, 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).
76. 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).
77. 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).
78. 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. U.S.A.* **111**, E4468–E4477 (2014).
79. 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).
80. 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).

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

**Funding:** This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement no. 755816 (ERC Starting Grant to M.V.L.). This work was also supported by the following grants: NIMH R01-MH080134 (to K.P.), NIMH R01-MH104446 (to K.P.), NFAR grant (to K.P.), NIMH Autism Center of Excellence grant P50-MH081755 (to E.C. and K.P.), NIMH R01-MH036840 (to E.C.), NIMH R01-MH110558 (to E.C. and N.E.L.), NIMH U01-MH108898 (to E.C.), NIDCD R01-DC016385 (to E.C., K.P., L.E., and M.V.L.), NIMH R01-MH118281 (to C.-H.C.), CDMRP AR130409 (to E.C.), and the Simons Foundation 176540 (to E.C.). K.C. was supported by the Utah Stimulating Access to Research in Residency Transition Scholar (STARRTS) under award number 1R38HL143605-01. **Author contributions:** Conceptualization: M.V.L., E.C., K.P., L.E., and T.P. Methodology: M.V.L., T.P., V.H.G., C.-H.C., N.E.L., E.C., K.P., and L.E. Software: M.V.L., D.J.H., A.M.D., and V.H.G. Formal analysis: M.V.L., V.H.G., D.J.H., T.P., J.S., R.A.I.B., and N.B. Investigation: L.E., K.C., C.C.B., L.L., K.P., and E.C. Data curation: E.C., K.P., L.E., T.P., V.H.G., and M.V.L. Writing—original draft preparation: M.V.L. and E.C. Writing—review and editing: M.V.L., E.C., K.P., L.E., T.P., J.S., R.A.I.B., N.B., K.C., N.E.L., V.H.G., D.J.H., C.-H.C., and A.M.D. Visualization: M.V.L. Supervision: E.C., K.P., L.E., and N.E.L. Project administration: E.C., K.P., L.E., and N.E.L. Funding acquisition: M.V.L., E.C., K.P., L.E., and N.E.L. **Competing interests:** A.M.D. is a founder of and holds equity in CorTechs Labs Inc. and serves on its Scientific Advisory Board. A.M.D. is also a member of the Scientific Advisory Board of Human Longevity Inc. (HLI) and the Mohn Medical Imaging and Visualization Centre. A.M.D. receives funding through a research agreement with General Electric Healthcare (GEHC). The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. None of the other authors have any biomedical financial interests or potential conflicts of interest to report. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Analysis code is available at [https://github.com/IIT-LAND/genomic\\_cortical\\_patterning\\_autisms](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 at <http://development.psychencode.org>. GTEx data can be found at [www.gtexportal.org](http://www.gtexportal.org). Microarray data from the songbird vocal learning model can be found in Gene Expression Omnibus (GSE34819).

Submitted 19 February 2021

Accepted 15 July 2021

Published 3 September 2021

10.1126/sciadv.abh1663

**Citation:** M. V. Lombardo, L. Eyley, T. Pramparo, V. H. Gazestani, D. J. Hagler Jr., C.-H. Chen, A. M. Dale, J. Seidlitz, R. A. I. Bethlehem, N. Bertelsen, C. C. Barnes, L. Lopez, K. Campbell, N. E. Lewis, K. Pierce, E. Courchesne, Atypical genomic cortical patterning in autism with poor early language outcome. *Sci. Adv.* **7**, eabh1663 (2021).