Transcriptional organization of autism spectrum disorder and its connection to ASD risk genes and phenotypic variation

Vahid H. Gazestani[1,2,3], Tiziano Pramparo[1], Srinivasa Nalabolu[1], Benjamin P. Kellman[2,4], Sarah Murray[5], Linda Lopez[1], Karen Pierce[1], Eric Courchesne[1,*], Nathan E. Lewis[2,3,4,6,*]

[1] Autism Center of Excellence, Department of Neuroscience

[2] Department of Pediatrics

[3] Novo Nordisk Foundation Center for Biosustainability

[4] Bioinformatics and Systems Biology Program

[5] Department of Pathology

[6] Department of Bioengineering

University of California San Diego, La Jolla, CA 92093, USA

[*] Correspondence to Nathan E. Lewis (nlewisres@ucsd.edu) and Eric Courchesne (ecourchesne@ucsd.edu)

ABSTRACT

Hundreds of genes are implicated as risk factors for autism spectrum disorder (ASD). However, the mechanisms through which they are associated with ASD remain unclear. Here, we analyzed transcriptomics from ASD toddlers and discovered a core gene network with dysregulated gene co-expression. The identified network includes highly expressed processes in fetal-stage brain development and is dysregulated in neuron models of ASD. We found ASD risk genes across diverse functions are upstream and regulate this core network. In particular, many risk genes impact the network through the RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways. Finally, the dysregulation degree of this network positively correlates with early-age ASD clinical severity. Thus, our results provide insights into how the heterogeneous genetic basis of ASD could converge on a core network with consequence on the postnatal outcome of toddlers with ASD. Deeper study into this may help decipher the molecular basis of ASD and decode the complex link between its genetic and phenotypic variation.

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with prenatal and early postnatal biological onset [1-3]. Genetic factors contribute to the predisposition and development of ASD with estimated heritability rates of 50-83% [4,5]. Large-scale genetic studies have implicated several hundred risk (rASD) genes that could be associated with many different pathways, cell processes, and neurodevelopmental stages [6-8]. This highly heterogeneous genetic landscape has raised challenges in elucidating the biological mechanisms involved in the disorder. While rigorous proof remains lacking, current evidence suggests that rASD genes fall into networks and biological processes [6,7,9-13] that modulate one or more critical stages of prenatal and early postnatal brain development, including neuronal proliferation, migration, neurite growth, synapse formation and function [3,8]. However, these insights are mostly gained from focused studies on single rASD genes (see Courchesne et al.[3] for a recent review) or based on transcriptome data of non-ASD brains [9-11], leaving an incomplete picture of molecular changes at the individual level and relationships with early-age clinical heterogeneity.

To further complicate efforts to discern the molecular bases of ASD, the implicated rASD genes are largely identified through *de novo* loss-of-function mutations in their coding sequence. Such events account for 5-10% of the ASD population, and most of heritability is estimated to reside in common variants also seen in the typically developing population [5,14-17]. Currently, there is a paucity of data on whether ASD cases with known rASD gene mutations manifest as special subtypes of ASD with distinct molecular etiology, or whether they share mechanisms with the general ASD population.

To address these fundamental questions, it is important to understand which molecular processes are perturbed in prenatal and early postnatal life in ASD individuals, assess how they vary among subjects, and evaluate how these perturbations relate to rASD genes and early-age ASD clinical symptoms. It is expected

that the genetic changes in ASD alter gene expression and signaling in the early-age developing brain [3,7,11,18]. Therefore, capturing dysregulated gene expression at prenatal and early postnatal ages may help unravel the underlying molecular organization of ASD. Unfortunately, doing so is particularly challenging as ASD brain tissue cannot be obtained at these early stages, and all available postmortem ASD brains are from much older ages, well beyond the ages when rASD genes are at peak expression and the disorder begins. However, in contrast to living neurons that have a limited time window for proliferation and maturation, other cell types constantly regenerate, such as blood cells. Given the strong genetic basis of ASD, some dysregulated developmental signals may continually reoccur in blood cells and thus be studied postnatally[19-21].

Reinforcing this notion, it was recently demonstrated that genes that are broadly expressed across many tissues are major contributors to the overall heritability of complex traits [22], and it was postulated that this could be relevant to ASD. Lending credence to this, previous studies have reported the enrichment of differentially expressed genes in ASD blood for the regulatory targets of CHD8 [20] and FMR1 [23] genes, two well-known rASD genes. Similarly, lymphoblastoid cells of ASD cases and iPS-derived models of fragile-X syndrome show over-expression of mir-181 with a potential role in the disorder [24]. Likewise, leukocytes from ASD toddlers show perturbations in biological processes, such as cell proliferation, differentiation, and microtubules [25-29], and these coincide with dysregulated processes seen in neural progenitor cells (NPCs) and neurons, derived from iPS cells from ASD subjects [30,31]. Ultimately, establishing the signatures of ASD in other tissues will be important to facilitate the study of the molecular basis of the disorder in living ASD subjects in the first years of life.

Here we leverage transcriptomic data from leukocytes, stem cell models, and the developing brain to study the underlying architecture of transcriptional dysregulation in ASD, its connection to rASD genes, and its association with prenatal development and clinical outcomes of ASD toddlers. Specifically, we discovered a conserved dysregulated gene network by analyzing leukocyte transcriptomic data from 1-4 years old ASD and typically developing (TD) toddlers. The dysregulated network is enriched for pathways known to be perturbed in ASD neurons, impacts highly expressed processes in prenatal brain development, and is dysregulated in iPS cell-derived neurons from ASD cases. Consistent with the postulated structure of complex traits [22,32], we show that rASD genes across diverse functional groups converge upon and regulate this core network. Importantly, this core network is disrupted to different levels of severity across ASD individuals, and is correlated with clinical severity in individual ASD toddlers. Thus, our results demonstrate how the heterogeneous genetic basis of ASD converges on a biologically relevant core network, capturing the underlying possible molecular etiology of ASD.

RESULTS

Leukocytes display transcriptome over-activity in ASD male toddlers

To identify the unique transcriptional response of ASD subjects, we analyzed 253 leukocyte gene expression profiles obtained from 226 male toddlers (119 ASD and 107 TD, Table S1). Robust linear regression modeling of the data identified 1236 unique differentially expressed (DE) genes (437 downregulated and 799 upregulated; FDR < 0.05). Jack-knife resampling demonstrated that the expression pattern of DE genes was not driven by a small number of cases, but rather shared with the vast majority of ASD subjects (Fig S1). The expression patterns were validated in a replicate dataset of 56 randomly re-sampled toddlers. We further confirmed the expression patterns of DE genes on another partially independent and one entirely independent cohort (Fig S1-S4).

We employed a systems approach to decipher how the transcriptional perturbations in leukocytes of ASD toddlers are organized in gene networks (Fig 1.a). We reasoned that ASD associated interactome rewiring is most pronounced in networks of DE genes. To identify such rewiring, we first extracted a static network (that is, the network is indifferent to the cell context) composed of high confidence physical and regulatory interactions among DE genes, as obtained from multiple databases (Methods). We next pruned the static network using our leukocyte transcriptome data to obtain context-specific networks of each study group separately (that is, the networks differ based on their cognate gene expression data). The context specific network of each study group was obtained by only retaining interactions in the static network that were significantly co-expressed within that group with FDR <0.05. To ensure the robustness of our conclusions, we replicated all presented results on two other networks with different numbers of genes and interactions obtained from additional resources (Methods).

The context-specific networks (DE-ASD and DE-TD) include published physical and regulatory interactions among DE genes that exhibit within-group co-expression in our data. DE-ASD and DE-TD networks are composed of a similar set of genes (i.e., those expressed in the leukocytes that are differentially expressed between ASD and TD samples), but the wiring of the two networks differ based on the co-expression patterns within each study group. To assess the possibility that intracellular pathways were being specifically modulated in ASD, we created a merged network by considering the union of interactions in the DE-ASD and DE-TD networks. We next examined the co-expression strength of the merged network in ASD and TD individuals (Methods) [33-35]. This proxy for the transcriptional activity of gene networks [9] demonstrated that co-expression strength was higher in the ASD than the TD samples (Fig 1.b; p-value < 0.01; paired Wilcoxon-Mann-Whitney test). The stronger co-expression that is driven by the DE-ASD network, suggests a higher level of concerted activation or suppression of pathways involving DE genes in ASD toddlers. This elevated co-expression activity (herein referred to as over-activity) of the

network was reproducible in the other two ASD datasets and replicable across alternative analysis methods (Fig S1-S4).

In summary, the leukocyte transcriptional networks of the DE genes show higher than normal co-expression activity in ASD toddlers. Moreover, the dysregulation pattern is present in a large percentage of ASD toddlers, as evidenced by the resampling analyses and the other two ASD datasets.

The leukocyte-based gene network captures transcriptional programs of brain development

We next assessed the potential association of the leukocyte-based network to the spatiotemporal neurodevelopmental signals relevant to ASD. By overlaying our network on the *in vivo* neurodevelopmental RNA-Seq transcriptome data from BrainSpan[36,37], we found that the DE-ASD network was enriched for genes that are strongly expressed in the neocortex at prenatal and early postnatal periods (p-value $<4.3\times10^{-30}$; Fig 1.c).

To investigate the spatiotemporal activity pattern during brain development, we measured the co-expression strength of interactions in the leukocyte-based network at different neurodevelopmental time windows across brain regions using BrainSpan. We found that the highest co-expression activity of the DE-ASD network temporally coincided with peak neural proliferation in brain development (10-19 post conception weeks [3,8]) across the brain and then decreased in activity at later time points (Fig 1.d). Further supporting the transcriptional activity of the leukocyte-derived network in prenatal brain, we found evidence that the DE-ASD network is preserved at the co-expression level between ASD leukocytes and prenatal brain (Fig 1.e).

Networks of rASD genes are associated with the DE-ASD network

We next analyzed the DE-ASD network in the context of other studies to explore the relevance of our leukocyte-based signature to neocortical development. Parikshak et al. previously reported gene co-expression modules that are responsive to the developmental trajectories of cortical laminae during prenatal and early postnatal ages [10]. A subset of these modules show enrichment in rASD genes [10]. We examined the overlap of our leukocyte-derived network with all modules from Parikshak et al[10]. The DE-ASD network preferentially overlapped with rASD gene-enriched modules from that study (Fig 1.f; Table S2). This suggests that our DE-ASD network is functionally related to rASD genes during neocortical development. We confirmed the significant overlap of our DE-ASD network with the networks of rASD genes reported in two other studies [7,9], indicating the robustness of the results (Fig 1.f). Intriguingly, the prenatal brain co-expression network of high confidence rASD genes was more similar to that of ASD leukocytes than TD leukocytes (Fig 1.g), suggesting that neurodevelopmental transcriptional programs related to rASD genes might be more represented in the ASD leukocyte transcriptome than TD samples.

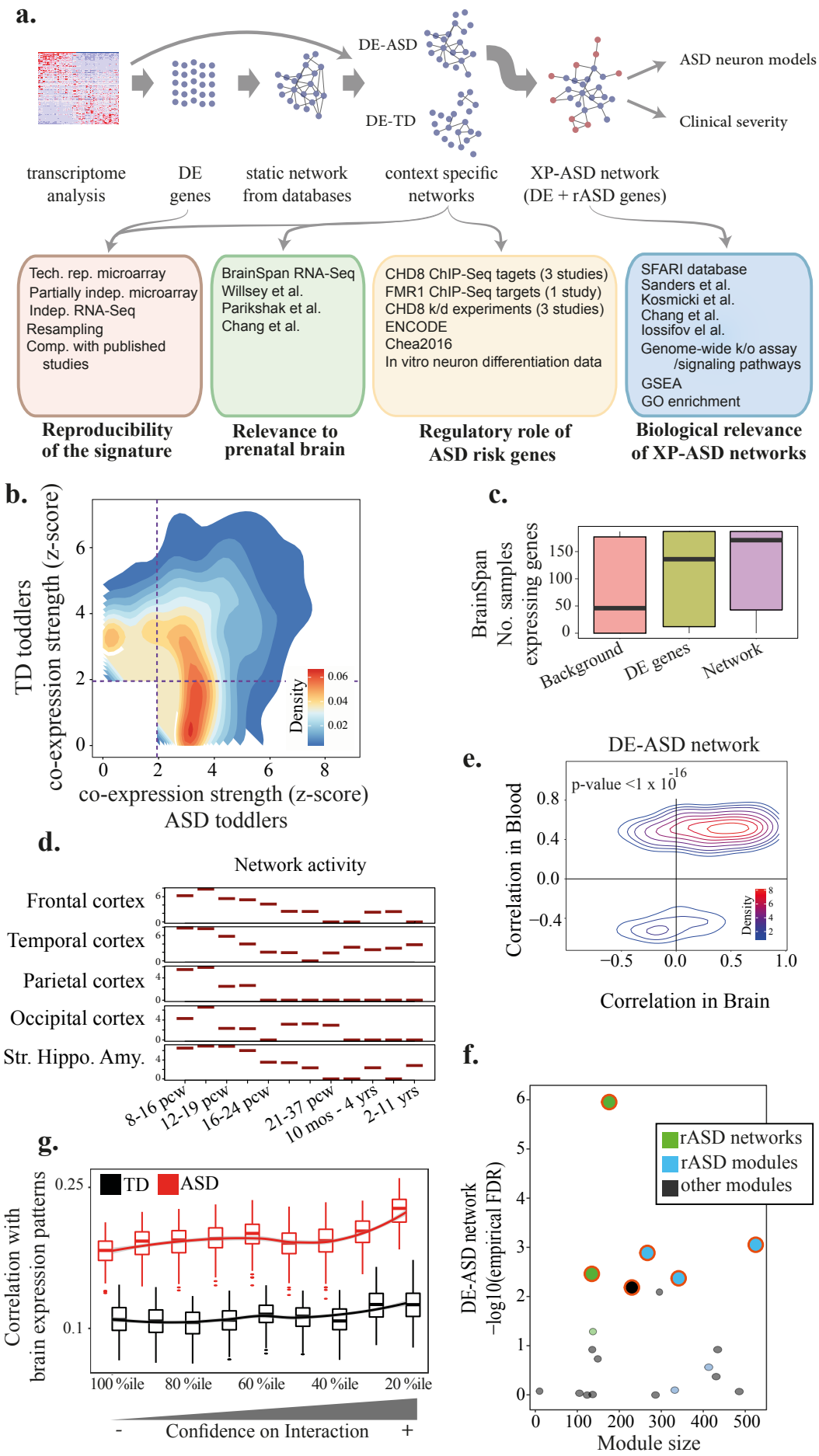

Fig 1. Elevated co-expression activity of the DE-ASD network in ASD leukocytes and its preservation in prenatal brain.
a) Overview of this study. Transcriptome analysis of 226 ASD and TD toddlers identified 1236 DE genes. We built a comprehensive "static" network of DE genes from high confidence physical and regulatory interactions. The static network was next pruned to only include links supported by measured gene co-expression within the ASD and TD transcriptome data. This yielded context specific DE-ASD and DE-TD networks, and allowed the comparison of gene co-expression strength in ASD and TD subjects. To understand the link of the DE-ASD network to ASD risk genes, an XP-ASD network was constructed using both DE and ASD risk genes. The DE-ASD and XP-ASD networks were analyzed in the context of neural differentiation, ASD neuron models and ASD clinical severity. b) The DE networks are more strongly co-expressed in the ASD toddlers compared to TD toddlers. For an unbiased analysis, the union of DE-ASD and DE-TD networks was considered for this analysis. DE networks are composed of high confidence physical and regulatory interactions. c) Genes in the DE-ASD network are highly expressed in the brain between 8 post conception weeks to 1 year old. For each gene, the number of samples strongly expressing the gene (RPKM >5) was counted based on BrainSpan normalized data [82]. The background genes were composed of all protein coding genes that were expressed in our microarray experiment and were present in BrainSpan RNA-Seq dataset. See also Fig S6. d) The activity pattern of the DE-ASD network during the neurodevelopmental period across brain regions. At each time window, the distribution of co-expression strength of interacting gene pairs in DE-ASD network was measured using Pearson's correlation coefficient. The co-expression distribution was next compared with a background distribution using Wilcoxon-Mann-Whitney test. The y-axis indicates the z-transformed p-value of this comparison. e) Leukocyte gene co-expression pattern of interactions in DE-ASD network is conserved in prenatal and early postnatal neocortex transcriptome data. The correlation of interacting gene pairs in the DE-ASD network was calculated from neocortex transcriptome data (8 post conception weeks until 1 year old, postnatal). The correlation patterns were next paired with those observed in ASD leukocytes. A p-value was estimated by comparing the observed preservation of DE-ASD with that of DE-TD (Fig S6). f) Overlaps of DE-ASD network with brain developmental modules and networks. As illustrated, modules and networks enriched for rASD genes significantly overlap with the DE-ASD network (FDR <0.1; Table S2). rASD networks: networks constructed around high confidence rASD genes [7,9]; rASD modules: co-expression modules enriched for rASD genes [10]; other modules: modules that are not enriched for rASD genes [10]. g) Similarity of interactions of a brain co-expression network around rASD genes [9] with ASD and TD samples as measured by Pearson's correlation coefficient. Boxplots represent the observed similarity based on 100 random sub-samplings of 75 ASD and 75 TD samples (~70% of samples in each diagnosis group). The x-axis represents the top percentile of positive and negative interactions based on the brain transcriptome interaction correlation value (see also Fig S5).

With the observed overlap patterns, we next tested for enrichment of rASD genes in our DE-ASD network. For this analysis, we assessed the overlap of DE-ASD network with different rASD gene lists of different size and varying confidence levels. Surprisingly, this analysis demonstrated that rASD genes are not enriched in the DE-ASD network (p-value >0.19; Methods).

The DE-ASD network is enriched for the regulatory targets of rASD genes

Many high confidence rASD genes have regulatory functions [3,7,11,18]. Although the perturbed DE-ASD network is not enriched for rASD genes, it overlaps with co-expression modules and networks of known rASD genes. At the mechanistic level, the observed co-expression of rASD and DE genes in the prenatal brain could be due to the regulatory influence of rASD genes on the DE-ASD network, and thereby mutations in rASD genes could cause the network over-activity and brain maldevelopment in ASD.

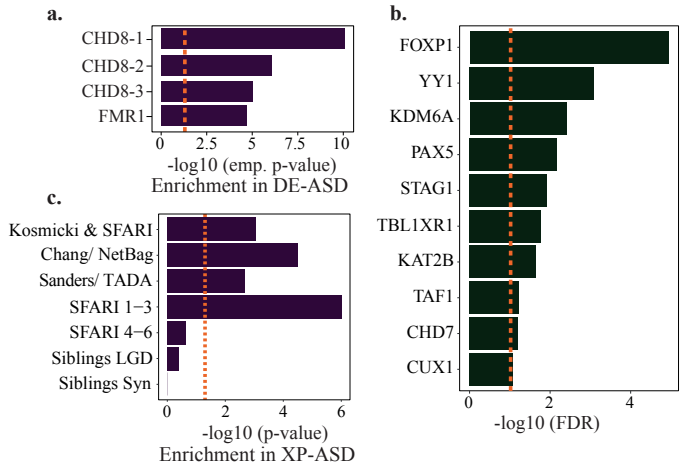

Fig 2. rASD genes are enriched for the regulators of the DE-ASD network.
a) Genes identified by ChIP-Seq as regulatory targets of CHD8 (CHD8-1: Sugathan et al. [38]; CHD8-2: Gompers et al[40]; CHD8-3: Cotney et al. [39]) and FMR1 [41], two high confidence rASD genes, are enriched in the DE-ASD network. Enrichment was assessed empirically (Methods). b) The DE-ASD network significantly overlaps with the regulatory targets of rASD genes based on the ENCODE and Chea2016 repositories. FDR <0.1 was considered as significant. c) High confidence genes are significantly enriched in the XP-ASD network (hypergeometric test). The lists of high confidence rASD genes were extracted from SFARI database [44], Kosmicki et al. [16], Chang et al. [7], and Sanders et al. [17]. List of likely gene damaging (LGD) and synonymous (Syn) mutations in typical siblings were extracted from Iossifov et al[15].

To elucidate if rASD genes could regulate the DE-ASD network, we examined if the regulatory targets of rASD genes are enriched in the DE-ASD network. Indeed, we observed that the DE-ASD network is enriched for genes regulated by two high confidence rASD genes, CHD8 [38-40] and FMR1 [41](Fig2.a). To more systematically identify regulators of the network, we evaluated the overlap of the DE-ASD network with the regulatory targets from 845 assays in the ENCODE project [42] and 615 manually curated assays in Chea2016[43]. Strikingly, we observed DE-ASD network is significantly enriched for 11 out of 20 high confidence and suggestive confidence rASD genes (OR: 2.54; p-value: 0.05; Fig 2.b; Table S3).

The DE-ASD network is preferentially linked to high confidence rASD genes

The rASD genes were often not differentially expressed in ASD leukocytes, and therefore the DE-ASD network was not enriched in rASD genes. However, to explore if rASD genes could regulate the DE-ASD network, we expanded the DE-ASD network by including rASD genes. Thus, we obtained an expanded-ASD, XP-ASD, network (Table S4). To construct XP-ASD network, we used a similar approach to that of the DE-ASD network. We first curated a high confidence static network of DE and 965 speculated rASD genes. The context-specific XP-ASD network was next inferred by retaining only the significantly co-expressed interacting pairs in ASD samples. This pruning step results in removal of genes from the static network that do not show significant co-expression patterns with their known partners or regulatory targets in ASD leukocytes. Accordingly, the XP-ASD network included a total of 316 out of 965 (36%) likely rASD genes.

Our list of 965 rASD genes included rASD genes of both high confidence (e.g., recurrently mutated in ASD individuals) and low confidence (some even found in typical siblings of ASD individuals). We reasoned that if the XP-ASD network is truly relevant to the prenatal etiology of ASD, a preferential incorporation of high confidence rASD genes would be expected in the leukocyte-derived XP-ASD network. By following different analytical methods, different groups have separately categorized rASD genes into high and low confidence [7,16,44]. Importantly, we found a reproducible enrichment of high confidence rASD genes in the XP-ASD network (Fig 2.c) with a significant enrichment for strong evidence rASD genes with *de novo* protein truncating variants in ASD subjects (hypergeometric p-value $<3.06\times10^{-6}$). Further corroborating the regulatory role of rASD genes on DE-ASD network, we found a significant enrichment of rASD genes with DNA binding activities in the XP-ASD network (OR: 3.1; p-value $<2.1\times10^{-12}$; Fig S7). Furthermore, the XP-ASD network was not enriched for rASD genes classified as low confidence (p-value >0.24). As negative controls, we constructed two other networks by including genes with likely deleterious and synonymous mutations in siblings of ASD individuals. Consistent with a possible role of XP-ASD networks in ASD, we found these negative control genes are not significantly associated with the DE genes (p-values >0.41; Fig 2.c). The preferential addition of high confidence and regulatory rASD genes supports the relevance of the XP-ASD network to the pathobiology of ASD, and the likelihood that the high confidence rASD genes are regulating the DE-ASD network.

rASD genes show potential suppressing effects on the DE-ASD network

To explore the regulatory effect of the rASD genes on the DE genes, we analyzed their interaction types (i.e., positive or negative correlations, alluding to activator or repressor activity). Comparative analysis of DE- and XP-ASD networks indicated a significant enrichment of negative correlations between rASD and DE genes (p-value $<3.1\times10^{-4}$; Fisher's exact test), suggesting more of an inhibitory role of rASD genes on the DE genes (Fig 3.a).

Supporting the inhibitory role of rASD genes, the DE-ASD network was enriched for genes that were up-regulated by the knock-down of CHD8 in neural progenitor and stem cells; but not for those that were down-regulated (Fig 3.b) [38-40]. Consistently, we observed in our dataset an overall up-regulation of genes that are also up-regulated in knock-down experiments of the transcriptional repressor CHD8 (p-value <0.039 across three different studies; GSEA), but not for those that are down-regulated. We observed a similar up-regulation pattern for the binding targets of the FMR1 rASD gene in the ASD transcriptome (p-value: 0.078; GSEA). The potential inhibitory role of rASD genes on the DE-ASD network was further supported in an independent dataset on neural differentiation. Specifically, we observed an anti-correlated expression pattern between the rASD and the DE genes from the XP-ASD network in *in vitro*-differentiated human neural progenitors (Fig 3.c).

Signaling pathways are central to the leukocyte-based networks

We next identified key pathways involved in the XP-ASD and DE-ASD networks. Biological process enrichment analysis of the XP-ASD network demonstrated it is highly enriched for signaling pathways (Fig 4.a; Table S5). Moreover, the DE-ASD network was highly enriched for PI3K/AKT, mTOR, and related pathways (Fig 4.b). To delineate mechanisms by which rASD genes could dysregulate DE genes, we compared enriched biological processes of DE and rASD genes involved in the XP-ASD network. DE genes were more enriched for cell proliferation related processes, particularly PI3K/AKT and its downstream pathways such as mTOR, autophagy, viral translation, and FC receptor signaling (Fig 4.a-b). However, the rASD genes were better enriched for processes involved in neuron differentiation and maturation including neurogenesis, dendrite development and synapse assembly (Fig 4.a).

Our results suggest up-regulation and elevated co-expression activity of PI3K/AKT and its down-stream pathways in ASD leukocytes (Fig 4.a-b). These processes are involved in brain development and growth during prenatal and early postnatal ages [3,45,46] and focused studies on rASD genes have implicated

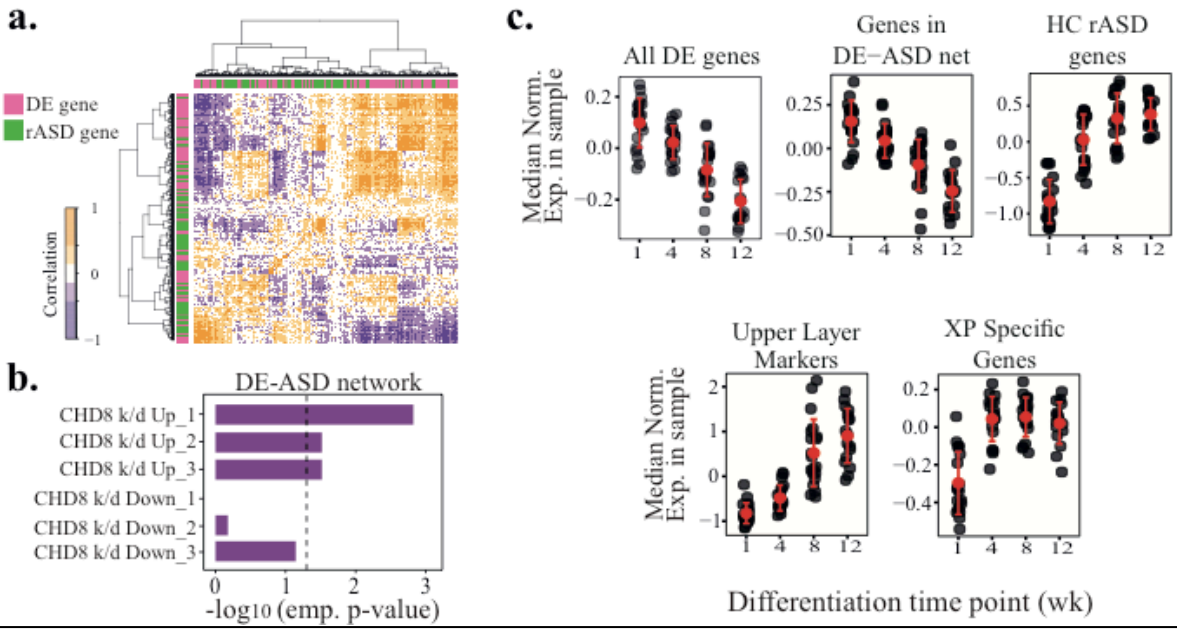

Fig 3. rASD genes potentially suppress the DE genes.
a) Interactions between DE and rASD genes are enriched for negative interactions in the ASD leukocyte transcriptome. See Fig S7 for more details. b) The DE-ASD network is significantly enriched for genes that are up-regulated in response to the knock-down of CHD8 gene. Data were extracted from three studies: Sugathan et al [38] (CHD8 k/d_1), Gompers et al (CHD8 k/d_2)[40], and Cotney et al [39] (CHD8 k/d_3). See also Fig S8. c) Expression patterns of DE genes are negatively correlated with those of rASD genes based on *in vitro* differentiation of human primary neural precursor cells [84]. In each panel, gray circles represent the median expression of associated genes. Expression levels of each gene was normalized to have mean of zero and standard deviation of one across samples. While genes in the DE-ASD network are significantly down-regulated during neuron differentiation (p-value = 4.4×10^{-6}), XP specific genes (i.e., rASD genes present only XP-ASD network, but not DE-ASD) are significantly up-regulated (p-value = 1.2×10^{-3}). The expression levels of CACNA1E, PRSS12, and CARTPT were considered as the markers of upper layer neurons[83] (late stage of neural differentiation). See Fig S7 for the related details.

them in ASD [3,8,47,48]. Further supporting the over-activity of the PI3K/AKT and its down-stream pathways in our cohort of ASD toddlers, gene set enrichment analysis demonstrated genes involved in PI3K/AKT signaling, mTOR pathway and the targets of the FOXO1 transcriptional repressor (the two main downstream processes of the PI3K/AKT) are altered in ASD leukocytes in directions that are consistent with PI3K/AKT over-activity (Supplementary Notes).

We further investigated the DE-ASD and XP-ASD networks using an integrated hub analysis approach (Methods). In the DE-ASD network, 63% of hub genes were involved in or regulated by the PI3K/AKT pathway including PIK3CD, AKT1 and GSK3B (Fig 4.c). The PI3K/AKT pathway is known to be active in the prenatal brain and involve in neural cell proliferation and maturation[3]. Consistent with a potential regulatory role of rASD genes on the DE-ASD networks, genes that were only hubs in the XP-ASD network were highly enriched for the regulatory genes associated with neuronal proliferation and

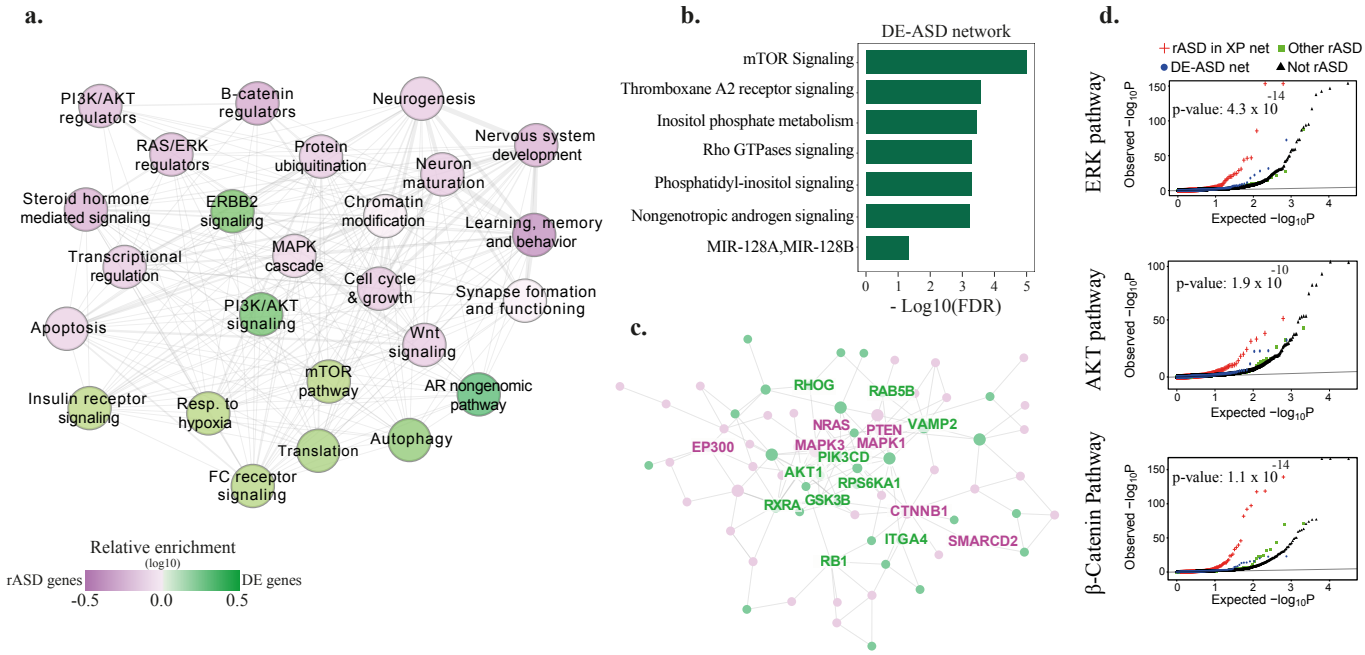

Fig 4. The architecture of the XP-ASD network.
a) Summary of enriched biological processes in the XP-ASD network. Each node represents a biological process that is significantly enriched in the XP-ASD network (Fisher's exact test; Table S5). Nodes that preferentially include rASD and DE genes are represented by purple and green colors, respectively. The interactions among terms represent the connection patterns of their cognate genes in the XP-ASD network with thicker interactions indicating more significant connections (hypergeometric test). Only connections with p-value <0.05 are shown. This illustration covers 86% of genes involved in the XP-ASD network. b) All processes that are significantly enriched in the DE-ASD network and up-regulated in ASD leukocytes based on GSEA. c) The connected graph of hubs in the XP-ASD network. Green nodes represents genes that were hubs in both XP-ASD and DE-ASD networks. Genes that were hubs only in the XP-ASD network are in purple. d) Significant enrichment of rASD genes in the XP-ASD network for the regulators of RAS/ERK, PI3K/AKT, WNT and β-catenin pathways. The x-axis indicates the p-value that gene mutations would dysregulate the corresponding signaling pathways. The background is composed of all genes that were assayed in Brockmann et al [52], excluding rASD and DE genes. The significance of enrichment of rASD genes in XP-ASD network for the regulators of signaling pathways were measured using Wilcoxon-Mann-Whitney test with background genes (illustrated in black) as control. See Fig S8-S9 for more details.

maturation, including regulatory members of the RAS/ERK (e.g., NRAS, ERK2, ERK1, SHC1), PI3K/AKT (e.g., PTEN, PIK3R1, EP300), and WNT/β-catenin (e.g., CTNNB1, SMARCC2, CSNK1G2) pathways (Fig 4.c; Table S6-S7). While PI3K/AKT (a hub in DE-ASD and XP-ASD networks) promotes proliferation and survival, the ERK pathway (a hub in the XP-ASD network) can trigger differentiation of neural progenitor cells by mediating PI3K/AKT associated signaling pathways[3,49-51].

rASD genes regulate DE-ASD genes through specific signaling pathways

We further explored if perturbation to the rASD genes lead to the perturbation of the DE-ASD network through changes in the RAS/ERK, PI3K/AKT, and WNT/β-catenin pathways. To assess this, we leveraged genome-wide mutational screening data in which gene mutations were scored based on their effects on the activity of the RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways [52]. The activity of the signaling pathways was directly measured based on the phosphorylation state of ERK, AKT, and β-catenin proteins [52]. Consistent with functional enrichment and hub analysis results, we found that rASD genes in the XP-ASD network are significantly enriched for regulators of RAS/ERK, PI3K/AKT, and WNT/β-catenin pathways (Fig 4.d; p-value $<1.9\times10^{-10}$; Wilcoxon-Mann-Whitney test). Specifically, regulators of these pathways (FDR<0.1) accounted for inclusion of 39% rASD genes in the XP-ASD. As the control, no significant enrichment for the regulators of RAS/ERK, PI3K/AKT, and WNT/β-catenin pathways were observed among rASD genes that were not included in the XP-ASD network (Fig 4.d). These results support the regulatory role of rASD genes on the DE-ASD network through perturbation of RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways.

In summary, our XP-ASD network decomposition suggests a modular regulatory structure for the XP-ASD network in which diverse rASD genes converge upon and dysregulate activity of the DE genes (Fig 4.a). Importantly, for a large percentage of rASD genes, the dysregulation flow to the DE genes is canalized through highly inter-connected signaling pathways including RAS/ERK, PI3K/AKT, and WNT/β-catenin.

The DE-ASD network is over-active in neuron models of ASD

Our results demonstrate the presence of an over-active network in leukocytes of living ASD toddlers. Furthermore, they implicate the over-activity of the DE-ASD network in the prenatal etiology of ASD by demonstrating the activity of the perturbed network during brain development and its associations with high confidence rASD genes. Also, our results suggest that the network over-activity signal is present in a large percentage of our ASD toddlers and is associated with neural proliferation and maturation.

To further validate these results, we first examined if the DE-ASD network is over-active in iPS cell-derived neural progenitors and neurons of ASD toddlers compared to those of TD cases. For this, we

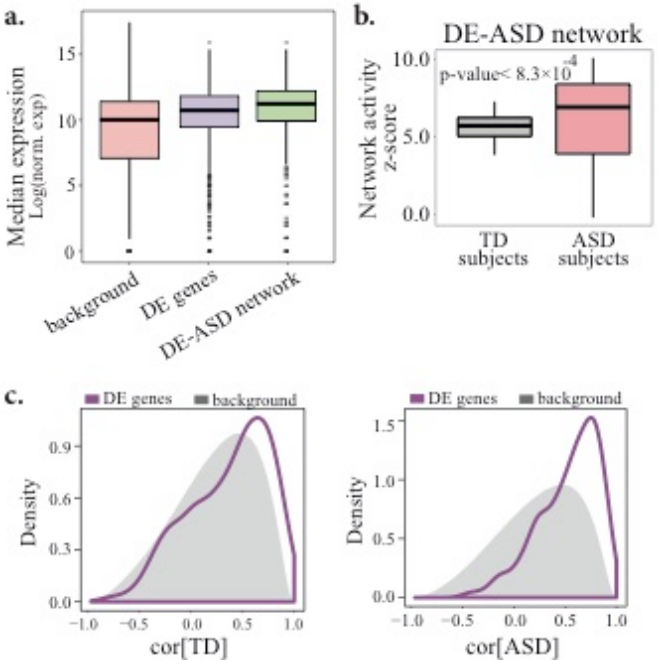

Fig 5. The DE-ASD network is over-active in differentiating neurons of ASD cases.
a) The DE-ASD network is more highly expressed during neural differentiation of iPSCs from ASD and TD cases, as measured by RNA-Seq. Median expression of the genes at neural progenitors and neurons stages were considered. b) The DE-ASD network shows higher co-expression level in ASD derived neural progenitors and neurons. To estimate the co-expression strength of interacting gene pairs in DE-ASD network in neural progenitor and neurons of ASD and TD cases, we sub-sampled the dataset 100 times and measured the activity level at each iteration (a balanced number of ASD and TD samples were selected at each iteration; see Methods). The boxplots represent the distribution of z-transformed p-values of co-expression strength as measured by Wilcoxon-Mann-Whitney test. c) Change in the over-activity of interactions present in the DE-ASD network as measured by co-expression strength. The background distribution is based on the co-expression distribution of randomly selected genes that show the same mean expression pattern as those in of DE-ASD network.

analyzed the transcriptomes of iPS cells from 8 ASD individuals with macrocephaly and 6 TD individuals [30], which were differentiated into neural progenitor and neuron stages. Analysis of the DE-ASD at neural stages demonstrated that the network is over-active in these ASD neuron models (Fig 5), suggesting the functional relevance of identified leukocyte molecular signatures to the abnormal ASD brain development.

Network dysregulation is associated with ASD severity

We evaluated the potential role of the DE-ASD network activity on the development of early-age ASD symptoms. For this, we first tested if the same gene dysregulation patterns exist across individuals at different levels of ASD severity. Indeed, we observed that the fold change patterns of DE genes are almost identical across different ASD severity levels (Fig S11). The implicated RAS/ERK, PI3K/AKT, WNT and β-catenin pathways in our model are well known to have pleotropic roles during brain development from neural proliferation and neurogenesis to neural migration and maturation with implications in ASD [3], suggesting the DE-ASD network is involved in various neurodevelopmental related processes. At the mechanistic level, this suggests that the spectrum of autism could be mediated through the extent of dysregulation of the DE-ASD network, as it is composed of high confidence physical and regulatory interactions. Hence, we examined whether the magnitude of the co-expression activity level of the DE-ASD network correlated with clinical severity across individual ASD toddlers. Indeed, we found that the extent of gene co-expression activity within the DE-ASD network was correlated with ASD toddlers' ADOS social affect deficit scores, the ASD diagnostic gold standard (Fig 6). To assess the significance of observed correlation patterns, we repeated the analysis with 10,000 permutations of the ADOS social affect scores of ASD individuals (see inset boxplots in Fig 6). This analysis demonstrated the significance of the observed

correlations (Fig 6). Our results suggest the perturbation of the same network at different extents can potentially result in a spectrum of postnatal clinical severity levels in ASD toddlers.

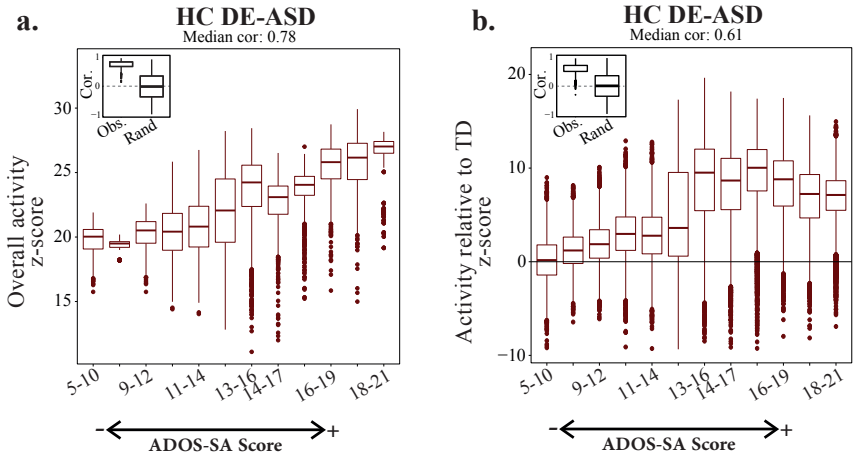

Fig 6. Activity level of DE-ASD networks correlates with ASD severity.
a) ASD toddlers were sorted by their ADOS social affect scores (ADOS-SA) with higher scores representing more severe cases. The network activity was measured in a running window on ADOS-SA scores. The overall activity of DE-ASD network in a set of samples was measured by comparing the co-expression strength of interactions in the network with the background derived from the same set of samples (Methods). To ensure robustness of the results, we measured activity level of DE-ASD network at each severity group by randomly selecting 20 samples from that severity level, iterating 1000 times. The left inset panel illustrates the distribution of observed correlation values of DE-ASD network for ranges of ADOS severity, and compares it with permuted data, with 10,000 random shuffling of ADOS-SA scores of ASD cases. b) The relative activity of DE-ASD networks compared to TD cases. The relative activity level was estimated by comparing the co-expression strength of interactions in DE-ASD network between ASD and TD toddlers. For each severity group, 20 samples were randomly selected from each of ASD and TD samples, iterating 1000 times. Significance of the trend was evaluated by 10,000 permutations of the ADOS-SA scores of ASD cases.

Conclusion

While ASD demonstrates a strong genetic basis, it remains elusive how implicated genes are connected to the molecular dysregulations that underlie the disorder at prenatal and early postnatal ages. Towards this, we developed a systems biology framework that integrates transcriptomic dysregulations in living ASD toddlers with current knowledge on ASD risk genes to explain ASD associated fetal-stage brain transcriptomic changes and clinical outcomes. Specifically, we found a dysregulated transcriptional network that shows elevated gene co-expression activity in ASD toddlers. This core network was robustly associated with rASD genes with likely deleterious mutations in ASD subjects. Such rASD genes have potentially large effect sizes on the etiology but occur in a small percentage of the ASD population [53,54]. We show that many rASD genes may exert their regulatory effect on this DE-ASD core network through the inter-connected RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways. The connection of the DE-

ASD network (constructed with data from the general ASD pediatric population) with high confidence rASD genes provides empirical evidence of shared mechanisms underlying ASD in both those with highly penetrant rASD genes and those of other etiologies (e.g., common variants) in the wider ASD population.

The key aspect of our signature is that it is constructed based on transcriptomic data from young living ASD toddlers. This allows us to correlate its variations with the core clinical features of the same ASD toddlers. Indeed, the dysregulation degree of the DE-ASD network correlated with deficits in the toddlers' ADOS social affect scores. Social and behavioral deficits are also suggested to be correlated with the genetic variations in ASD subjects [55,56]; and previous studies have established the effect of the PI3K/AKT signaling pathway (central to the DE-ASD core network and significantly altered in ASD leukocytes) on social behaviors of mouse models [47,48]. Together, these observations suggest that the etiological roots of ASD converge on gene networks that correlate with the symptom severity in ASD individuals. Moreover, our results reinforce the hypothesis that stronger dysregulation of the same core network could lead to higher severity in the ASD cases. The DE-ASD core network is enriched for pathways implicated in ASD, strongly associated with high confidence rASD genes, and correlate with ASD severity. However, we note that a direct causal relationship between the co-expression activity of the network and ASD remains to be established. Moreover, our network co-expression activity measure is a summary score from the strongest signal in our dataset (i.e., differentially expressed genes) at a group level (i.e., severity level). Therefore, by design, it may not capture the heterogeneity that could exist within each group. As detailed below, future work is needed to explore the causal relationship of our gene network to ASD development, symptoms, and the potential existence of other dysregulation mechanisms in ASD individuals.

The emerging architecture of complex traits suggests that gene mutations often propagate their effects through regulatory networks and converge on core pathways relevant to the trait [22,32]. Our findings support the existence of an analogous architecture for ASD, wherein rASD genes with diverse biological roles overlap in their down-stream function. Although not significantly overlapping with rASD genes, we found that the DE-ASD network is significantly co-expressed with rASD genes in both leukocyte and brain. We also illustrated that the DE-ASD network could be controlled by rASD genes through direct transcriptional regulation or highly interconnected signaling pathways. We postulate that the DE-ASD network is a primary convergence point of ASD etiologies, including its genetic basis as we elaborated for rASD genes, in a large portion of the ASD population. This predicts that the spectrum of autism in such cases is correlated with the degree and mechanism of the perturbation of the DE-ASD network. A detailed analysis of iPS cell-derived ASD neurons demonstrated the dysregulation of the leukocyte-based DE-ASD network in ASD neurons, supporting the neural-level relevance of the findings to ASD etiology and its prevalence in the ASD population. Furthermore, direct clinical-level relevance is demonstrated by the high

correlation we found between degree of dysregulation in the DE-ASD core network and ASD symptom severity in the ASD toddlers.

The currently recognized rASD genes are not fully penetrant to the disorder, except for a handful of syndromic genes [53,54,57,58]. Our analysis of the XP-ASD network provides some insights on how the effects of rASD genes could potentially combine to result in ASD. Although some rASD genes could directly modulate the DE-ASD network at the transcriptional level, our results suggest that the regulatory consequence of many rASD genes on the DE-ASD network is canalized through the PI3K/AKT, RAS/ERK, WNT and β-catenin signaling pathways. The structural and functional interrogation of the DE-ASD network localized the PI3K/AKT pathway to its epicenter and demonstrated enrichment for processes down-stream of this pathway. Moreover, we found that high confidence rASD genes are better connected to the DE-ASD core network, suggesting that the closeness and influence of genes on these signaling pathways is correlated with their effect size on the disorder. These results articulate that perturbation of the PI3K/AKT, RAS/ERK, WNT and β-catenin signaling pathways through gene regulatory networks may be an important etiological route for ASD that could be associated with the disorder severity level in a relatively large fraction of the ASD population. Congruent with this hypothesis, cell and animal models of ASD have demonstrated the enrichment of high confidence rASD genes for the regulators of the RAS/ERK, PI3K/AKT, WNT and β-catenin signaling pathways [3,8,11,18,47,48,51]. These signaling pathways are highly conserved and pleiotropic, impacting multiple prenatal and early postnatal neural development stages from proliferation/differentiation to synaptic and neural circuit development [3]. Such multi-functionalities could be the underlying reason that we detected the signal in ASD leukocytes.

It is necessary to analyze large subject cohorts from unbiased, general pediatric community settings to capture the heterogeneity that underlies ASD at early ages. This study presents the largest transcriptome analysis on early-age ASD cases thus far from such settings. However, the analyzed dataset is still of a modest size, and as such our analysis was focused on the strongest signal that best differentiates ASD cases from TD individuals (i.e., differentially expressed genes). Here we illustrate that the captured signal is informative about the transcriptional organization of ASD and shows promise in bridging the gap between genetic and clinical outcomes. Future studies with larger datasets are required to not only replicate these results, but also explore other long-standing questions in the field, such as the basis of gender bias that exists in ASD or the potential molecular mechanisms that differentiate high functioning from low functioning cases. However, perhaps the most exciting direction is to expand the presented framework to systematically diagnose, classify and prognostically stratify ASD cases at early postnatal ages based on the underlying molecular mechanisms. The concept of precision molecular medicine for ASD can only be actualized via approaches that illuminate the early-age living biology of ASD [3,18,21]. ASD toddler-derived iPS cell studies show ASD is a progressive prenatal and early postnatal disorder that involves a cascade of

352 diverse and varying molecular and cellular changes such as those resulting from dysregulation of the
353 pathways and networks highlighted herein [3,30,31]. As such, dynamic, individual-based molecular assays in
354 infants and toddlers will be essential to develop. The presented framework could prove invaluable for the
355 development of quantitative, molecular-based measures for the ASD diagnosis and prognosis by identifying
356 specific molecular dysregulations that we show are observable in leukocytes of a large fraction of living
357 ASD toddlers at young ages.

Materials and Methods

Participant recruitment and clinical evaluation

The primary aim of this study was to associate the transcriptome dysregulations present in ASD leukocytes with the ASD risk genes. However, the currently available genetic information is mostly based on males, and less is known about the genetic basis of ASD females. Therefore, we focused on male toddlers for the transcriptome analysis in this study, specifically 264 male toddlers with the age range of 1 to 4 years. Part of the transcriptome data of this study (153 individuals) was reported previously [21,59] and a similar methodology was employed for participant recruitment and sample collection from 111 new cases [21]. Research procedures were approved by the Institutional Review Board of the University of California, San Diego. Parents of subjects underwent Informed Consent Procedures with a psychologist or study coordinator at the time of their child's enrollment.

About 70% of toddlers were recruited from the general population as young as 12 months using an early detection strategy called the 1-Year Well-Baby Check-Up Approach [60]. Using this approach, toddlers who failed a broadband screen, the CSBS IT Checklist [61], at well-baby visits in the general pediatric community settings were referred to our Center for a comprehensive evaluation. The remainder of the sample was obtained by general community referrals. All toddlers received a battery of standardized psychometric tests by highly experienced Ph.D. level psychologists including the Autism Diagnostic Observation Schedule (ADOS; Module T, 1 or 2), the Mullen Scales of Early Learning and the Vineland Adaptive Behavior Scales. Testing sessions routinely lasted 4 hours and occurred across 2 separate days. Toddlers younger than 36 months in age at the time of initial clinical evaluation were followed longitudinally approximately every 9 months until a final diagnosis was determined at age 2-4 years. For analysis purposes, toddlers (median age, 27 months) were categorized into two groups based on their *final* diagnosis assessment: 1) ASD: subjects with ASD diagnosis or ASD features; 2) TD: typically developing (TD) controls. For more information see Table S1.

ADOS scores at each toddler's final visit were used for correlation analyses with DE-ASD network activity scores. All but 4 toddlers were tracked and diagnosed using the appropriate module of the ADOS (i.e., Toddler, 1, or 2) between the ages of 24-49 months (Table S1), an age where the diagnosis of ASD is relatively stable [62-64]; the remaining 4 toddlers had their final diagnostic evaluation between the ages of 18 to 24 months.

Blood sample collection and microarray gene expression processing

Blood samples were usually taken at the end of the clinical evaluation sessions. In order to monitor health status, the temperature of each toddler was monitored using an ear digital thermometer immediately preceding the blood draw. The blood draw was scheduled for a different day in cases that the temperature

was higher than 99 Fahrenheit. Moreover, blood draw was not taken if a toddler had some illness (e.g., cold or flu), as observed or stated by parents. We collected four to six milliliters of blood into ethylenediaminetetraacetic-coated tubes from all toddlers. Blood leukocytes were captured and stabilized by LeukoLOCK filters (Ambion) and were immediately placed in a −20°C freezer. Total RNA was extracted following standard procedures and manufacturer's instructions (Ambion).

RNA labeling, hybridization, and scanning was conducted at Scripps Genomic Medicine center, (CA, USA) using Illumina BeadChip technology. All arrays were scanned with the Illumina BeadArray Reader and read into Illumina GenomeStudio software (version 1.1.1). Raw Illumina probe intensities were converted to expression values using the lumi package[65]. We employed a three-step procedure to filter for probes with reliable expression levels. First, we only retained probes that met the detection p-value <0.05 cut-off threshold in at least 3 samples. Second, we required the probes to have expression levels above 95 percentile of negative probes in at least 50% of samples. The probes with detection p-value >0.1 across all samples were selected as negative probes and their expression levels were pooled together to estimate the 95 percentile expression level. Third, for genes represented by multiple probes, we considered the probe with highest mean expression level across our dataset, after quantile normalization of the data. These criteria led to the selection of 14,854 protein coding genes as expressed in our leukocyte transcriptome data, which is similar to the previously reported estimate of 14,631 protein coding genes (chosen based on Entrez Ids) for whole blood by GTex consortium[66]. To ensure results are not affected by the variations in the procedure of selecting expressed genes, we replicated all of our analyses (redoing DE analysis and re-constructing HC DE and XP networks) by choosing 13,032 protein coding genes as expressed (Fig S14).

Data processing and differential gene expression analysis of microarray datasets

We subdivided our microarray samples into three datasets to assess the reproducibility of the results. The primary dataset included 253 high quality samples and was used for the discovery of the dysregulation signal. The second dataset replicated 56 randomly selected male toddlers from the primary dataset (35 ASD and 21 TD). The third dataset was composed of 48 male toddlers with 24 independent, non-overlapping ASD cases, while 21 out of 24 TD cases overlapped with the primary dataset. The second and third datasets were microarrays generated at the same time, but included different subjects not in the primary dataset. All three datasets used Illumina microarray technology. However, the primary dataset was analyzed by Illumina HT-12 Chips, while the second and third datasets used Illumina WG-6 Chips. The pre-processing and downstream analysis of the three datasets were conducted separately. The data are available in the Gene Expression Omnibus database (GSE42133;GSE111175).

The primary dataset was originally composed of 275 samples from 240 male ASD and TD individuals. Quality control analysis was performed to identify and remove 22 outlier samples from the

dataset. Samples were marked as outlier if they showed low signal intensity of the microarray (average signal of two standard deviations lower than the overall mean), deviant pairwise correlations, deviant cumulative distributions, deviant multi-dimensional scaling plots, or poor hierarchical clustering, as described elsewhere [20]. After removing low quality samples, the primary dataset had 253 samples from 226 male toddlers including 27 technical replicates. High reproducibility was observed across technical replicates (mean Spearman correlation of 0.917 and median of 0.925). We randomly removed one of each of two technical replicates from the dataset.

The limma package [67] was then applied on quantile normalized data for differential expression analysis in which moderated t-statistics was calculate by robust empirical Bayes methods[68]. Sample batch was used as a categorical covariate (total of two batches; both Illumina HT-12 platforms). Exploration graphs indicated that linear modeling of batch covariate was effective at removing its influence on expression values (Fig S13). MA-plots of the primary dataset did not show existence of overall bias in the fold change estimates (Fig S1). DE analysis identified 1236 differentially expressed genes with Benjamini-Hochberg FDR <0.05.

We performed multiple analyses to confirm that our results (1) are replicable in the other two microarray datasets, (2) are robust to alterations in the analysis pipeline, (3) are not affected by the batches or potential hidden covariates, (4) are present in the vast majority of samples, and (5) are not driven by changes in the blood cell type composition between ASD and TD toddlers (Figs S1-S4).

Reproducibility of transcriptional over-activity of DE-ASD networks in an independent RNA-Seq dataset

We performed RNA-Seq experiments on 56 samples from an independent cohort of 12 (19 samples) TD and 23 (37 samples) ASD male toddlers. None of subjects overlapped with those in the primary dataset. This allowed us to ensure our results are not subject nor platform (i.e., microarray vs. RNA-Seq) specific.

RNA-Seq libraries were sequenced at the UCSD IGM genomics core on a HiSeq 4000. We processed the raw RNA-Seq data with our pipeline that starts with quality control with FastQC[69]. Low quality bases and adapters were removed using trimmomatic[70]. Reads were aligned to the genome using STAR[71]. STAR results were processed using Samtools[72], and transcript quantification is done with HTseq-count[73]. Subsequently, low expressed genes were removed and data were log count per million (cpm) normalized (with prior read count of 1) using limma[67]. We performed SVA analysis[74] on the normalized expression data and included the first surrogate variable as covariate to account for potential hidden confounding variables. Differential expression analysis was performed using Limma package with subjects modeled as random effects.

458 **ASD risk genes**

459 ASD risk genes were extracted from the SFARI database [44] on Dec. 7, 2016. We also included the
460 reported risk genes from a recent meta-analysis of two large-scale genetic studies, containing genes mutated
461 in ASD individuals but not present in Exome Aggregation Consortium database (ExAC)[16]. Together, these
462 two resources provided 965 likely rASD genes that were used for the construction of XP-ASD networks
463 (Table S8). Previously published genes with likely gene damaging and synonymous mutations in ASD
464 siblings were retrieved from Iossifov et al.[15].

465 ASD high confidence risk genes were extracted from the SFARI database (genes with confidence
466 levels of 1 and 2), Kosmicki et al. [16] (recurrent gene mutations in ASD individuals, but not present in ExAC
467 database), Sanders et al.[17], and Chang et al. [7]. Strong evidence genes with *de novo* protein truncating
468 variants in ASD subjects were extracted from Kosmicki et al.[16] and included rASD genes that were not in
469 ExAC database and with a probability of loss-of-function intolerance (pLI) score of above 0.9. Gene names
470 in these datasets were converted to Entrez gene ids using DAVID tools [75].

471 To assess the overlap of DE-ASD networks with rASD genes, we considered our list of all rASD
472 genes (965 genes), different lists of high confidence rASD genes (varying in size and composition) and
473 their combinations, including all SFARI rASD genes, SFARI genes levels 1-to-3, SFARI genes levels 1 and
474 2, strong evidence rASD genes from Kosmicki et al.[16], and strong evidence rASD genes from Sanders et
475 al.[17]

476

477 **Functional characterization of DE-ASD networks**

478 We set two criteria to identify biological processes that are differentially expressed between ASD
479 and TD samples and are enriched in the DE-ASD networks. First, we required the biological process to be
480 significantly changed between ASD and TD transcriptome samples. Second, we required the biological
481 process to be significantly enriched in the DE-ASD networks.

482 GSEA identified multiple gene sets that were significantly upregulated in ASD samples (FDR
483 <0.12; Table S9), using the R version of the GSEA package and msigdb.v5.1 database (downloaded on Oct.
484 20, 2016) [76,77]. Significantly enriched processes in the DE-ASD networks were identified by examined the
485 overlap of GSEA-identified significantly altered gene sets with the DE-ASD networks based on empirical
486 permutation tests, and p-values were corrected for multiple testing using Benjamini-Hochberg procedure.
487 We excluded gene sets annotated as associated with specific reference datasets in MSigDB since their
488 generalizability to our dataset has not been established (Table S9).

489

Biological enrichment analysis of XP-ASD networks

Significantly enriched Gene Ontology biological processes (GO-BP) were identified by Fisher's exact test on terms with the 10-2000 annotated genes. The terms with Benjamini-Hochberg estimated FDR <0.1 were deemed as significant. The enriched terms were next clustered based on the GO-BP tree, extracted from Amigo database using RamiGO package in R [78]. The general terms with more than 1000 annotated genes that spanned two or more clusters were removed. The list of enriched GO-BP terms and their clustering are provided in Table S5.

Deciphering potential regulators of DE-ASD networks

To identify genes that potentially regulate DE-ASD networks, we examined the overlap of DE-ASD networks with identified targets of human transcription factors as part of ENCODE project[42] and curated Chea2016 database[43]. We performed overlap analysis with each of the three DE-ASD networks separately using the EnrichR portal. Some of the transcription factors were assayed multiple times. To obviate potential biases, we used Fisher's method to combine the enrichment p-values across assays related to a given transcription factor during the analysis of each DE-ASD networks. Next, p-values were corrected using the Benjamini-Hochberg procedure. Only transcription factors whose targets were significantly enriched in all three DE-ASD networks were considered as significantly overlapping with the DE-ASD networks (FDR <0.1).

Brain developmental gene expression data

Normalized RNA-Seq transcriptome data during human neurodevelopmental time periods were downloaded from the BrainSpan database on Dec. 20, 2016[36,37]. To calculate correlations, normalized RPKM gene expression values were log2(x+1) transformed.

Neural progenitor differentiation data

Microarray transcriptome data from differentiation of primary human neural progenitor cells to neural cells [79] were downloaded from the NCBI GEO database (GSE57595). The data were already quantile normalized and ComBat batch-corrected [80]. For genes with multiple probes, we retained the probe with the highest mean expression value.

To observe the transcriptome response of XP-ASD networks during neuron differentiation, we correlated the gene expression patterns with the developmental time points, considering the differentiation time as an ordinal variable. The results are represented in Fig S7.

519 **ASD induced pluripotent stem cells (iPSC) data**

520 ASD iPSC data [30] were downloaded from GEO (GSE67528). Gene expression counts were
521 normalized with the TMM method [81] and filtered to exclude low-expressed genes (genes with count per
522 million greater than 1 were retained). To calculate the correlations, normalized RNA-Seq gene expression
523 values were log(x+1) transformed.

524 **Regulatory effect of gene mutations on signaling pathways**

525 Data were extracted from a genome-wide mutational study that monitored the regulatory effect of
526 gene mutations on phosphorylation status of 10 core genes of different signaling pathways and processes [52].
527 Genes whose mutations affected the phosphorylation status of the core signaling genes with FDR <0.1 were
528 considered as the regulators of the cognate signaling pathway.

Acknowledgments

Authors would like to thank Dr. Lilia Iakoucheva for the critical review of this manuscript. This work was supported by NIMH R01-MH110558 (EC, NEL), NIMH R01-MH080134 (KP), NIMH R01-MH104446 (KP), NFAR grant (KP), NIMH P50-MH081755 (EC), Brain & Behavior Research Foundation NARSAD (TP), and generous funding from the Novo Nordisk Foundation through Center for Biosustainability at the Technical University of Denmark (NNF10CC1016517).

References

1. Stoner, R. *et al.* Patches of disorganization in the neocortex of children with autism. *N Engl J Med* **370**, 1209-1219 (2014).
2. Courchesne, E. *et al.* Neuron number and size in prefrontal cortex of children with autism. *JAMA* **306**, 2001-10 (2011).
3. Courchesne, E. *et al.* The ASD Living Biology: from cell proliferation to clinical phenotype. *Mol Psychiatry* (2018).
4. Sandin, S. *et al.* The Heritability of Autism Spectrum Disorder. *JAMA* **318**, 1182-1184 (2017).
5. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-5 (2014).
6. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci* **19**, 1454-1462 (2016).
7. Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J. & Vitkup, D. Genotype to phenotype relationships in autism spectrum disorders. *Nat Neurosci* **18**, 191-8 (2015).
8. de la Torre-Ubieta, L., Won, H., Stein, J.L. & Geschwind, D.H. Advancing the understanding of autism disease mechanisms through genetics. *Nat Med* **22**, 345-61 (2016).
9. Willsey, A.J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
10. Parikshak, N.N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-21 (2013).
11. Hormozdiari, F., Penn, O., Borenstein, E. & Eichler, E.E. The discovery of integrated gene networks for autism and related disorders. *Genome Res* **25**, 142-54 (2015).
12. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* **94**, 677-94 (2014).
13. Lin, G.N. *et al.* Spatiotemporal 16p11.2 protein network implicates cortical late mid-fetal brain development and KCTD13-Cul3-RhoA pathway in psychiatric diseases. *Neuron* **85**, 742-54 (2015).
14. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat Genet* **47**, 582-8 (2015).
15. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216-21 (2014).
16. Kosmicki, J.A. *et al.* Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat Genet* **49**, 504-510 (2017).
17. Sanders, S.J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-33 (2015).
18. Sahin, M. & Sur, M. Genes, circuits, and precision therapies for autism and related neurodevelopmental disorders. *Science* **350**(2015).
19. Wright, F.A. *et al.* Heritability and genomics of gene expression in peripheral blood. *Nat Genet* **46**, 430-7 (2014).
20. Pramparo, T. *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol* **11**, 841 (2015).
21. Pramparo, T. *et al.* Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* **72**, 386-94 (2015).
22. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
23. Nishimura, Y. *et al.* Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. *Hum Mol Genet* **16**, 1682-98 (2007).
24. Achuta, V.S. *et al.* Functional changes of AMPA responses in human induced pluripotent stem cell-derived neural progenitors in fragile X syndrome. *Sci Signal* **11**(2018).
25. Hu, V.W. *et al.* Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. *PLoS One* **4**, e5775 (2009).

26. Hu, V.W., Frank, B.C., Heine, S., Lee, N.H. & Quackenbush, J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC Genomics* **7**, 118 (2006).
27. Hu, V.W. *et al.* Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Res* **2**, 78-97 (2009).
28. Kong, S.W. *et al.* Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One* **7**, e49475 (2012).
29. Diaz-Beltran, L. *et al.* Cross-disorder comparative analysis of comorbid conditions reveals novel autism candidate genes. *BMC Genomics* **18**, 315 (2017).
30. Marchetto, M.C. *et al.* Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol Psychiatry* **22**, 820-835 (2017).
31. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375-390 (2015).
32. Califano, A. & Alvarez, M.J. The recurrent architecture of tumour initiation, progression and drug sensitivity. *Nat Rev Cancer* **17**, 116-130 (2017).
33. Ideker, T. & Krogan, N.J. Differential network biology. *Mol Syst Biol* **8**, 565 (2012).
34. Yang, B. *et al.* Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. *Nat Commun* **9**, 678 (2018).
35. Chen, L., Liu, R., Liu, Z.P., Li, M. & Aihara, K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* **2**, 342 (2012).
36. BrainSpan. BrainSpan: Atlas of the Developing Human Brain. (2016).
37. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-9 (2011).
38. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci U S A* **111**, E4468-77 (2014).
39. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun* **6**, 6404 (2015).
40. Gompers, A.L. *et al.* Germline Chd8 haploinsufficiency alters brain development in mouse. *Nat Neurosci* **20**, 1062-1073 (2017).
41. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247-61 (2011).
42. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
43. Lachmann, A. *et al.* ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* **26**, 2438-44 (2010).
44. Abrahams, B.S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
45. Hur, E.M. & Zhou, F.Q. GSK3 signalling in neural development. *Nat Rev Neurosci* **11**, 539-51 (2010).
46. Lipton, J.O. & Sahin, M. The neurology of mTOR. *Neuron* **84**, 275-91 (2014).
47. Clipperton-Allen, A.E. & Page, D.T. Pten haploinsufficient mice show broad brain overgrowth but selective impairments in autism-relevant behavioral tests. *Hum Mol Genet* **23**, 3490-505 (2014).
48. Cupolillo, D. *et al.* Autistic-Like Traits and Cerebellar Dysfunction in Purkinje Cell PTEN Knock-Out Mice. *Neuropsychopharmacology* **41**, 1457-66 (2016).
49. Chen, Y., Huang, W.C., Sejourne, J., Clipperton-Allen, A.E. & Page, D.T. Pten Mutations Alter Brain Growth Trajectory and Allocation of Cell Types through Elevated beta-Catenin Signaling. *J Neurosci* **35**, 10252-67 (2015).
50. Rhim, J.H. *et al.* Cell type-dependent Erk-Akt pathway crosstalk regulates the proliferation of fetal neural progenitor cells. *Sci Rep* **6**, 26547 (2016).
51. Mellios, N. *et al.* MeCP2-regulated miRNAs control early human neurogenesis through differential effects on ERK and AKT signaling. *Mol Psychiatry* (2017).

52. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for GPCR signalling. *Nature* **546**, 307-311 (2017).
53. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* **1380**, 42-77 (2011).
54. Huguet, G., Ey, E. & Bourgeron, T. The genetic landscapes of autism spectrum disorders. *Annu Rev Genomics Hum Genet* **14**, 191-213 (2013).
55. Robinson, E.B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet* **48**, 552-5 (2016).
56. Wang, Y. *et al.* Heritable aspects of biological motion perception and its covariation with autistic traits. *Proc Natl Acad Sci U S A* (2018).
57. Amir, R.E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat Genet* **23**, 185-8 (1999).
58. Neale, B.M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* **485**, 242-5 (2012).
59. Glatt, S.J. *et al.* Blood-based gene expression signatures of infants and toddlers with autism. *J Am Acad Child Adolesc Psychiatry* **51**, 934-44 e2 (2012).
60. Pierce, K. *et al.* Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *J Pediatr* **159**, 458-465 e1-6 (2011).
61. Wetherby, A.M. & Prizant, B.M. *Communication and symbolic behavior scales: Developmental profile*, (Paul H Brookes Publishing, 2002).
62. Chawarska, K., Klin, A., Paul, R., Macari, S. & Volkmar, F. A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *J Child Psychol Psychiatry* **50**, 1235-45 (2009).
63. Cox, A. *et al.* Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *J Child Psychol Psychiatry* **40**, 719-32 (1999).
64. Kleinman, J.M. *et al.* Diagnostic stability in very young children with autism spectrum disorders. *J Autism Dev Disord* **38**, 606-15 (2008).
65. Du, P., Kibbe, W.A. & Lin, S.M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-8 (2008).
66. Consortium, G.T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-60 (2015).
67. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
68. Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S. & Smyth, G.K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The annals of applied statistics* **10**, 946 (2016).
69. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).
70. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014).
71. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
72. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
73. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
74. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, 1724-35 (2007).
75. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
76. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).

77. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-50 (2005).
78. Schroder, M.S., Gusenleitner, D., Quackenbush, J., Culhane, A.C. & Haibe-Kains, B. RamiGO: an R/Bioconductor package providing an AmiGO visualize interface. *Bioinformatics* **29**, 666-8 (2013).
79. Stein, J.L. *et al.* A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* **83**, 69-86 (2014).
80. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-27 (2007).
81. Robinson, M.D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25 (2010).
82. Miller, J.A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).
83. Zeng, H. *et al.* Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483-96 (2012).