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 these genes exert their effects at early ages in ASD remain unclear. To identify such mechanisms, 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 core network through the RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways. Finally, the dysregulation degree of this core 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 enable deciphering 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 appear to 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 rASD-induced 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-to-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 what molecular processes are perturbed in prenatal and early postnatal life in ASD individuals, assess how they vary among subjects, and

35 evaluate how these perturbations relate to rASD genes and early-age ASD clinical symptoms. It is expected
36 that the genetic changes in ASD alter gene expression and signaling in the early-age developing brain
37 [3,7,11,18]. Therefore, capturing dysregulated gene expression at prenatal and early postnatal ages may help
38 unravel the underlying molecular organization of ASD. Unfortunately, doing so is particularly challenging
39 as ASD brain tissue cannot be obtained at these early stages, and all available postmortem ASD brains are
40 from much older ages, well beyond the ages when rASD genes are at peak expression and the disorder
41 begins. However, in contrast to living neurons that have a limited time window for proliferation and
42 maturation, other cell types constantly regenerate, such as blood cells. Given the strong genetic basis of
43 ASD, some dysregulated developmental signals may continually reoccur in blood cells and thus be studied
44 postnatally [19-21].
45 Reinforcing this notion, it was recently demonstrated that genes that are broadly expressed across
46 many tissues are major contributors to the overall heritability of complex traits [22], and it was postulated that
47 this could be relevant to ASD. Lending credence to this, previous studies have reported the enrichment of
48 differentially expressed genes in ASD blood for the regulatory targets of CHD8 [20] and FMR1 [23] genes, two
49 well-known rASD genes. Similarly, lymphoblastoid cells of ASD cases and iPS-derived models of fragile-
50 X syndrome show over-expression of mir-181 with a potential role in the disorder [24]. Likewise, leukocytes
51 from ASD toddlers show perturbations in biological processes, such as cell proliferation, differentiation,
52 and microtubules [25-29], and these coincide with dysregulated processes seen in neural progenitor cells
53 (NPCs) and neurons, derived from iPS cells from ASD subjects [30,31]. Ultimately, establishing the signatures
54 of ASD in other tissues will be important to facilitate the study of the molecular basis of the disorder in
55 living ASD subjects in the first years of life.
56 Here we leverage transcriptomic data from leukocytes, stems cell models, and the developing brain
57 to study the underlying architecture of transcriptional dysregulation in ASD, its connection to rASD genes,
58 and its association with prenatal development and clinical outcomes of ASD toddlers. Specifically, we
59 discovered a conserved dysregulated gene network by analyzing leukocyte transcriptomic data from 1-4
60 years old ASD and typically developing (TD) toddlers. The dysregulated network is enriched for pathways
61 known to be perturbed in ASD neurons, impacts highly expressed processes in prenatal brain development,
62 and is dysregulated in iPS cell-derived neurons from ASD cases. Consistent with the omnigenic model of
63 complex traits [22], we show that rASD genes across diverse functional groups converge upon and regulate
64 this core network. Importantly, this core network is disrupted to different levels of severity across ASD
65 individuals, and is correlated with clinical severity in individual ASD toddlers. Thus, our results
66 demonstrate how the heterogeneous genetic basis of ASD converges on a biologically relevant core
67 network, capturing the underlying possible molecular etiology of ASD.

68 **RESULTS**

69 **Leukocytes display transcriptome over-activity in ASD male toddlers**

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

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

89 The context-specific networks (DE-ASD and DE-TD) include published physical and regulatory
90 interactions among DE genes that exhibit within-group co-expression in our data. Notably, although DE-
91 ASD and DE-TD networks are composed of a similar set of genes (i.e., those expressed in the blood and are
92 differentially expressed between ASD and TD samples), the wiring of the two networks differ based on the
93 co-expression patterns within each study group. To assess the possibility that intracellular pathways were
94 being specifically modulated in ASD, we created a merged network by considering the interactions that are
95 present in at least one of DE-ASD and DE-TD networks. We next examined the co-expression strength of
96 the merged network in ASD and TD individuals (Methods) [32-34]. This proxy for the transcriptional activity
97 of gene networks [9] demonstrated that co-expression strength was higher in the ASD than the TD samples
98 (Figure 1.b; p-value < 0.01; paired Wilcoxon-Mann-Whitney test). The stronger co-expression that is driven
99 by the DE-ASD network, suggests a higher level of concerted activation or suppression of pathways
100 involving DE genes in ASD toddlers. This elevated co-expression activity (herein referred to as over-

101 activity) of the network was reproducible in the other two ASD datasets, specific to the DE genes and
102 replicable across alternative analysis methods (Figures S3-S4).
103 In summary, the leukocyte transcriptional networks of the DE genes show higher than normal co-
104 expression activity in ASD toddlers. Moreover, the dysregulation pattern is present in a large percentage of
105 ASD toddlers, as evidenced by the resampling analyses and the other two ASD datasets.

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

107 We next assessed the potential association of the leukocyte-based network to the spatiotemporal
108 neurodevelopmental signals relevant to ASD. By overlaying our network on the *in vivo*
109 neurodevelopmental RNA-seq transcriptome data from BrainSpan[35,36], we found that the DE-ASD network
110 was enriched for genes that are strongly expressed in the neocortex at prenatal and early postnatal periods
111 (p-value $<4.3\times10^{-30}$; Figure 1.c).
112 To investigate the spatiotemporal activity pattern during brain development, we measured the co-
113 expression strength of interactions in the leukocyte-based network at different neurodevelopmental time
114 windows across brain regions using BrainSpan. We found that the highest co-expression activity of the DE-
115 ASD network temporally coincided with peak neural proliferation in brain development (10-19 post
116 conception weeks [3,8]) across the brain and then decreased in activity at later time points (Figure 1.d).
117 Further supporting the transcriptional activity of the blood-derived network in prenatal brain, we found
118 evidence that the DE-ASD network is preserved at the co-expression level between ASD blood and prenatal
119 brain (Figure 1.e).

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

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

With the observed overlap patterns, we next examined the potential 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. This analysis demonstrated that rASD genes are not enriched in the DE-ASD network (p-value >0.19; Methods).

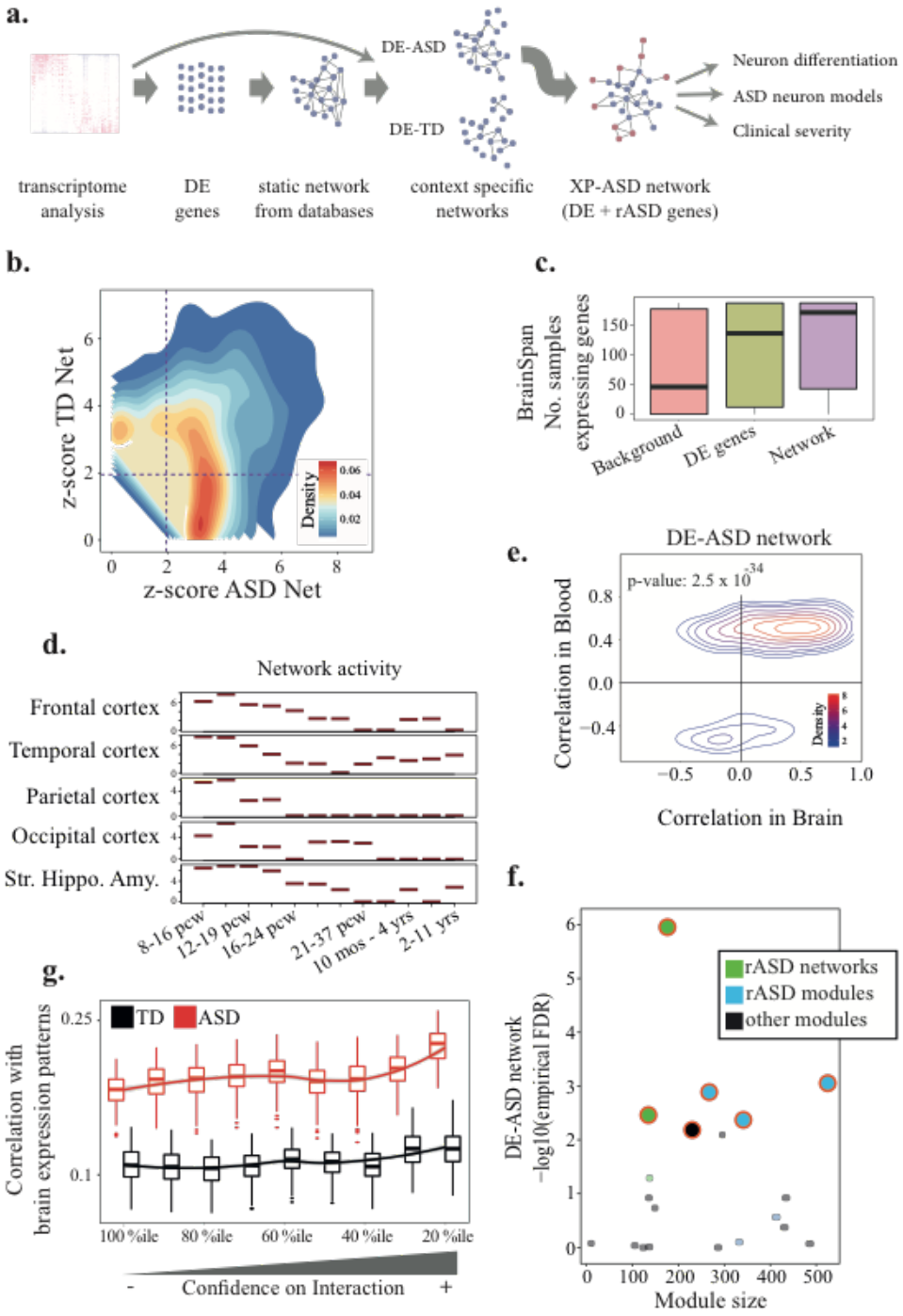

Fig 1. Elevated co-expression activity of the DE-ASD network in ASD leukocytes and its preservation in prenatal brain.
a) Overview of this study. Transcriptome analysis of 226 ASD and TD toddlers identified 1236 DE genes. We built a comprehensive static network of DE genes from high confidence physical and regulatory interactions. The static network was next pruned using measured gene co-expression within the ASD and TD transcriptome data to obtain context specific DE-ASD and DE-TD networks, separately, to compare the strength of gene co-expression 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) Interacting DE genes are considerably more strongly co-expressed in the ASD toddlers. The plot illustrates the distribution of co-expression strength for a merged network composed of interactions that were significant and present in at least one of DE-ASD or DE-TD networks. 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 RPKM data. 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. 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 (Methods). The y-axis indicates the z-transformed p-value of this comparison. e) Blood 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 were calculated based on neocortex transcriptome data from 8 post conception weeks until the postnatal age of 1 year old. 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. see 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; rASD modules: co-expression modules enriched for rASD genes; other modules: modules that are not enriched for rASD genes. g) Similarity of interactions of a brain co-expression network around rASD genes 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 Figure S3 for more details.

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

138 High confidence rASD genes are profoundly enriched for genes with regulatory functions [3,7,11,18].
139 Although the perturbed DE-ASD network is not enriched for rASD genes, it overlaps with co-expression
140 modules and networks of known rASD genes. At the mechanistic level, the observed co-expression of
141 rASD and DE genes in the prenatal brain could be due to the regulatory influence of rASD genes on the
142 DE-ASD network, and thereby mutations in rASD genes could cause the network over-activity and brain
143 maldevelopment in ASD.
144
145
146
147

148

149

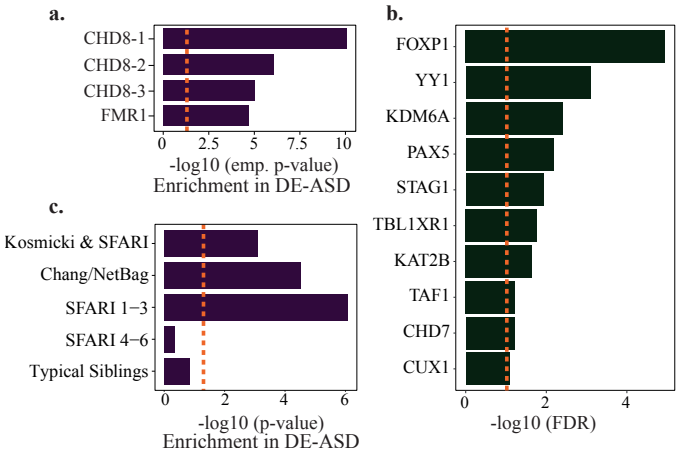

Fig 2. rASD genes are enriched for the regulators of the DE-ASD network.
a) Genes identified as regulatory targets of CHD8 (CHD8-1: Sugathan et al.; CHD8-2: Gompers et al; CHD8-3: Cotney et al.) and FMR1, two high confidence rASD genes, are enriched in the DE-ASD network. The target genes were previously identified by ChIP-Seq experiments and the enrichment was assessed empirically (Methods). b) Several rASD genes have their regulatory targets significantly enriched among the DE–ASD network based on the ENCODE data. Here, each cell represents one experiment on the corresponding factor. The significance of overlap with DE-ASD was measured by hypergeometric test and 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, Kosmicki et al. and Chang et al. The compiled list of mutations among typical siblings of ASD individuals in four genetic studies were extracted from Chang et al. study.

150

151

152

153

154

155 To elucidate if rASD genes could regulate the DE-ASD network, we first examined if the regulatory
156 targets of rASD genes are enriched in the DE-ASD network. Indeed, we observed that the DE-ASD
157 network is enriched for genes potentially regulated by two high confidence rASD genes, CHD8 [37-39] and
158 FMR1 [40](Figure 2.a). To more systematically identify potential regulators of the network, we examined the
159 overlap of the DE-ASD network with the regulatory targets identified across 845 assays as part of the
160 ENCODE project [41] and 615 assays manually curated and deposited at Chea2016 database. Strikingly, we
161 observed DE-ASD network is significantly enriched for 11 out of 20 high confidence and suggestive
162 confidence rASD genes (OR: 2.54; p-value: 0.05; Figure 2.b; Table S3).

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

164 The rASD genes were often not differentially expressed in ASD leukocytes, and therefore the DE-
165 ASD network was not enriched in rASD genes. However, to explore if rASD genes could regulate the DE-
166 ASD network, we expanded the DE-ASD network by including rASD genes. Thus, we obtained an
167 expanded-ASD, XP-ASD, network (Table S4). For this, we used a similar approach to that of the DE-ASD
168 network. We first curated a high confidence static network of DE and 965 speculated rASD genes. The
169 context-specific XP-ASD network was next inferred by retaining only the significantly co-expressed
170 interacting pairs in ASD samples. This pruning step results in removal of genes from the static network that

171 do not show significant co-expression patterns with their known partners or regulatory targets in ASD
172 leukocytes. Accordingly, the XP-ASD network included a total of 316 out of 965 (36%) likely rASD genes.
173 Our list of 965 rASD genes included rASD genes of both high confidence (e.g., recurrently mutated
174 in ASD individuals) and low confidence (some even found in typical siblings of ASD individuals). We
175 reasoned that if the XP-ASD network is truly relevant to the prenatal etiology of ASD, a preferential
176 incorporation of high confidence rASD genes would be expected in the leukocyte-derived XP-ASD
177 network. By following different analytical methods, different groups have separately categorized rASD
178 genes into high and low confidence [7,16,42]. Importantly, we found a reproducible enrichment of high
179 confidence rASD genes in the XP-ASD network (Figure 2.c) with a significant enrichment for strong
180 evidence rASD genes with *de novo* protein truncating variants in ASD subjects (p-value $<3.06\times10^{-6}$;
181 hypergeometric test). Intuitively, the XP-ASD network did not show enrichment for rASD genes that were
182 also detected in TD siblings of ASD subjects (p-value >0.14) or are currently ranked as low confidence (p-
183 value >0.44). Further corroborating the regulatory role of rASD genes on DE-ASD network, we found a
184 significant enrichment of the XP-ASD network for rASD genes with DNA binding activities (OR: 3.1; p-
185 value $<2.1\times10^{-12}$; Figure S7). This preferential addition of high confidence and regulatory rASD genes
186 supports the relevance of the XP-ASD network to the pathobiology of ASD, and the likelihood that the high
187 confidence rASD genes are regulating the DE-ASD network.

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

189 To explore the regulatory effect of the rASD genes on the DE genes, we analyzed their interaction
190 types (i.e., positive or negative correlations, alluding to activator or repressor activity). Comparative
191 analysis of DE- and XP-ASD networks indicated a significant enrichment of negative correlations between
192 rASD and DE genes (p-value $<3.1\times10^{-4}$; Fisher's exact test), suggesting more of an inhibitory role of rASD
193 genes on the DE genes (Figure 3.a).
194 Supporting the inhibitory role of rASD genes, the DE-ASD network was enriched for genes that
195 were up-regulated by the knock-down of CHD8 in neural progenitor and stem cells; but not for those that
196 were down-regulated (Figure 3.b) [37-39]. Consistently, we observed an overall up-regulation of genes that are
197 up-regulated in CHD8 knock-down background (p-value <0.039 across three different datasets; GSEA), but
198 not for those that are down-regulated. We observed a similar up-regulation pattern for the binding targets of
199 FMR1 rASD gene (p-value <0.078; GSEA) in the ASD transcriptome. The potential inhibitory role of
200 rASD genes on the DE-ASD network was further supported in an independent dataset on neural
201 differentiation. Specifically, we observed an anti-correlated expression pattern between the rASD and the
202 DE genes from the XP-ASD network in *in vitro*-differentiated human neural progenitors (Figure 3.c).

Signaling pathways are central to the leukocyte-based networks

We next sought to identify 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 (Figure 4.a; Table S5). Moreover, the DE-ASD network was highly enriched for PI3K/AKT, mTOR, and related pathways (Figure 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, response to oxidative stress, and FC receptor signaling (Figure 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 (Figure 4.a).

Our results suggest up-regulation and elevated co-expression activity of PI3K/AKT and its down-stream pathways in ASD leukocytes (Figure 4.a-b). These processes are involved in brain development and

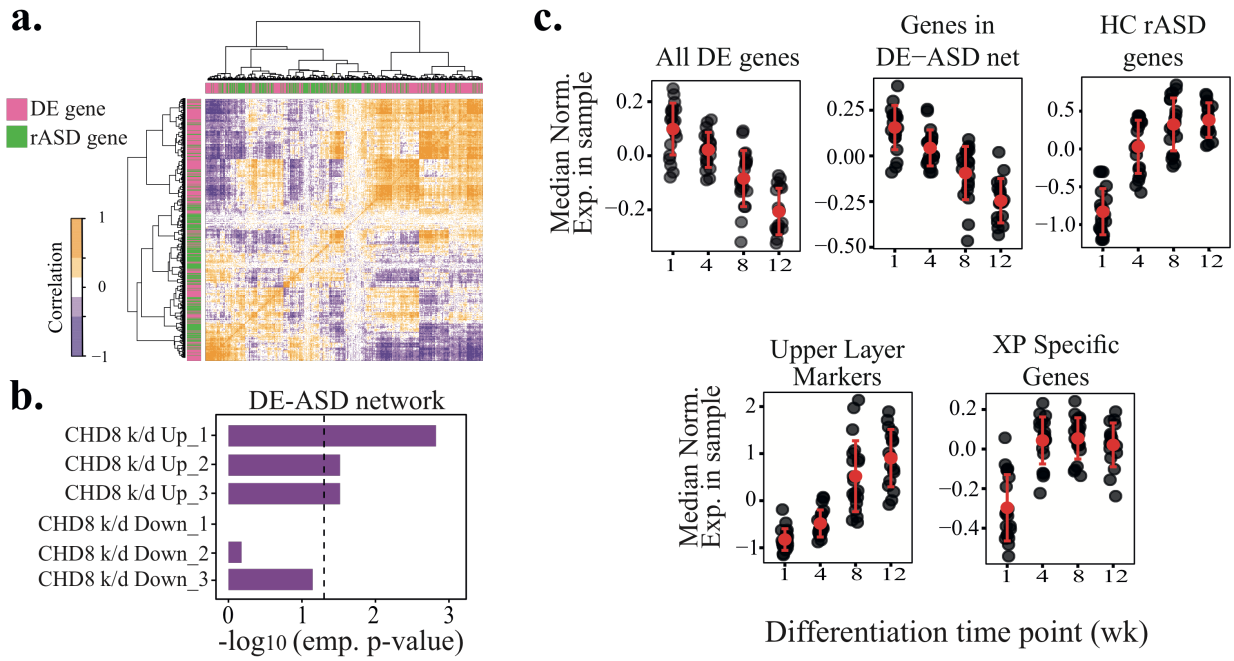

Fig 3. rASD genes potentially suppress the DE genes.
a) The expression of rASD genes is negatively correlated with the DE genes in leukocyte transcriptome data. See Figure S5 for more details. b) The DE-ASD network is significantly enriched for genes that are up-regulated in response to the knock-down of CHD8 gene. Data were extracted from two studies of Sugathan et al (CHD8 k/d_1), Gompers et al (CHD8 k/d_2), and Cotney et al (CHD8 k/d_3). 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. In each panel, the gray circles represent the median expression pattern of associated genes. Gene expressions were 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 the neuron differentiation process (p-value = 0.007), rASD genes show a significant up-regulation pattern (p-value = 1.12×10^{-5}). The expression levels of three genes of CACNA1E, PRSS12, and CARTPT were considered as the markers of upper layer neurons (late stage of neural differentiation). See Figure S5 for the related supplementary figures.

growth during prenatal and early postnatal ages [3,43,44] and focused studies on rASD genes have implicated them in ASD [3,8,45,46]. Further supporting the over-activity of the PI3K/AKT and its down-stream pathways in 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 (Methods).

We further investigated the DE-ASD and XP-ASD networks using an integrated hub analysis approach (Methods). In the DE-ASD network, 53% of hub genes were involved in the PI3K/AKT pathway including PIK3CD, AKT1 and GSK3B (Figure 4.c). Moreover, 72% of hubs in the DE-ASD network were also hub in the XP-ASD network (Figure S6). Genes that were hub only in the XP-ASD network were highly enriched for the regulatory genes associated with neuronal proliferation and maturation processes including members of the RAS/ERK (e.g., ERK2, ERK1, RAF1), PI3K/AKT (e.g., PTEN, RHOG, ERBB2,

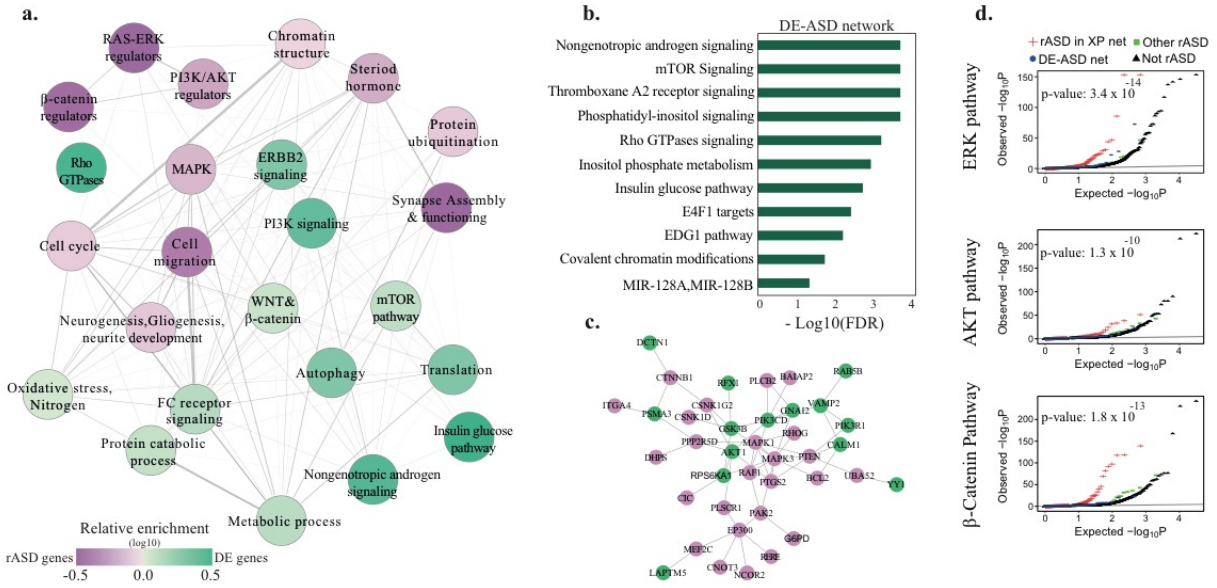

Fig 4. The architecture of the XP-ASD network.
a) Summary of enriched biological processes in the XP-ASD network. Each node represents a biological process that is significantly enriched in the XP-ASD network (Fisher's exact test; Table S5). Nodes that preferentially include rASD and DE genes are represented by purple and green colors, respectively. The interactions among terms represent the connection patterns of their cognate genes in XP-ASD network with thicker interactions indicating more significant connections (hypergeometric test). Only connections with p-value <0.05 are represented. This illustration covers 83% 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. Purple and green color represent hubs in the XP-ASD and DE-ASD networks, respectively. 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. Background is composed of all genes that were assayed in Brockmann et al, 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 Figure S6 for more details.

228 EP300), and WNT/β-catenin (e.g., CTNNB1, SMARCC2, CSNK1G2) pathways (Figure 4.c; Table S6).
229 While PI3K/AKT (hub in DE-ASD and XP-ASD networks) promotes proliferation and survival, the ERK
230 pathway (hub in the XP-ASD network) triggers differentiation of neural progenitor cells [3,47-49].

231 **rASD genes regulate DE-ASD genes through specific signaling pathways**

232 We further explored if perturbation to the rASD genes lead to the perturbation of the DE-ASD
233 network through changes in the RAS/ERK, PI3K/AKT, and WNT/β-catenin pathways. To assess this, we
234 leveraged genome-wide mutational screening data in which gene mutations were scored based on their
235 effects on the activity of the RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways [50]. The
236 activity of the signaling pathways was measured based on the phosphorylation state of ERK, AKT, and β-
237 catenin proteins [50]. Consistent with functional enrichment and hub analysis results, we found that rASD
238 genes in the XP-ASD network are significantly enriched for the potential regulators of RAS/ERK,
239 PI3K/AKT, and WNT/β-catenin pathways (Figure 4.d; p-value $<1.9\times10^{-10}$; Wilcoxon-Mann-Whitney test).
240 Specifically, regulators of these pathways (FDR<0.1) accounted for inclusion of 51% rASD genes in the
241 XP-ASD. As the control, no significant enrichment for the regulators of RAS/ERK, PI3K/AKT, and
242 WNT/β-catenin pathways were observed among rASD genes that were not included in the XP-ASD
243 network (Figure 4.d). These results support the regulatory role of rASD genes on the DE-ASD network
244 through perturbation of RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways.

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

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

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

256 To further validate these results, we first examined if the DE-ASD network is over-active in iPS
257 cell-derived neural progenitors and neurons of ASD toddlers compared to those of TD cases. For this
258 analysis, we considered the available transcriptome data from Marchetto et al study where fibroblast cells
259 from 8 ASD individuals with macrocephaly and 6 TD individuals were reprogrammed to iPS cells [30]. The

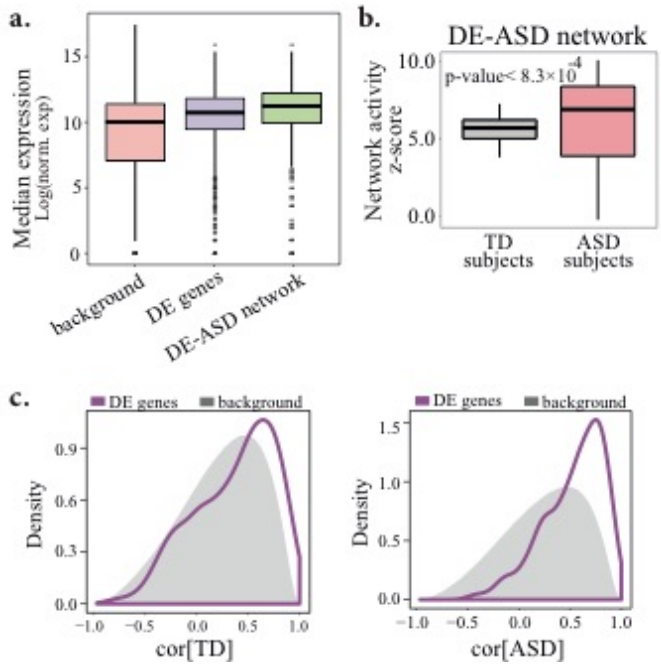

Fig 5. The DE-ASD network is over-active in differentiating neurons of ASD cases.
a) The DE-ASD network is more highly expressed during neural differentiation of iPSCs from ASD and TD cases, as measured by RNA-Seq. Median expression of the genes at neural progenitors and neurons stages were considered. 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.

generated iPS cells were next differentiated into the neural progenitor and then neuron stages. Analysis of the DE-ASD at neural stages demonstrated that the network is indeed strongly expressed and over-active in these ASD neuron models (Figure 5), suggesting the functional relevance of identified blood molecular signatures to the abnormal ASD brain development.

Network dysregulation is associated with ASD severity

We sought to directly assess 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 exists 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 S8). 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 deficits, the ASD diagnostic gold standard (Fig 6). To assess the significance of observed correlation patterns, we repeated the analysis with 10,000 random shuffling of the ADOS social affect scores of ASD individuals (The inset boxplots in Fig 6). This analysis demonstrated the significance

of the observed correlations (Figure 6). Therefore, the discovered molecular signature connects the genetic bases of autism to transcriptional changes and shows diagnostic and prognostic potential as it correlates with social and communication severity on ASD individuals.

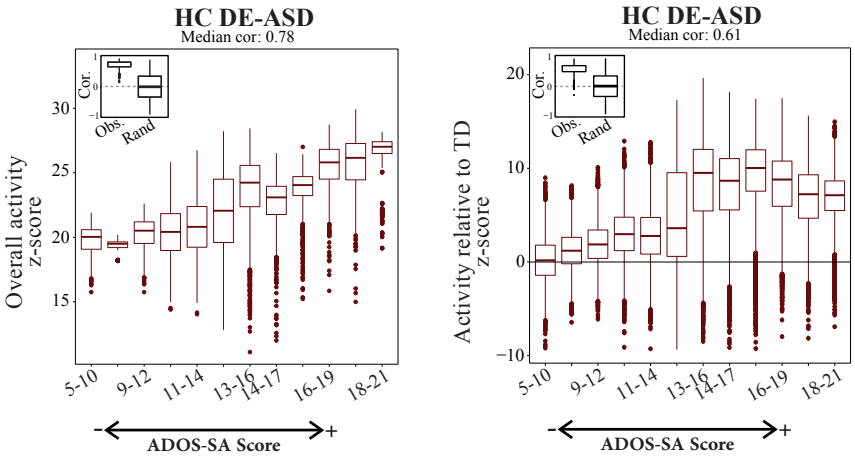

Fig 6. Activity level of DE-ASD networks are correlated with ASD severity.
a) ASD toddlers were sorted based on their ADOS Social Affect scores (ADOS-SA) with higher scores representing more severe cases. The network activity was measured in a running window on ADOS-SA scores. The overall activity of 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. The background was defined as the co-expression distribution of genes with the same mean expression level as genes in the DE-ASD network in samples under investigation (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 with ADOS severity levels and compares it with those after randomization (i.e., expected by chance). The randomization was based on 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. The randomization was conducted by 10,000 random shuffling 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 size on the etiology but occur in a small percentage of the ASD population [51,52]. We

276 show that many rASD genes may exert their regulatory effect on this DE-ASD core network through the
277 inter-connected RAS/ERK, PI3K/AKT, and WNT/β-catenin signaling pathways. The connection of the DE-
278 ASD network (that is constructed based on data drawn from the general ASD pediatric population) with
279 high confidence rASD genes provides empirical evidence of shared mechanisms underlying ASD in both
280 those with high penetrant rASD genes and those of other etiologies (e.g., common variants, environmental
281 factors) in the wider ASD population.

282 The key aspect of our signature is that it is constructed based on transcriptomic data from young
283 living ASD toddlers. This allows us to correlate its variations with the core clinical features of the same
284 ASD toddlers. Indeed, the dysregulation degree of DE-ASD network correlated with deficits in the
285 toddlers' ADOS social affects. Social and behavioral deficits are also suggested to be correlated with the
286 genetic variations in ASD subjects [53,54]; and previous studies have established the effect of the PI3K/AKT
287 signaling pathway (central to the DE-ASD core network and significantly altered in ASD leukocytes) on
288 social behaviors of mouse models [45,46]. Together, these observations suggest that etiological roots of ASD
289 converge on gene networks that correlate with the symptom severity in ASD individuals. Moreover, our
290 results reinforce the hypothesis that stronger dysregulation of the same core network could lead to higher
291 severity in the ASD cases. The DE-ASD core network is enriched for pathways implicated in ASD,
292 strongly associated with high confidence rASD genes, and correlate with ASD severity. However, we note
293 that a direct causal relationship between the co-expression activity of the network and ASD remains to be
294 established. Moreover, our network co-expression activity measure is a summary score from the strongest
295 signal in our dataset (i.e., differentially expressed genes) at a group level (i.e., severity level). Therefore, by
296 design, it may not capture the heterogeneity that could exist within each group. As detailed below, future
297 work is needed to explore the causal relationship of our gene network to ASD development, symptoms, and
298 the potential existence of other dysregulation mechanisms in ASD individuals.

299 The emerging architecture of complex traits suggests that gene mutations often propagate their
300 effects through regulatory networks and converge on core pathways relevant to the trait [22,55]. Our findings
301 support the existence of an analogous architecture for ASD, wherein rASD genes with diverse biological
302 roles overlap in their down-stream function. Although not significantly overlapping with rASD genes, we
303 found that the DE-ASD network is significantly co-expressed with rASD genes in both blood and brain
304 tissues. We also illustrated that the DE-ASD network could be controlled by rASD genes through direct
305 transcriptional regulation or highly interconnected signaling pathways. We postulate that the DE-ASD
306 network is a primary convergence point of ASD etiologies, including its genetic basis as we elaborated for
307 rASD genes, in a large portion of the ASD population. This predicts that the spectrum of autism in such
308 cases is correlated with the degree and mechanism of the perturbation of the DE-ASD network. A detailed
309 analysis of iPS cell-derived ASD neurons demonstrated the dysregulation of the leukocyte-based DE-ASD

310 network in ASD neurons, supporting the neural-level relevance of the findings to ASD etiology and its
311 prevalence in the ASD population. Furthermore, direct clinical-level relevance is demonstrated by the high
312 correlation we found between degree of dysregulation in the DE-ASD core network and ASD symptom
313 severity in the ASD toddlers.

314 The currently recognized rASD genes are not fully penetrant to the disorder, except for a handful of
315 syndromic genes [51,52,56,57]. Our analysis of the XP-ASD network provides some insights on how the effects
316 of rASD genes could potentially combine to result in ASD. Although some rASD genes could directly
317 modulate the DE-ASD network at the transcriptional level, our results suggest that the regulatory
318 consequence of many rASD genes on the DE-ASD network is canalized through the PI3K/AKT,
319 RAS/ERK, WNT and β-catenin signaling pathways. The structural and functional interrogation of the DE-
320 ASD network localized the PI3K/AKT pathway to its epicenter and demonstrated enrichments for
321 processes down-stream of this pathway. Moreover, we found that high confidence rASD genes are better
322 connected to the DE-ASD core network, suggesting that the closeness and influence of genes on these
323 signaling pathways is correlated with their effect size on the disorder. These results articulate that
324 perturbation of the PI3K/AKT, RAS/ERK, WNT and β-catenin signaling pathways through gene regulatory
325 networks may be an important etiological route for ASD that could be associated with the disorder severity
326 level in a relatively large fraction of the ASD population. Congruent with this hypothesis, cell and animal
327 models of ASD have demonstrated the enrichment of high confidence rASD genes for the regulators of the
328 RAS/ERK, PI3K/AKT, WNT and β-catenin signaling pathways [3,8,11,18,45,46,49]. These signaling pathways are
329 highly conserved and pleiotropic, impacting multiple prenatal and early postnatal neural development
330 stages from proliferation/differentiation to synaptic and neural circuit development [3]. Such multi-
331 functionalities could be the underlying reason that we detected the signal in ASD leukocytes.

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

344 actualized via approaches that illuminate the early-age living biology of ASD [3,18,21]. ASD toddler-derived
345 iPS cell studies show ASD is a progressive prenatal and early postnatal disorder that involves a cascade of
346 diverse and varying molecular and cellular changes such as those resulting from dysregulation of the
347 pathways and networks highlighted herein [3,30,31]. As such, dynamic, individual-based molecular assays in
348 infants and toddlers will be essential to develop. The presented framework could prove invaluable for the
349 development of quantitative, molecular-based measures for the ASD diagnosis and prognosis by identifying
350 specific molecular dysregulations that we show are observable in leukocytes of a large fraction of living
351 ASD toddlers at young ages.

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* (in press).
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).

401 26. Hu, V.W., Frank, B.C., Heine, S., Lee, N.H. & Quackenbush, J. Gene expression profiling of
402 lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals
403 differential regulation of neurologically relevant genes. *BMC Genomics* **7**, 118 (2006).
404 27. Hu, V.W. *et al.* Gene expression profiling differentiates autism case-controls and phenotypic
405 variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism.
406 *Autism Res* **2**, 78-97 (2009).
407 28. Kong, S.W. *et al.* Characteristics and predictive value of blood transcriptome signature in males
408 with autism spectrum disorders. *PLoS One* **7**, e49475 (2012).
409 29. Diaz-Beltran, L. *et al.* Cross-disorder comparative analysis of comorbid conditions reveals novel
410 autism candidate genes. *BMC Genomics* **18**, 315 (2017).
411 30. Marchetto, M.C. *et al.* Altered proliferation and networks in neural cells derived from idiopathic
412 autistic individuals. *Mol Psychiatry* **22**, 820-835 (2017).
413 31. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in
414 Autism Spectrum Disorders. *Cell* **162**, 375-390 (2015).
415 32. Ideker, T. & Krogan, N.J. Differential network biology. *Mol Syst Biol* **8**, 565 (2012).
416 33. Yang, B. *et al.* Dynamic network biomarker indicates pulmonary metastasis at the tipping point of
417 hepatocellular carcinoma. *Nat Commun* **9**, 678 (2018).
418 34. Chen, L., Liu, R., Liu, Z.P., Li, M. & Aihara, K. Detecting early-warning signals for sudden
419 deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* **2**, 342 (2012).
420 35. BrainSpan. BrainSpan: Atlas of the Developing Human Brain. (2016).
421 36. Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-9 (2011).
422 37. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum
423 disorder in neural progenitors. *Proc Natl Acad Sci U S A* **111**, E4468-77 (2014).
424 38. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes
425 during human neurodevelopment. *Nat Commun* **6**, 6404 (2015).
426 39. Gompers, A.L. *et al.* Germline Chd8 haploinsufficiency alters brain development in mouse. *Nat*
427 *Neurosci* **20**, 1062-1073 (2017).
428 40. Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and
429 autism. *Cell* **146**, 247-61 (2011).
430 41. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**,
431 57-74 (2012).
432 42. Abrahams, B.S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism
433 spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
434 43. Hur, E.M. & Zhou, F.Q. GSK3 signalling in neural development. *Nat Rev Neurosci* **11**, 539-51
435 (2010).
436 44. Lipton, J.O. & Sahin, M. The neurology of mTOR. *Neuron* **84**, 275-91 (2014).
437 45. Clipperton-Allen, A.E. & Page, D.T. Pten haploinsufficient mice show broad brain overgrowth but
438 selective impairments in autism-relevant behavioral tests. *Hum Mol Genet* **23**, 3490-505 (2014).
439 46. Cupolillo, D. *et al.* Autistic-Like Traits and Cerebellar Dysfunction in Purkinje Cell PTEN Knock-
440 Out Mice. *Neuropsychopharmacology* **41**, 1457-66 (2016).
441 47. Chen, Y., Huang, W.C., Sejourne, J., Clipperton-Allen, A.E. & Page, D.T. Pten Mutations Alter
442 Brain Growth Trajectory and Allocation of Cell Types through Elevated beta-Catenin Signaling. *J*
443 *Neurosci* **35**, 10252-67 (2015).
444 48. Rhim, J.H. *et al.* Cell type-dependent Erk-Akt pathway crosstalk regulates the proliferation of fetal
445 neural progenitor cells. *Sci Rep* **6**, 26547 (2016).
446 49. Mellios, N. *et al.* MeCP2-regulated miRNAs control early human neurogenesis through differential
447 effects on ERK and AKT signaling. *Mol Psychiatry* (2017).
448 50. Brockmann, M. *et al.* Genetic wiring maps of single-cell protein states reveal an off-switch for
449 GPCR signalling. *Nature* **546**, 307-311 (2017).
450 51. Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and
451 genomic disorders and still counting. *Brain Res* **1380**, 42-77 (2011).

452 52. Huguet, G., Ey, E. & Bourgeron, T. The genetic landscapes of autism spectrum disorders. *Annu Rev*
453 *Genomics Hum Genet* **14**, 191-213 (2013).
454 53. Robinson, E.B. *et al.* Genetic risk for autism spectrum disorders and neuropsychiatric variation in
455 the general population. *Nat Genet* **48**, 552-5 (2016).
456 54. Wang, Y. *et al.* Heritable aspects of biological motion perception and its covariation with autistic
457 traits. *Proc Natl Acad Sci U S A* (2018).
458 55. Califano, A. & Alvarez, M.J. The recurrent architecture of tumour initiation, progression and drug
459 sensitivity. *Nat Rev Cancer* **17**, 116-130 (2017).
460 56. Amir, R.E. *et al.* Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-
461 binding protein 2. *Nat Genet* **23**, 185-8 (1999).
462 57. Neale, B.M. *et al.* Patterns and rates of exonic de novo mutations in autism spectrum disorders.
463 *Nature* **485**, 242-5 (2012).