bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Title:** Hierarchical cortical transcriptome disorganization in autism

**Authors:** Michael V. Lombardo[1,2,3], Eric Courchesne[4], Nathan E. Lewis[5,6], & Tiziano Pramparo[4]

**Affiliations**

[1] Centre for Integrative Neuroscience and Neurodynamics, School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[2] Center for Applied Neuroscience, Department of Psychology, University of Cyprus, Nicosia, Cyprus

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

[4] Department of Neuroscience, University of California, San Diego, San Diego CA, USA

[5] Department of Pediatrics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA, USA

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

**Corresponding Author**: Michael V. Lombardo (mvlombardo@gmail.com)

**Running title:** ASD hierarchical transcriptome disruption

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Abstract

Autism spectrum disorder (ASD) is genetically heterogeneous and complex. Functional genomics work has begun to identify specific dysregulated transcriptomic pathways. However, it remains unclear whether such dysregulation reveals evidence for independent or higher-level coordinated systems-level pathology. Here we find replicable evidence across 2 datasets for 10 gene co-expression modules that are differentially expressed in ASD cortex. Rather than being distinct non-interacting pathology, these modules work in synergy and interact at the protein level. This systems-level pathology is characterized by downregulated synaptic and neural developmental processes and upregulated catabolism, viral processes, translation, protein targeting and localization, interferon signaling, glia-relevant, and apoptosis processes. Hierarchical organization of meta-modules (clusters of highly correlated modules) is also highly affected in ASD. These results support a new viewpoint of pathophysiology affecting ASD characterized by multiple coordinated dysregulated transcriptomic processes that produce emergent systems-level pathology and which may be overlooked by focusing on dysregulated elements in isolation.

**Keywords:** autism / immune / synapse / transcriptome / translation

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The pathophysiology behind atypical brain development in autism spectrum disorders (ASD) is highly complex. Elegant genetics work is continually unveiling an ever more diverse array of biological mechanisms associated with ASD (Geschwind and Levitt, 2007; Geschwind and State, 2015). With such diversity, a key question arises as to whether such mechanisms point to many independent disrupted pathways or some convergence on a few common pathways affecting large-scale biological systems and/or interactions between such systems (Gokoolparsadh et al., 2016). One way to test this question is to examine pathophysiology at a level above genetics, such as the transcriptome, and examine whether the diversity of disrupted transcriptomic signals converge onto independent or interacting systems.

Past work examining systems-level cortical transcriptome dysregulation has highlighted two important gene modules (i.e. collections of genes whose expression levels are highly correlated) that are dysregulated in ASD. The first module is downregulated in expression in ASD and enriched for synaptic processes and neuronal markers, while the second module is upregulated in ASD and enriched for immune/inflammation processes and astrocyte and M2 microglia activation state markers (Gupta et al., 2014; Voineagu et al., 2011). While this work was seminal in furthering our understanding of some specific pathways dysregulated in the ASD cortical transcriptome, it is unclear if the pathways are independently dysregulated. Pointing towards the idea that such modules may not be independent, Gupta and colleagues found negative correlations between these modules when collapsing data across both groups (Gupta et al., 2014). However, this observation does not clearly point to converging hierarchical systems-level pathology because it critically does not test whether the groups differed in their interactions between each other. If the ASD group differed in the relationship between such modules, this would point towards transcriptome dysregulation that extends beyond the level of single modules, and involves dysregulation spanning interactions between larger pathological processes. To further solidify evidence of dysregulated systems-level pathology resulting from dysregulation between disparate modules, the evidence of statistical dependency in correlations between co-expression modules would require evidence of physical interactions between the protein products of such modules.

In this work we test the hypothesis that diverse molecular mechanisms are hierarchically disrupted in the cortical transcriptome of ASD and point towards interacting systems-level pathology rather than multiple independent types of pathology in synaptic and immune processes. Specifically, we hypothesize that dysregulated gene co-expression modules may work in synergy to form emergent pathology not visible by looking at single modules in isolation. We predict that differentially expressed modules will be highly correlated and that such correlations may be increased in ASD. Going beyond correlations between co-expression modules, we further predict that if such systems are interacting we would see strong evidence for physical interactions in a protein-protein interaction analysis. Furthermore, given the emerging literature on neuronal/synaptic-immune interactions (Choi et al., 2016; Coiro et al., 2015; Elmer et al., 2013; Glynn et al., 2011; Le Belle et al., 2014; Oskvig et al., 2012; Stephan et al., 2012; Stevens et al., 2007), we also hypothesized that specific dysregulated modules enriched in immune/inflammation and synaptic processes will be aberrantly connected in ASD compared to Controls. This work represents the first study examining hierarchical disruption of the cortical

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

transcriptome in ASD at both the level of single dysregulated gene modules, but also dysregulation in higher-level interactions between modules. We provide the first look at the full organization of correlations between gene modules across the transcriptome (i.e. eigengene networks) and examine how such connections manifest differently both at the level of inter-modular connectivity (i.e. connections between specific modules) as well as connectivity relevant to organization of clusters of highly correlated modules (i.e. meta-modules) (Langfelder and Horvath, 2007; Oldham et al., 2006; Oldham et al., 2008). Subtle and specific changes in eigengene network organization or global patterns of network reorganization are both plausible predictions regarding how eigengene networks are organized differently in ASD. Both scenarios would lead to the prediction that the composition of meta-modules as well as connectivity within and outside of normative meta-module boundaries would differ in ASD.

## Materials and Methods

### *Datasets*

We re-analyzed two existing datasets probing cortical gene expression in ASD. The first dataset utilized microarrays on frontal (BA9; n =16 ASD; n = 16 Controls) and temporal cortex (BA 41/42; n = 13 ASD; n= 13 Controls) tissue and was first described by Voineagu and colleagues (Gene Expression Omnibus (GEO) Accession ID: GSE28521) (Voineagu et al., 2011). The second dataset utilized RNAseq on frontal (BA10, n = 6 ASD, n = 8 Controls; BA44, n = 16, n = 11 Controls) and occipital cortex (BA19, n = 24 ASD, n = 38 Controls) tissue and was first described by Gupta and colleagues (http://www.arkinglab.org/resources/) (Gupta et al., 2014). These datasets were selected because they were relatively the largest studies in the literature. For each dataset we utilized the already pre-processed and quality controlled datasets publicly available in order to be as congruent as possible with prior published work. For genes with multiple probes in the Voineagu dataset we selected the probe with the highest mean expression value across the full dataset using the collapseRows function in R (Miller et al., 2011). Within the Gupta dataset, missing values were present for some genes in some subjects and these missing values were imputed using the impute.knn function within the impute R library. This procedure was done in order to maximize the total number of genes possible for inclusion into further WGCNA analysis. All further analyses utilize a subset of the 8,075 genes that were common across both datasets.

### *Weighted Gene Co-Expression Network Analysis (WGCNA)*

Co-expression analysis was implemented with the WGCNA package in R (Langfelder and Horvath, 2008). A consensus WGCNA analysis was implemented in order to detect consensus modules for cross-dataset comparisons (implemented with the blockwiseConsensusModules function) (Langfelder and Horvath, 2007). Consensus WGCNA analysis consisted of construction of correlation matrices, which were then converted into adjacency matrices that retain information about the sign of the correlation (i.e. signed networks use a transformation of 0.5*(r+1)). Adjacency matrices were raised to a soft power threshold selected based on an analysis across various soft power thresholds and choosing the soft power threshold based on a measure of $R^2$ scale-free topology model fit that maximized and plateaued well above 0.8 (i.e. soft power = 14 for both datasets; see Fig S1). Soft power thresholded

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

adjacency matrices were then converted into a topological overlap matrix (TOM) and a TOM dissimilarity matrix (i.e. 1-TOM). The TOM dissimilarity matrix was then input into agglomerative hierarchical clustering using the average linkage method. Gene modules were defined from the resulting clustering tree and branches were cut using a hybrid dynamic tree cutting algorithm (deepSplit = 2) (Langfelder et al., 2008). Modules were merged at a cut height of 0.2 and the minimum module size was set to 30. For each gene module a summary measure called the module eigengene (ME) was computed as the first principal component of the scaled (standardized) module expression profiles. Genes that cannot be clustered into any specific module are left within the M0 module, and this module is not considered in any further analyses.

To test for differential expression at the level of ME variation we used linear mixed effect models implemented with the lme function in the nlme R library. These models included diagnosis as the fixed effect of interest and additionally included age, sex, RIN, PMI, and median 5' to 3' prime bias (specific to Gupta dataset) as fixed effect covariates. Brain region was also included in each model as a within-subject random effect modeled with random intercepts to account for the correlation of multiple brain regions from the same individual. To identify MEs with replicable differential expression across both datasets, we utilized t-statistics from the linear mixed models to compute replication Bayes Factor (repBF) statistics (Verhagen and Wagenmakers, 2014) that quantify evidence for or against replication (see here for R code: http://bit.ly/1GHiPRe). Replication Bayes Factors greater than 10 are generally considered as strong evidence for replication. To identify replicable modules we first considered modules that possessed a significant effect passing FDR (Storey, 2002) $q<0.05$ within the Voineagu dataset and then also required these modules possess significant effects in the Gupta dataset (FDR $q<0.05$) and that this evidence quantitatively produces evidence for replication with a replication Bayes Factor statistic > 10.

### *Process Level Gene Set Enrichment Analyses*

To characterize specific biological processes for all modules, we performed process level (i.e. Process Networks) enrichment analyses within the MetaCore GeneGO software platform. To identify emergent processes from collections of highly correlated dysregulated modules we used GO biological processes enrichment analysis (AmiGO 2; http://amigo.geneontology.org/) in order to leverage GO's relatively broader hierarchical structure (compared to MetaCore GeneGO). REVIGO (Supek et al., 2011) was then utilized on the top 50 GO terms ranked by fold enrichment in order to assist in reducing the large number of GO terms into semantically similar clusters of terms. We manually edited the REVIGO output by inserting custom descriptive terms for each cluster and to correct for obvious errors in semantic clustering (e.g., a term like 'synaptic organization' occurring outside of the synaptic cluster).

### *Cell Type/Cellular Compartment Enrichment Analyses*

To characterize differentially expressed modules by enrichments in specific cell types (neuron, astrocyte, oligodendrocyte, M1 and M2 microglia states), and cellular compartments (synapse, postsynaptic density, ribosomal subunits), we utilized lists of markers previously used by Gupta and colleagues (Gupta et al., 2014). The exception to this was lists of ribosomal subunit markers. These were obtained

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

from lists contained in GO. Enrichment tests were hypergeometric tests (i.e. sum(dhyper()) in R) using the total number of genes (8,075) as the background pool total.

### *Eigengene Network Analysis*

Eigengene network analysis proceeded by constructing robust ME partial correlation matrices separately for each group. These matrices were computed in MATLAB using robust regression to be insensitive to outliers (Wager et al., 2005) and the robust regression models incorporated the removal of variation from nuisance covariates (i.e. age, sex, RIN, PMI, median 5' to 3' bias, brain region). Partial correlation matrices were then converted into adjacency matrices that retain information about the sign of the correlation. ME adjacency matrices were converted into topological overlap dissimilarity matrices (1-TOM) and then were inserted into agglomerative hierarchical clustering using the ward.D linkage method. The resulting cluster tree was then cleaved into meta-modules using the same dynamic hybrid tree cutting algorithm utilized in WGCNA. We used a deepSplit parameter of 3 since this selection was optimal over and above other options for being able to accurately capture the major branch divisions that are apparent upon visual inspection of the dendrograms.

To visualize eigengene network topology we utilized the qgraph library in R (Epskamp et al., 2012) to construct weighted graphs of the ME adjacency matrices for each group. These graphs are depicted using a spring embedded layout algorithm (Fruchterman and Reingold, 1991) whereby highly connected nodes are attracted to each other and less highly connected nodes are repulsed away from each other. Because these plots are constructed from the adjacency matrices, distance is furthest apart when the correlation is r = -1 and closest when r = 1.

All hypothesis tests on connectivity strength between replicable differentially expressed modules, within and outside meta-module connectivity, and specific inter-modular (i.e. between-module) connectivity were implemented with permutation tests (10,000 iterations). The test statistic in each case was the difference in connectivity strength between ASD and Controls. On each iteration we randomized group labels and recomputed the test statistic. FDR (Storey, 2002) $q<0.05$ was used as the threshold for multiple comparisons correction.

### *Protein-Protein Interaction Analysis Between Dysregulated Co-Expression Modules*

To further underscore that statistical dependencies in highly correlated co-expression modules indicate true direct interactions between modules, we implemented a protein-protein interaction analysis to test whether the degree of physical interactions between proteins from dysregulated modules interact more than interactions between random sets of genes sampled from the background pool. To implement this analysis, we used Java-based command line tools for GeneMANIA (Warde-Farley et al., 2010) to query the latest protein-protein interaction database (Data Set ID: 2014-08-12; Database Version: 1 June 2014) and pull out all connections between proteins from a merged gene list of all downregulated and upregulated gene modules. We then utilized custom code to compute the number of interactions between a seed module and other modules either dysregulated in either the same direction as the seed module (i.e. downregulated seed module connections

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with other downregulated modules) or in the other direction as the seed module (i.e. downregulated seed module connections with upregulated modules). To test these results against the null hypothesis that any randomly selected list of genes would show similar degree protein-protein interactions, we ran a permutation analysis (1000 iterations) whereby on each iteration, we kept the seed module's genes intact, but randomly selected from the background pool (i.e. all 8,075 genes included in the co-expression analysis) the remaining genes to total the same number of genes as in the unpermuted analysis. On each iteration we then recomputed the number of interactions with the seed module. A p-value was then computed as the proportion of times in the permuted data that we observed interactions as great or greater than the interactions observed in the unpermuted data.

## Results

### *Replicable Dysregulation of Specific Gene Modules in ASD*

Consensus WGCNA on the 8,075 genes common to both the Voineagu and Gupta datasets identified 27 co-expression modules. Information regarding the enrichments for each of these modules can be found in Table S1. Module membership (i.e. the correlation between a gene and its module eigengene) and the top 10 hub genes based on module membership for each module are reported in Table S2. Ten of the 27 modules were identified as differentially expressed in a replicable fashion across datasets (i.e. replication Bayes Factor > 10; see Table S3 for full statistical information on these comparisons). Five of these 10 modules were on-average upregulated in ASD, while the remaining 5 were on-average downregulated in ASD. Three of the 5 ASD-upregulated modules (M12, M24, M27) were enriched in a variety of processes related to the immune system and inflammation; processes such as interferon signaling, complement system, phagocytosis, innate immune response to RNA viral infection, among several others (Fig 1). Interestingly, M12 and M27 are also enriched in M1 microglia markers, while M24 is enriched in M2 microglia markers (Fig 3; Table S4). The ASD-upregulated M25 module was heavily enriched for translation initiation and this enrichment is driven by a large number of genes coding for ribosomal proteins for the 40 and 60S ribosomal subunits (Fig 1). These genes also contributed to a significant enrichment in markers for the postsynaptic density (Fig 3; Table S4). The ASD-upregulated M1 module showed a mixed set of enrichment terms spanning cell signaling processes (i.e. NOTCH, Hedgehog signaling), axonal guidance, regulation of angiogenesis, integrin-mediated cell adhesion, cell cycle G1-S growth factor regulation, ESR2 signal transduction, among several others (Fig 1). Module M1 was enriched in astrocyte and M2 microglia markers (Fig 3; Table S4). In contrast to the ASD-upregulated modules, the replicable ASD-downregulated modules were enriched in a variety of synaptic, neuronal, cytoskeletal, and hormonal processes. These processes were diverse across modules and ranged from processes such as synaptic cell adhesion, synaptogenesis, neurogenesis, axonal guidance, synaptic vesicle exocytosis, transmission of nerve impulse, calcium transport, cell adhesion amyloid proteins, gonadotropin regulation, cytoskeleton spindle and cytoplasmic microtubules, actin filaments, and regulation of cytoskeleton rearrangement, amongst several others (Fig 2). In terms of cell type and cellular component enrichment, downregulated modules are enriched in neuronal (M3, M14), synaptic (M9), and postsynaptic density markers (M9) (Fig 3; Table S4).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

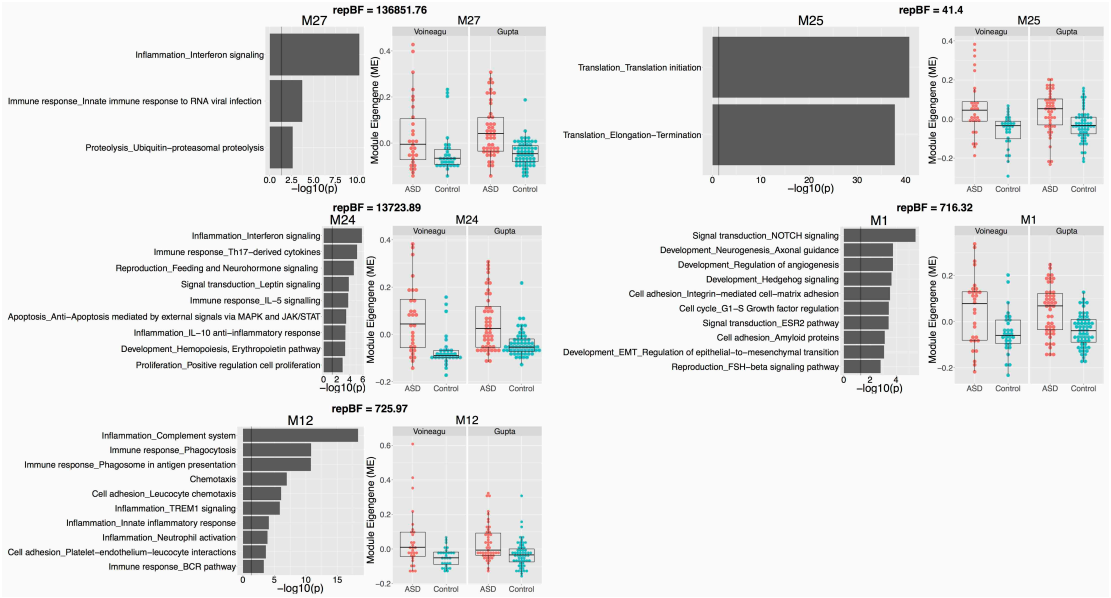


***Fig 1: Upregulated gene co-expression modules in ASD.*** *This figure shows gene co-expression modules that were on-average elevated in ME expression in ASD and in a replicable manner across datasets. Each module has a dot-boxplot whereby each individual is represented by a dot and the central tendency (median) and dispersion (interquartile range) is shown with the boxplot. Next to each dot-boxplot are the process-level enrichment terms passing FDR q<0.05 (limited to the top 10 terms) from MetaCore GeneGO. The vertical black line on the enrichment bar plots represents p = 0.05. For each module, the replication Bayes Factor statistic (repBF) is cited above the scatter-boxplot (repBF > 10 indicates strong evidence for replication).*

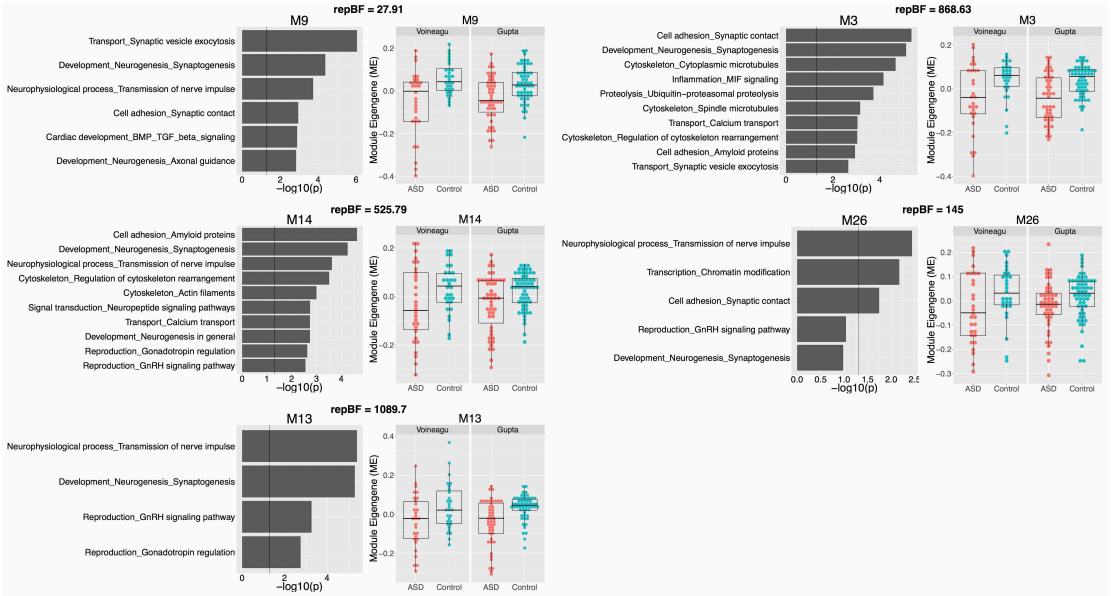


***Fig 2: Downregulated gene co-expression modules in ASD.*** *This figure shows gene co-expression modules that were on-average decreased in ME expression in ASD and in a replicable manner across datasets. Each module has a dot-boxplot whereby each individual is represented by a dot and the central tendency (median) and dispersion (interquartile range) is shown with the boxplot. Next to each scatter-boxplot are the process-level enrichment terms passing FDR q<0.05 (limited to the top 10 terms) from MetaCore GeneGO. The exception here is M26, whereby none of the terms passed FDR q<0.05. In this instance, we plot the first 5 terms for descriptive purposes. The vertical black line on the enrichment bar plots represents p = 0.05. For each module, the replication Bayes Factor statistic (repBF) is cited above the scatter-boxplot (repBF > 10 indicates strong evidence for replication).*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

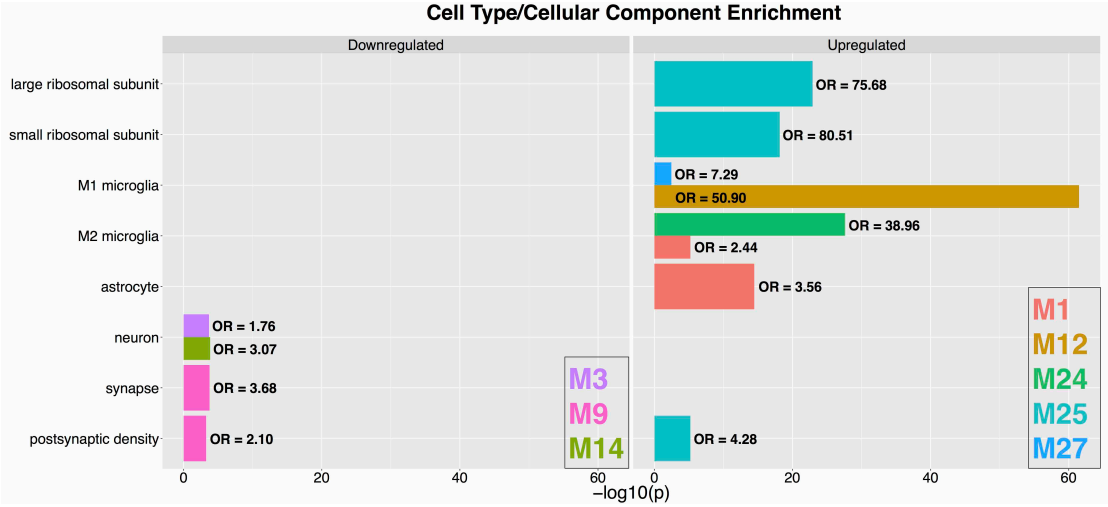


***Fig 3: Cell type/cellular compartment enrichments for dysregulated modules.*** *This figure shows enrichments in a variety of cell types and cellular components for the modules that are replicably dysregulated in ASD. The left panel shows enrichments for downregulated modules, while the right panel shows enrichments for the upregulated modules. The coloring of the bars denote which specific module shows the enrichment and the color legend is shown in the bottom right box for each panel. The x-axis plots the* $-\log_{10}$ *p-values while the y-axis indicates the specific cell type or cellular compartment. Next to each bar we indicate the enrichment odds ratio (OR).*

## *Differentially Expressed Modules are Highly Correlated in ASD*

Modules that are on-average differentially expressed (Figs 1-2) are highly correlated. This pattern of correlation was one of strong positive correlations within modules that share similar directionality of differential expression, but strong negative correlations between modules with different directionality of differential expression. Interestingly, these correlations become significantly enhanced in ASD compared to Controls in the Voineagu dataset (within downregulated modules $p = 0.012$; within upregulated modules $p = 0.042$; between downregulated and upregulated modules $p = 0.008$; Fig 4A-B). Within the Gupta dataset, this phenomenon of highly correlated differentially expressed modules as well as strong negative correlations between upregulated and downregulated modules is already present in Controls and stays present in ASD, though quantitative strengthening of such connectivity in ASD does not occur (within downregulated modules $p = 0.957$; within upregulated modules $p = 0.327$; between downregulated and upregulated modules $p = 0.667$; Fig 4C-D).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

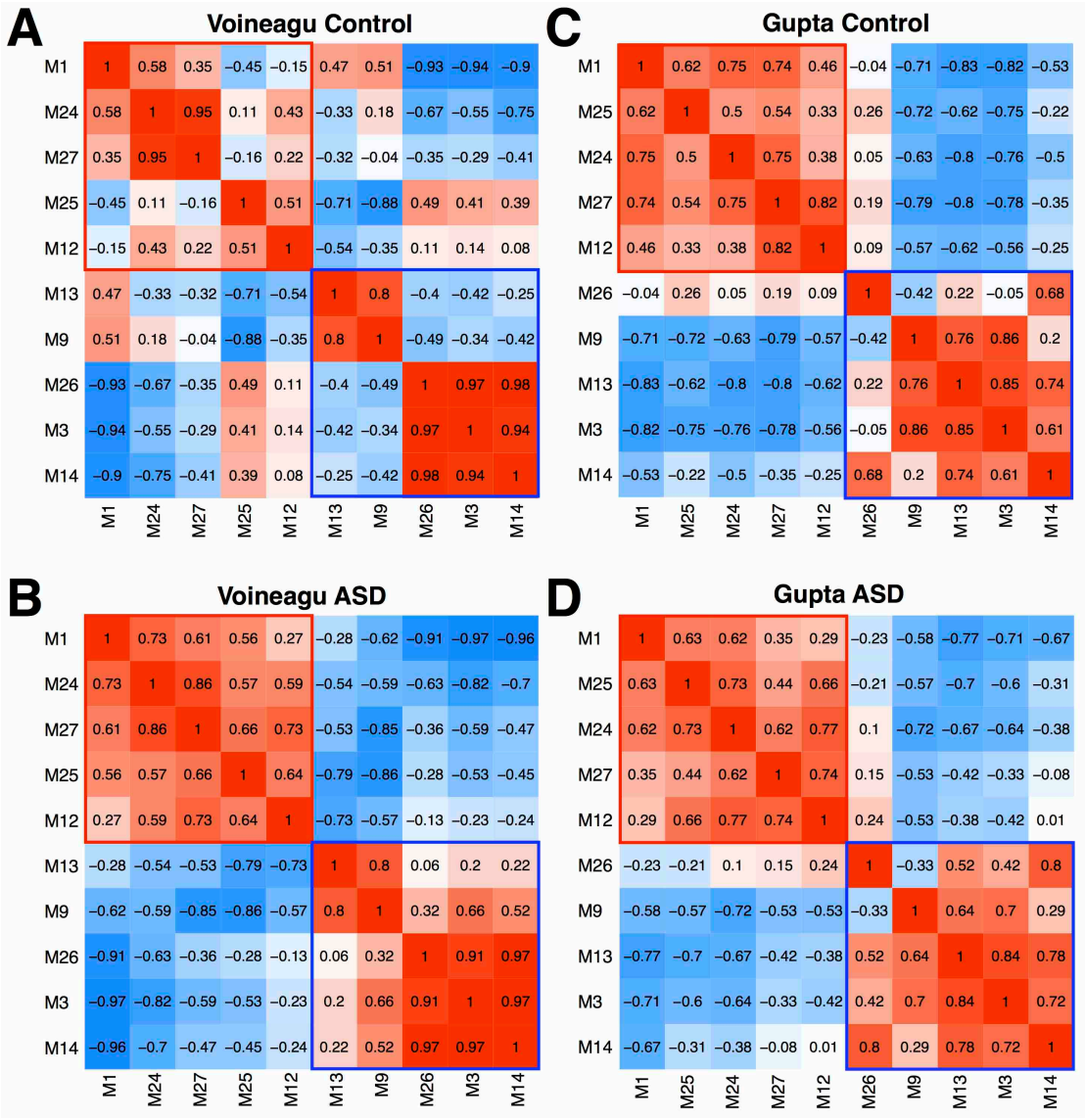


***Fig 4: Correlations between dysregulated modules.*** *Panels A and B show correlations between differentially expressed modules in the Voineagu Control (A) or ASD (B) datasets. Panels C and D show correlations between these same modules in the Gupta Control (A) or ASD (B) datasets.*

## *Highly Correlated Differentially Expressed Modules Highly Interact at the Level of Protein-Protein Interactions*

We next explicitly set out to test the hypothesis that correlations between dysregulated co-expression modules may be indicative of high levels of physical interactions between proteins. To answer this question, we queried the GeneMANIA protein-protein interaction database (Warde-Farley et al., 2010) and discovered that dysregulated modules do indeed show strong evidence for high levels of physical interaction at the protein level. Specifically, seed modules that are on-average dysregulated in ASD show a significantly higher number of protein interactions with other modules that are dysregulated both in similar or different directions (all $p < 9.99e-4$; Fig 5). For example, an ASD-downregulated module highly interacts with other downregulated modules that are show high positive correlations at the co-expression level. Similarly, the same downregulated module also highly interacts with other upregulated modules, whereby there is evidence in co-expression for negative,

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rather than positive correlations. This evidence bolsters the idea that these collections of highly correlated differentially expressed modules may represent connected systems biological phenomena that may not apparent by studying the co-expression modules in isolation.

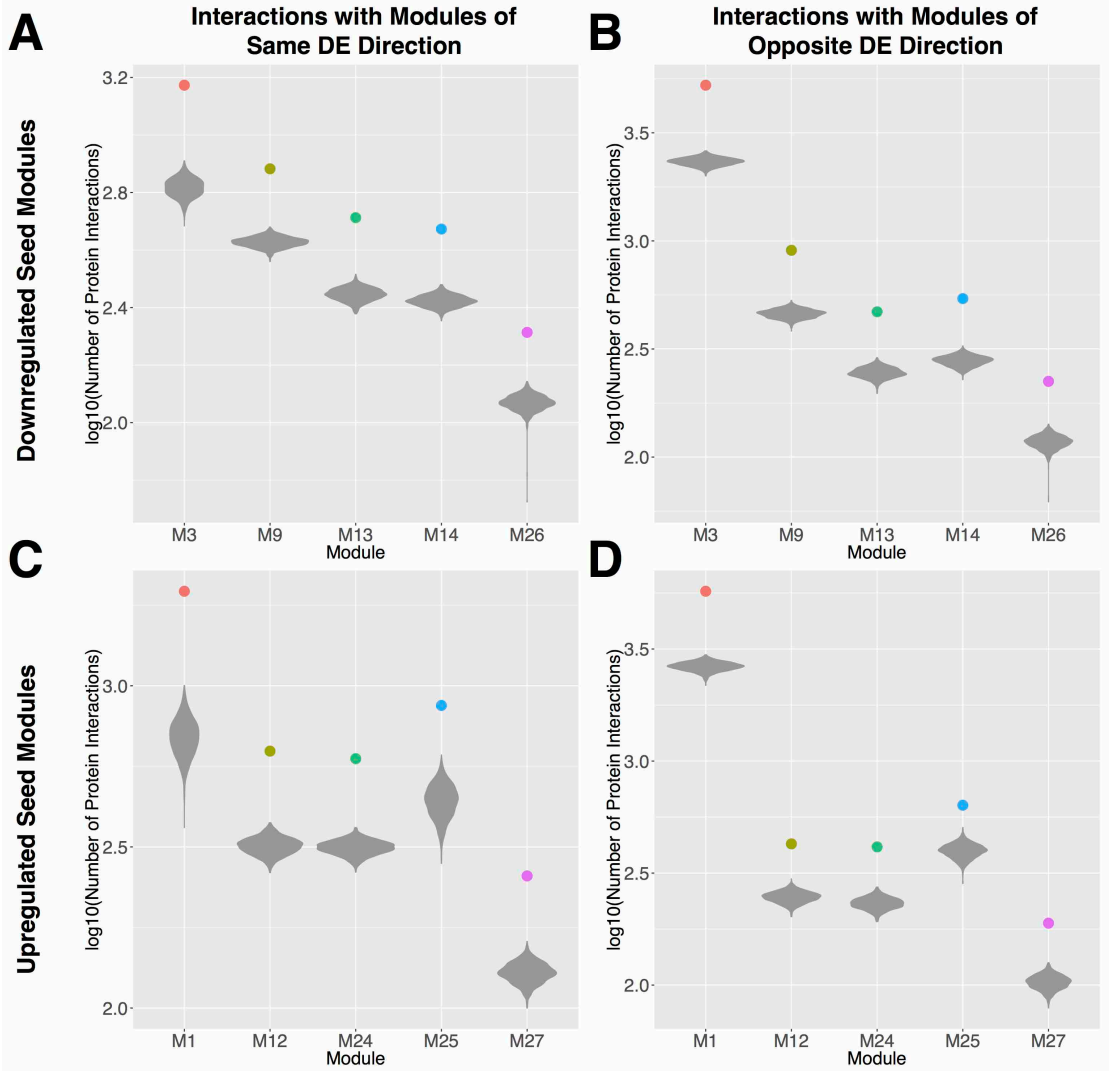


***Fig 5: Protein-protein interactions between dysregulated modules.*** *This figure shows the number of protein-protein interactions (on log10 scale) between a dysregulated seed module and groupings of other dysregulated modules that are either dysregulated in the same (panels A and C) or different direction (panels B and D) as the seed module. Panels A and B show results when using downregulated modules as seeds, while panels C and D show results when using upregulated modules as seeds. The colored dots represent the actual number of protein-protein interactions, while the grey colored violin plots indicate the null distribution of interactions between the seed module and randomly selected genes from the background pool.*

## *Processes Enriched within Dysregulated Modules*

We next asked the question of what biological processes might characterize such emergent phenomena. Leveraging the hierarchical structure of Gene Ontology (GO) we input merged lists of all differentially expressed modules together and computed GO biological process gene set enrichment and then clustered the top 50 enriched GO terms by semantic similarity (Supek et al., 2011). Here we find that combining highly connected downregulated modules results in enrichment in a

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

variety of synaptic and neural developmental processes (e.g., neuron projection, axonogenesis, neuron differentiation, regulation of cell growth) as well as very broad terms implicating higher organism-level disruption in cognition and behavior (Fig 6A). In contrast, combining highly connected upregulated modules results in enrichments in a variety of catabolic and viral processes, translation and protein targeting and localization, interferon signaling, glia processes, apoptosis and others (Fig 6B). These results suggest that highly connected differentially expressed modules spanning multiple cell types and cellular compartments, also highly interact at the protein level and result in emergent phenomena that are not visible simply by examining modules in isolation.

***Fig 6: GO biological process enrichments for collections of downregulated or upregulated modules.*** *This plot shows GO biological process enrichment terms for the combination of all downregulated (A) or upregulated (B) modules. The top 50 GO terms ranked by fold enrichment were input into REVIGO (Supek et al., 2011) in order to cluster GO terms by semantic similarity. These clusters are shown in different colors along with a descriptive label for each cluster. Plotted on the x-axis of each plot is the Bonferroni-corrected –log10 p-value for each term.*

## *Topological Reorganization of Eigengene Networks*

Moving beyond the 10 replicably dysregulated modules, we next examined all modules and how their hierarchical organization into eigengene networks may be topologically reorganized in ASD. To do this, we identified meta-modules (i.e. clusters of highly connected modules) and then visualized full eigengene network organization and meta-module membership with spring-embedded graphs that indicate topological change via distancing nodes based on strength of correlation between modules (i.e. shorter distance indicates stronger correlation, further distance indicates weaker correlation). In addition to this qualitative examination of eigengene network topology, we also quantitatively tested for differences with respect to

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

connectivity strength within and outside meta-module boundaries as well identifying specific modules with disrupted connectivity.

Within the Voineagu dataset, ASD-dysregulated modules are topologically arranged closer together and within the same meta-module, compared to the much more spread out and heterogeneous organization in Controls with respect to meta-module membership of dysregulated modules (Fig 7A-B). Quantitatively, network reorganization can be quantified by examining connectivity strength differences within and outside meta-module boundaries. Four modules (M25, M9, M21, and M23) show ASD-decreased connectivity within normative meta-module boundaries. These same modules along with one other module (M16) also show enhanced connectivity outside of normative meta-module boundaries in ASD (Fig 7C). At a nodal level, we further observed specific between-module connections that are prominently affected in ASD (Fig 7D). The ASD-upregulated M25 translation initiation module is normatively negatively correlated with the prominent ASD-upregulated M27 interferon signaling and M1 cell signaling (i.e. NOTCH, Hedgehog signaling), axonal guidance, cytoskeleton, and cell cycle enriched modules. However, in ASD, these negative correlations significantly reverse and turn into positive correlations, suggesting some abnormally heightened integration between these distinct biological processes/pathways. In another example, the ASD-downregulated M9 synaptic module is normatively positively correlated with M1, M15, and M16, but these relationships reverse into negative correlations in ASD. This suggests that what should typically be a natural integration between these modules ends up being an abnormal lack of integration in ASD. Furthermore, M9's connectivity with another ASD-downregulated synaptic module (M3) is normatively negative, yet in ASD is highly positively correlated. This is a particularly interesting effect given that both modules are on-average downregulated in ASD and share many synaptic enrichment terms (synaptogenesis, synaptic contact, synaptic vesicle exocytosis), yet are normatively lacking integration, but in ASD seem to be working together in tandem. Finally, while there is little to no normative relationship between the M9 synaptic module and the ASD-upregulated M27 interferon signaling module, in ASD this relationship turns into a strong negative correlation. This effect could potentially indicate an abnormal immune-synapse interaction between upregulation of inflammation interferon signaling processes and downregulation of important synaptic processes in ASD.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

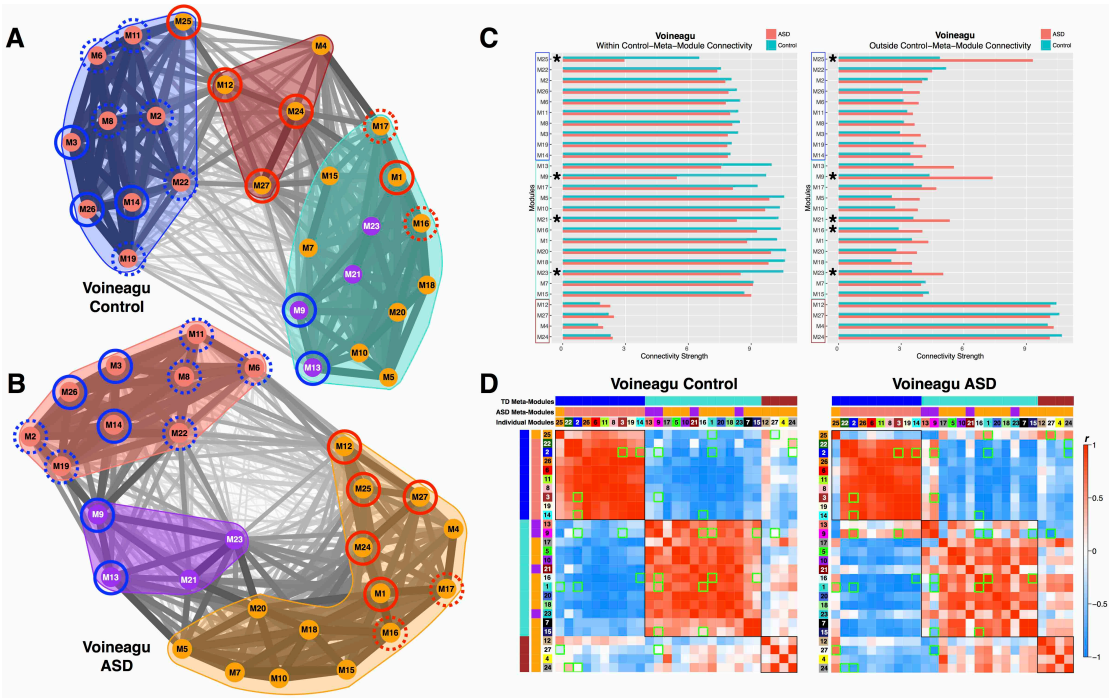


***Fig 7: Eigengene network topology and connectivity differences within the Voineagu dataset.*** *Panels A and B show eigengene networks as weighted graphs in a spring embedded layout for the Voineagu Control (A) or ASD (B) groups. The spring embedded layout places modules (nodes within the graphs) that are highly connected as much closer in space whereas modules that are less highly connected are repelled away from each other. The thickness of the connections (i.e. edges) between modules are scaled to connection strength whereby the thinnest line represents a correlation of r = -1 and the thickest line represents a correlation of r = 1. The color of each module node represents the ASD meta-module it belongs to. This was done to represent where the ASD meta-modules are located within the Control graph. The color-filled outlines around collections of modules represent the meta-module boundaries. Modules with a solid red or blue circle around it are modules that were identified in Figs 1-2 as being replicably dysregulated in ASD across both datasets (blue = ASD-downregulated; red = ASD-upregulated). The dotted circles represent differentially expressed modules (FDR q<0.05) present only within that specific dataset (see Table S3). Panel C shows within (C) and outside (D) normative meta-module connectivity strength for each seed module depicted on the y-axis. The normative (Control-defined) meta-modules are denoted by the color of the rectangular outlines on the y-axis. Connectivity strength is depicted on the x-axis and for within meta-module connectivity is defined as the sum of connection strength between the seed module and all other modules within the seed module's normative meta-module. Outside meta-module connectivity strength is defined as the sum of connection strength between the seed module and all other modules outside of the seed module's normative meta-module. Turquoise bars indicated Controls and salmon colored bars indicate ASD. The stars next to specific modules indicate a significant between-group difference in connectivity strength. Panel D illustrates eigengene networks as robust ME partial correlation matrices. Red coloring within the matrices indicates increasing positive correlation strength, while blue coloring indicates increasing negative correlation strength; see colorbar for key indicating how color corresponds to correlation strength. Matrices have rows and columns ordered by hierarchical clustering based on the Control group and the individual module numbers as well as meta-module colors are shown. Normative (Control-defined) meta-module boundaries are also clearly delineated by the black outlines over cells in the correlation matrices. Any cells with green outlines are those specific between-module connectivity comparisons that differed between-groups.*

Within the Gupta dataset there was also evidence of topological reorganization, which a much more fractionated organization of meta-modules in ASD compared to Controls (i.e. 6 meta-modules in ASD versus 4 in Controls). Similar to the Voineagu dataset, dysregulated modules again clustered close together and within the same meta-modules relative to a more heterogeneous organization in

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Controls (Fig 8A-B). Quantitatively, connectivity within and outside of normative meta-module boundaries was perturbed in ASD for nearly every single module (Fig 8C). This indicates that ASD eigengene network organization is highly perturbed with regard to connectivity of modules within normative eigengene network topology. In contrast to the numerous modules showing connectivity differences at the nodal level in the Voineagu dataset, very few nodal-level differences emerged. Most of these nodal differences were specific between-module connectivity for synaptic (M3, M5, M13), proteolysis, protein folding, and cell cycle (M11), and translation (M6) modules (Fig 8D). This subset of modules is normatively positively connected, but in ASD is abnormally negatively connected. This potentially indicates that normative interactions between these processes are potentially shut off or are disrupted in some fashion in ASD. For the Gupta dataset, it appears that overall eigengene network topology is reorganized in ASD in subtle ways that are spread across many modules and highly affect meta-modular organizational structure, but cannot be tied to very pronounced and specific differences within specific subsets of modules.

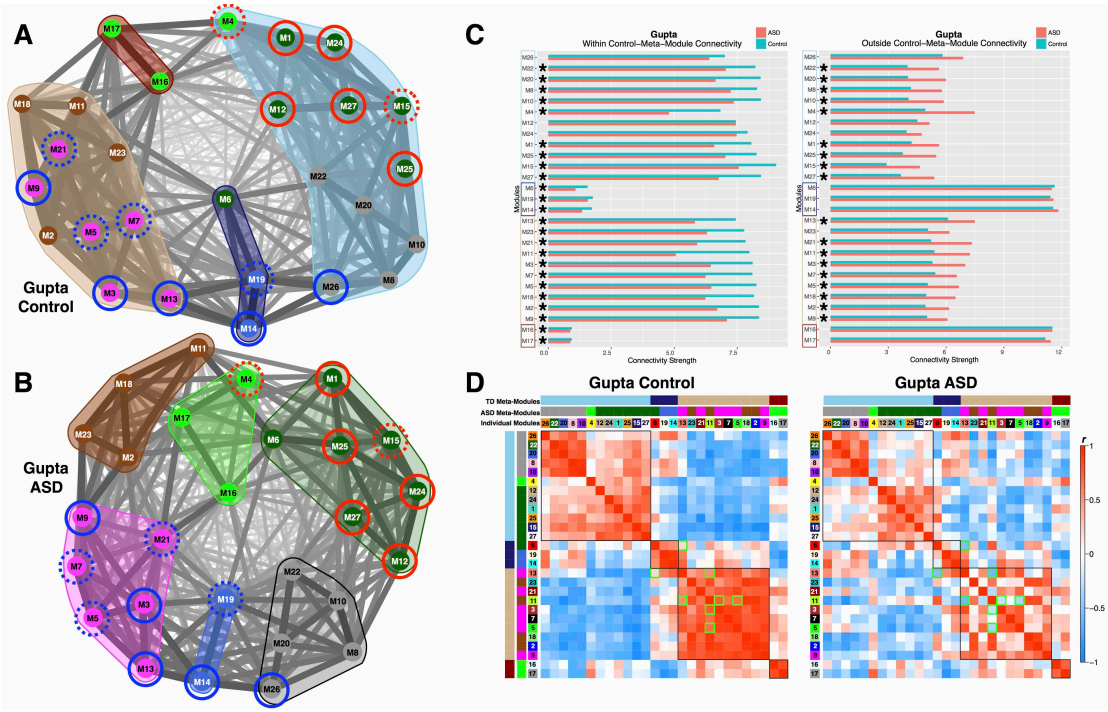


***Fig 8: Eigengene network topology and connectivity differences within the Gupta dataset.*** *Panels A and B show eigengene networks as weighted graphs in a spring embedded layout for the Gupta Control (A) or ASD (B) groups. The spring embedded layout places modules (nodes within the graphs) that are highly connected as much closer in space whereas modules that are less highly connected are repelled away from each other. The thickness of the connections (i.e. edges) between modules are scaled to connection strength whereby the thinnest line represents a correlation of r = -1 and the thickest line represents a correlation of r = 1. The color of each module node represents the ASD meta-module it belongs to. This was done to represent where the ASD meta-modules are located within the Control graph. The color-filled outlines around collections of modules represent the meta-module boundaries. Modules with a solid red or blue circle around it are modules that were identified in Figs 1-2 as being replicably dysregulated in ASD across both datasets (blue = ASD-downregulated; red = ASD-upregulated). The dotted circles represent differentially expressed modules (FDR q<0.05) present only within that specific dataset (see Table S3). Panel C shows within (C) and outside (D) normative meta-module connectivity strength for each seed module depicted on the y-axis. The normative (Control-defined) meta-modules are denoted by the color of the rectangular outlines on the y-axis. Connectivity strength is depicted on the x-axis and for within meta-module connectivity is defined as the sum of connection strength between the seed module and all other modules within the seed module's normative meta-module. Outside meta-module connectivity strength is defined as the sum of connection*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

*strength between the seed module and all other modules outside of the seed module's normative meta-module. Turquoise bars indicated Controls and salmon colored bars indicate ASD. The stars next to specific modules indicate a significant between-group difference in connectivity strength. Panel D illustrates eigengene networks as robust ME partial correlation matrices. Red coloring within the matrices indicates increasing positive correlation strength, while blue coloring indicates increasing negative correlation strength; see colorbar for key indicating how color corresponds to correlation strength. Matrices have rows and columns ordered by hierarchical clustering based on the Control group and the individual module numbers as well as meta-module colors are shown. Normative (Control-defined) meta-module boundaries are also clearly delineated by the black outlines over cells in the correlation matrices. Any cells with green outlines are those specific between-module connectivity comparisons that differed between-groups.*

## Discussion

Here we provide the first detailed characterization of how the ASD cortical transcriptome is hierarchically disorganized both at the level of specific co-expression modules and at higher levels of eigengene network organization (i.e. connectivity between modules and meta-modules). At the level of identifying dysregulation of individual co-expression modules, we have pinpointed several novel co-expression modules showing strong evidence for replicable dysregulation across datasets (Gupta et al., 2014; Voineagu et al., 2011). Methodologically, our joint consideration of multiple existing datasets in combination with discovery of strong Bayesian evidence for replicable dysregulation is a first within the literature, and would not have been possible by examining each dataset in isolation.

Unlike prior reports of two dysregulated modules, we have uncovered evidence for multiple downregulated synaptic modules and multiple upregulated immune/inflammation modules. The fact that these dysregulated signals do not converge into one module, but rather fractionate into several distinct modules suggests distinct co-regulated programs tuned to interact with various other biological processes and pathways active in the different cell types and cellular compartments. For example, M3 and M9 modules are on-average replicably downregulated in ASD and show similar biological process enrichment terms for synaptic processes. However, the difference between M3 and M9 becomes apparent when observing the different enrichments in cell type/compartment markers as well different interactions between modules. M3 is primarily enriched in neuronal markers, whereas M9 is specifically enriched in synaptic and postsynaptic density markers. Our examination of eigengene networks and connectivity between-modules allows for further novel distinctions between these modules. For example, within the Voineagu dataset, these modules are normatively negatively correlated, yet in ASD this relationship is significantly reversed to positive correlations. Within the context of normative brain development, these different modules may have unique roles to play but in ASD may be pulled together by some emergent pathophysiological and/or age-dependent processes taking place at the neuron, synapse, and postsynaptic density. Thus, these findings are an important distinction from prior work that implicated only a single synaptic module showing such ASD-downregulation. Furthermore, putting such results into the context of higher-order interactions at the eigengene network level is also critical, as it enables a better bird's eye view of how such multiple modules may interact differently, despite on-average showing the same directionality of a between-group difference and possessing similar biological process enrichment terms.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The findings of multiple types of ASD-upregulated immune/inflammation modules are also novel distinctions from past work. Although prior work has implicated interferon signaling, particularly with respect to M2 microglia markers (Gupta et al., 2014), here we find evidence for 2 upregulated interferon signaling modules (M24, M27). These modules differentiate by M1 and M2 microglia activation states, with M27 enriched in M1 microglia markers while M24 is enriched in M2 microglia markers. Between-module connectivity evidence also suggests that these two interferon signaling modules are disrupted in different ways. M27 is abnormally connected to an important ASD-upregulated translation initiation (M25) and ASD-downregulated synaptic module (M9). Given the enrichment in M27 for M1 microglia activation markers, this evidence suggests that cytotoxic M1 microglia processes may be affecting synaptic proteins in ASD. On the other hand, M24 shows intact connectivity between M25 and M9, but aberrant connectivity between other modules (M2, M22). These results suggest that while upregulated interferon signaling can be linked to both M1 and M2 microglia phenotypes, such aberrant processes may have differing impact on ASD brain function and structure.

In addition to the multiple dysregulated interferon signaling modules, we have also uncovered novel evidence for ASD-upregulation of an immune/inflammation module (M12) enriched in the complement system and phagocytosis processes and M1 microglia markers. In conjunction with effects from interferon signaling modules, the addition of the complement system may be of particular importance given the known links between the complement system and synaptic pruning (Stephan et al., 2012; Stevens et al., 2007) and remodeling as well as enhancing pro-inflammatory states of microglia activation in ASD (Morgan et al., 2010; Suzuki et al., 2013; Vargas et al., 2005). Recently, the complement system has been noted as a prominent player in the pathophysiology of schizophrenia, particularly for its role in synaptic pruning (Sekar et al., 2016). In the larger context of eigengene networks it is interesting that all of these important immune/inflammation modules are members of the same meta-module in ASD and that such a meta-module also includes other prominent modules such as the ASD-upregulated M25 translation initiation module. The current data present a role for complement system signaling alongside interferon signaling and other immune processes working together and potentially in concert with other important modules relating to translation and also for their role in various types of microglia activation states.

Translation has been an important topic in ASD primarily because of work on syndromic forms of autism related to mutations in *FMR1*, *TSC1/2*, and *PTEN* (Kelleher and Bear, 2008; Santini and Klann, 2014), as well as the important cap-dependent translation gene *EIF4E* (Gkogkas et al., 2013; Neves-Pereira et al., 2009; Santini et al., 2013; Yang et al., 2014). However, none of this work has specifically implicated ribosomal proteins themselves and no prior work on the cortical transcriptome in ASD has specifically implicated upregulation of translation initiation signals. Here we highlight one particularly prominent and novel ASD-upregulated module (M25) with heavy enrichment in translation initiation that is driven by a large number of ribosomal proteins for the 40S and 60S ribosomal subunits. A subset of these genes coding for ribosomal proteins also drive the enrichment in postsynaptic density markers. This module was also heavily dysregulated with respect to connectivity within and outside of normative meta-modular boundaries and showed specific abnormal interactions with other ASD-upregulated modules M1 and M27.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Additionally, this translation module was also a member of a meta-module in ASD that was composed of other upregulated immune/inflammation modules (M12, M24, M27), suggesting that it may play an important role integrating with upregulated immune/inflammation processes in ASD. Thus, not only have we discovered evidence for a novel and important upregulated signal in the ASD cortical transcriptome, but this finding also may have important implications with regards to its potential as a cross-cutting influence on other pathophysiological processes in ASD. From a systems point of view, work on blood leucocyte gene expression has also uncovered upregulated translation initiation as a prominent signal in young toddlers with ASD and this signal is present alongside other upregulated immune/inflammation signals, particularly interferon signaling and phagocytosis (Pramparo et al., 2015b). The presence of these dysregulated and highly connected translation initiation and immune/inflammation signals across brain and blood is potentially important because it may signal a unique opportunity to assay brain-relevant dysregulation in peripheral tissues and in-vivo in living patients. This peripheral window into potentially brain-relevant dysfunction that can be assayed in living patients may be particularly important given the recent discovery of a direct linkage between the brain and lymphatic vessels of the immune system (Louveau et al., 2015). Investigating this possible peripheral linkage to brain-relevant dysfunction in living patients using in-vivo techniques like functional and structural neuroimaging (Pramparo et al., 2015a) will be an important next step in understanding whether peripherally dysregulated signals in blood play some role in linking directly to important macro-level neural systems dysfunction in living patients (Lombardo et al., 2015). Another important direction for future work on this topic could be to better elucidate the role of these novel ribosomal protein genes via work with in-vivo or in-vitro models of key ribosomal proteins that are hub genes of this important M25 translation initiation module (see Table S2).

In addition to implicating several new gene co-expression modules of significance to ASD, this work provides primary evidence supporting the idea that the cortical transcriptome is dysregulated at hierarchical levels that cannot be understood from the vantage point of examining single co-expression modules in isolation. By identifying disruption in the interaction between-modules and in how eigengene networks are reconfigured into different meta-modular structures, this work presents a larger view on how multiple dysregulated signals may operate in conjunction with one another and potentially implicate important emergent interactions at the protein level. We show that a number of specific modules that are on-average up- or downregulated in ASD are also highly correlated and that this correlation can become stronger in ASD. This result is not apparent in prior work on this topic, with the closest result being the previous observation of a negative correlation when collapsing across both groups between single pair of modules enriched in synaptic and immune functions (Gupta et al., 2014). We have gone much further to show correlations between dysregulated modules including translation initiation modules and several other modules. We also demonstrated that beyond the statistical dependencies between co-expression modules, these dysregulated modules physically interact at the level of proteins. The disruption of these coordinated higher-order interactions at a protein level suggests that systems-level phenomena are disrupted in ASD that coordinates disparate biological processes and which cannot be adequately characterized by viewing smaller elements (e.g., single genes, single co-expression modules) in isolation. Thus, a primary conceptual advance from this aspect our work

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

suggests that we may need to move beyond arguments about single unitary processes, since the interactions between multiple dysregulated processes may underlie and better describe the pathology.

As a whole, the collection of ASD-downregulated modules appears to involve a widespread number of synaptic and broader neural developmental processes. These broader neural developmental processes in axonogenesis, positive regulation of cell growth, and regulation of neuron projection development are key new additions to previously implicated synaptic processes known to be dysregulated in ASD in prior transcriptome studies of ASD (Gupta et al., 2014; Voineagu et al., 2011). Relative to synaptic processes, these processes have developmentally prior roles that trace back to as early as the end of the first trimester of fetal brain development (Clancy et al., 2001) and could have key roles in ASD (Courchesne et al., 2011a; Courchesne et al., 2011b; Packer, 2016; Pramparo et al., 2015a). The nexus of both synaptic and these other neural developmental processes suggest an ongoing pathophysiology that extends throughout life in ASD and these processes may have important embedded roles in early and later brain development (Casanova and Casanova, 2014; Courchesne et al., 2011a; Courchesne et al., 2011b).

Alongside this downregulation of important synaptic and neural developmental processes, there are coordinated upregulated biological phenomena (i.e. immune/inflammation processes, translation, etc). To our knowledge, the novel signal of upregulated catabolism has not been implicated in any past work. Additionally, there are novel upregulated processes involved in protein targeting and localization that can be intertwined with translation processes (e.g., SRP-dependent cotranslational protein targeting to membrane). Finally, we also found enrichment in several viral processes, interferon signaling, glia-relevant, and apoptosis processes, and several other phenomena (e.g., catabolism, translation, protein targeting and localization). These highly coordinated processes are associated with multiple cell types/compartments, and the downregulation of synaptic and neuronal processes – as evidenced by the strong negative correlations between upregulated and downregulated modules. This evidence is generally in agreement with past theoretical ideas (Courchesne et al., 2011a) that suggested that early manifestations of pathophysiology potentially emerging in fetal development could then trigger a later corrective phase of development characterized by downregulation of synaptic and neuronal processes and potential upregulation immune/inflammation (e.g., microglia activation) (Morgan et al., 2010; Suzuki et al., 2013; Vargas et al., 2005), apoptotic, and other processes. A challenge for future research will be to unpack the relationships between known and novel upregulated processes with downregulated synaptic and neural developmental processes. However, it is important to underscore that these inferences emerge from the looking at the highly coordinated interactions between multiple dysregulated co-expression modules, and are not obvious by simply targeting specific modules and looking at such elements in isolation. Thus, these new insights produce new insights about systems level phenomena in ASD, and can guide reductionist studies to unravel specific mechanisms (e.g., targeting hub genes for many of the dysregulated modules we have implicated; Table S2).

In summary, this work highlights a hierarchical view of cortical transcriptome dysregulation in ASD. In doing so, we provide novel insight into new dysregulated processes coordinated with other previously described dysregulated signals. Our

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

approach allows for a better bird's eye view of how multiple pathophysiological processes may operate in ASD and may hint at new systems level phenomena as a potentially more accurate description of the pathophysiology affecting the brain in ASD. This perspective may have important translational and clinical implications as well as potential to help enable cross-level work connecting systems biology with systems neuroscience.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Acknowledgments

We would like to thank Bhismadev Chakrabarti, Bonnie Auyeung, Meng-Chuan Lai, Simon Baron-Cohen, and Peter Langfelder for helpful discussion on this work. This work is supported by grants (KL2TR00099 and 1KL2TR001444) from the University of California, San Diego Clinical and Translational Research Institute to Dr. Pramparo and a grant from the Simons Foundation Autism Research Initiative awarded to Prof. Courchesne (SFARI #176540).

## Author Contributions

MVL, EC, and TP conceived the idea for the study. MVL, NEL, and TP conceived all analyses. MVL implemented all data analyses. MVL, EC, NEL, and TP interpreted the results and wrote the manuscript. All authors read and approved the final manuscript.

## Conflict of Interest

The authors declare that they have no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# References

Casanova, E.L., and Casanova, M.F. (2014). Genetics studies indicate that neural induction and early neuronal maturation are disturbed in autism. Frontiers in cellular neuroscience *8*, 397.

Choi, G.B., Yim, Y.S., Wong, H., Kim, S., Kim, H., Kim, S.V., Hoeffer, C.A., Littman, D.R., and Huh, J.R. (2016). The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. Science (New York, NY *351*, 933-939.

Clancy, B., Darlington, R.B., and Finlay, B.L. (2001). Translating developmental time across mammalian species. Neuroscience *105*, 7-17.

Coiro, P., Padmashri, R., Suresh, A., Spartz, E., Pendyala, G., Chou, S., Jung, Y., Meays, B., Roy, S., Gautam, N.*, et al.* (2015). Impaired synaptic development in a maternal immune activation mouse model of neurodevelopmental disorders. Brain Behav Immun *50*, 249-258.

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

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

Elmer, B.M., Estes, M.L., Barrow, S.L., and McAllister, A.K. (2013). MHCI requires MEF2 transcription factors to negatively regulate synapse density during development and in disease. J Neurosci *33*, 13791-13804.

Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D., and Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. Journal of Statistical Software *48*, 1-18.

Fruchterman, T.M.J., and Reingold, E.M. (1991). Graph drawing by force‐directed placement. Software: Practice and experience *21*, 1129-1164.

Geschwind, D.H., and Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. Current opinion in neurobiology *17*, 103-111.

Geschwind, D.H., and State, M.W. (2015). Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gkogkas, C.G., Khoutorsky, A., Ran, I., Rampakakis, E., Nevarko, T., Weatherill, D.B., Vasuta, C., Yee, S., Truitt, M., Dallaire, P.*, et al.* (2013). Autism-related deficits via dysregulated eIF4E-dependent translational control. Nature *493*, 371-377.

Glynn, M.W., Elmer, B.M., Garay, P.A., Liu, X.B., Needleman, L.A., El-Sabeawy, F., and McAllister, A.K. (2011). MHCI negatively regulates synapse density during the establishment of cortical connections. Nature neuroscience *14*, 442-451.

Gokoolparsadh, A., Sutton, G.J., Charamko, A., Green, N.F., Pardy, C.J., and Voineagu, I. (2016). Searching for convergent pathways in autism spectrum disorders: insights from human brain transcriptome studies. Cell Mol Life Sci.

Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B., and Arking, D.E. (2014). Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun *5*, 5748.

Kelleher, R.J., 3rd, and Bear, M.F. (2008). The autistic neuron: troubled translation? Cell *135*, 401-406.

Langfelder, P., and Horvath, S. (2007). Eigengene networks for studying the relationships between co-expression modules. BMC Syst Biol *1*, 54.

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

Langfelder, P., Zhang, B., and Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. Bioinformatics *24*, 719-720.

Le Belle, J.E., Sperry, J., Ngo, A., Ghochani, Y., Laks, D.R., Lopez-Aranda, M., Silva, A.J., and Kornblum, H.I. (2014). Maternal inflammation contributes to brain overgrowth and autism-associated behaviors through altered redox signaling in stem and progenitor cells. Stem cell reports *3*, 725-734.

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

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

Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., and Horvath, S. (2011). Strategies for aggregating gene expression data: the collapseRows R function. BMC Bioinformatics *12*, 322.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Morgan, J.T., Chana, G., Pardo, C.A., Achim, C., Semendeferi, K., Buckwalter, J., Courchesne, E., and Everall, I.P. (2010). Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. Biol Psychiatry *68*, 368-376.

Neves-Pereira, M., Muller, B., Massie, D., Williams, J.H., O'Brien, P.C., Hughes, A., Shen, S.B., Clair, D.S., and Miedzybrodzka, Z. (2009). Deregulation of EIF4E: a novel mechanism for autism. J Med Genet *46*, 759-765.

Oldham, M.C., Horvath, S., and Geschwind, D.H. (2006). Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proceedings of the National Academy of Sciences of the United States of America *103*, 17973-17978.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. Nature neuroscience *11*, 1271-1282.

Oskvig, D.B., Elkahloun, A.G., Johnson, K.R., Phillips, T.M., and Herkenham, M. (2012). Maternal immune activation by LPS selectively alters specific gene expression profiles of interneuron migration and oxidative stress in the fetus without triggering a fetal immune response. Brain Behav Immun *26*, 623-634.

Packer, A. (2016). Neocortical Neurogenesis and the Etiology of Autism Spectrum Disorder. Neurosci Biobehav Rev.

Pramparo, T., Lombardo, M.V., Campbell, K., Carter Barnes, C., Marinero, S., Solso, S., Young, J., Mayo, M., Dale, A., Ahrens-Barbeau, C.*, et al.* (2015a). Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. Mol Syst Biol *11*, 841.

Pramparo, T., Pierce, K., Lombardo, M.V., Carter Barnes, C., Marinero, S., Ahrens-Barbeau, C., Murray, S.S., Lopez, L., Xu, R., and Courchesne, E. (2015b). Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. JAMA Psychiatry *72*, 386-394.

Santini, E., Huynh, T.N., MacAskill, A.F., Carter, A.G., Pierre, P., Ruggero, D., Kaphzan, H., and Klann, E. (2013). Exaggerated translation causes synaptic and behavioural aberrations associated with autism. Nature *493*, 411-415.

Santini, E., and Klann, E. (2014). Reciprocal signaling between translational control pathways and synaptic proteins in autism spectrum disorders. Sci Signal *7*, re10.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V.*, et al.* (2016). Schizophrenia risk from complex variation of complement component 4. Nature.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stephan, A.H., Barres, B.A., and Stevens, B. (2012). The complement system: an unexpected role in synaptic pruning during development and disease. Annual review of neuroscience *35*, 369-389.

Stevens, B., Allen, N.J., Vazquez, L.E., Howell, G.R., Christopherson, K.S., Nouri, N., Micheva, K.D., Mehalow, A.K., Huberman, A.D., Stafford, B.*, et al.* (2007). The classical complement cascade mediates CNS synapse elimination. Cell *131*, 1164-1178.

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

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

Suzuki, K., Sugihara, G., Ouchi, Y., Nakamura, K., Futatsubashi, M., Takebayashi, K., Yoshihara, Y., Omata, K., Matsumoto, K., Tsuchiya, K.J.*, et al.* (2013). Microglial activation in young adults with autism spectrum disorder. JAMA Psychiatry *70*, 49-58.

Vargas, D.L., Nascimbene, C., Krishnan, C., Zimmerman, A.W., and Pardo, C.A. (2005). Neuroglial activation and neuroinflammation in the brain of patients with autism. Annals of neurology *57*, 67-81.

Verhagen, J., and Wagenmakers, E.J. (2014). Bayesian tests to quantify the result of a replication attempt. J Exp Psychol Gen *143*, 1457-1475.

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

Wager, T.D., Keller, M.C., Lacey, S.C., and Jonides, J. (2005). Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage *26*, 99-113.

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T.*, et al.* (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research *38*, W214-220.

Yang, G., Smibert, C.A., Kaplan, D.R., and Miller, F.D. (2014). An eIF4E1/4E-T complex determines the genesis of neurons from precursors by translationally repressing a proneurogenic transcription program. Neuron *84*, 723-739.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Supplementary Figures

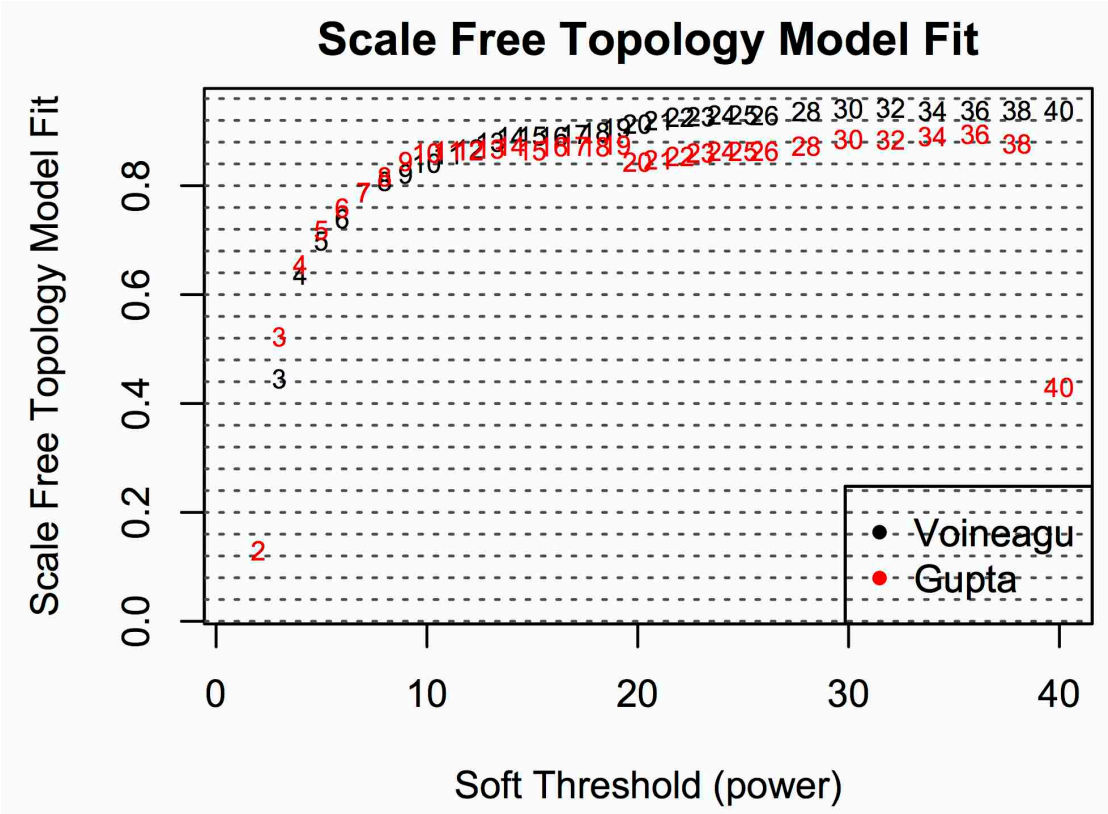


***Fig S1: Scale-free topology model fit across a range of soft power thresholds****. This plot shows the scale-free topology model fit scores ($R^2$) across a range of soft power thresholds. This analysis is done in order to choose a soft-power threshold to use in the main analyses. As a rule, we picked the soft power threshold whereby scale-free topology model fit $R^2$ is maximum and begins to plateau (i.e. soft power = 14).*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

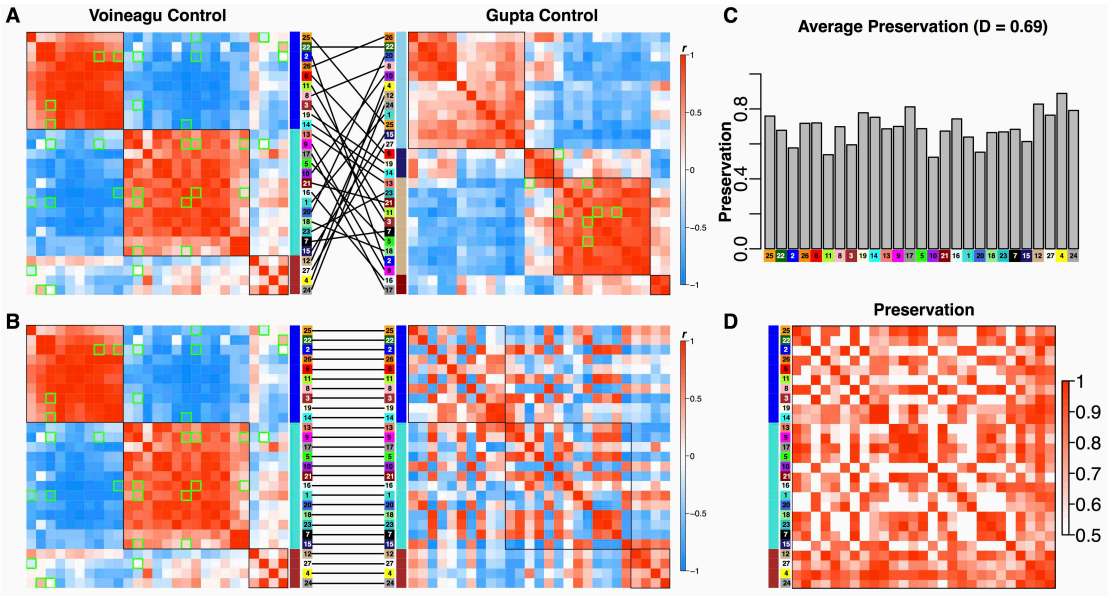


***Fig S2: Preservation of eigengene networks in the TD group***. *Panel A shows the eigengene networks for Voineagu and Gupta datasets when the rows and columns of the matrix are ordered by meta-module clustering. Panel B shows the matrices when ordered only by the Voineagu TD dataset clustering. Panel C shows average preservation levels across each module. Panel D shows preservation for all pairwise module comparisons. The plots in panels C and D were made using a modified version of the plotEigengeneNetworks function in the WGCNA R library. We modified this function to use ME robust partial correlation matrices.*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted July 20, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Supplemental Tables

***Table S1:*** *Enrichments for all modules, meta-modules, and collections of downregulated and upregulated modules.*

***Table S2:*** *Module membership and hub gene information for each module*

***Table S3:*** *Full result table of analysis examining on-average differential expression in ME values*

***Table S4:*** *Cell type and cellular component enrichment information*