bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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], & Tiziano Pramparo[4]

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

Corresponding Author: Michael V. Lombardo (mvlombardo@gmail.com), University of Cyprus, Department of Psychology, 1 Panepistimiou Avenue, Aglantzia, Nicosia 1678, Cyprus

Running title: Hierarchical transcriptome disruption in ASD

Keywords: autism; transcriptome; systems biology; synapse; immune; translation

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity.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

Elucidating how the cortical transcriptome is hierarchically disorganized may be key for understanding the pathophysiology behind of autism (ASD). Here we find replicable evidence across 2 datasets for 10 gene co-expression modules that are differentially expressed in ASD. These modules span multiple cell types and compartments including M1 and M2 microglia, astrocytes, ribosomal subunits, neuron, synapse, and postsynaptic density. These dysregulated modules are also strongly correlated and form higher-level emergent properties characterized by coordination between 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 at the level of clusters of highly connected modules (i.e. meta-modules) is also disrupted in ASD, as is connectivity between specific synaptic, immune, and translation modules. These results support a view of hierarchical cortical transcriptome disorganization in ASD characterized by emergent pathophysiology not readily apparent in smaller isolated elements.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 diversity of mechanisms implicated in ASD is large and this heterogeneity presents a complex puzzle with regards to how diverse molecular mechanisms may converge or diverge at different levels of analysis[1, 2]. At the level of the transcriptome, several studies have examined how gene expression is dysregulated in autism spectrum disorders (ASD) within blood and brain tissue samples[3, 4, 5, 6, 7, 8, 9, 10]. This work highlights some common themes regarding atypical synaptic and immune/inflammation processes. While this progress is very important, we also need to move further towards an understanding of how multiple dysregulated processes fit together within a larger systems biological context. The human brain is a hierarchically organized complex system and understanding the pathophysiology behind complex neurodevelopmental disorders like ASD requires an approach that elucidates how brain development in ASD may be hierarchically disrupted[11]. This more holistic systems biology understanding will allow us to move from important reductionist insights about single unitary mechanisms to understanding how such mechanisms may work at multiple hierarchical scales in concert to produce emergent phenomena that may help to better explain the pathophysiology behind ASD but also hone in on new ways to potentially treat affected individuals.

Elucidating principles of hierarchical organization of the cortical transcriptome in ASD has yet to be examined, but may significantly help in furthering our understanding about the pathophysiology of ASD. Past work has highlighted two important dysregulated gene modules - one downregulated in ASD and enriched for synaptic processes and neuronal markers and a second module upregulated in ASD and enriched for immune processes and astrocyte and M2 microglia activation state markers[5, 6]. Voineagu and colleagues found that the downregulated synaptic module was enriched in GWAS signals and thus interpreted this module to be of primary importance, given its likely genetic etiology. In contrast, the upregulated immune module was not enriched in GWAS signals and was interpreted as a secondary phenomenon relative to the synaptic dysregulation and/or could potentially be mediated by non-genetic etiological factors[5]. While this work was seminal in furthering our understanding of how the transcriptome is dysregulated in ASD, the interpretation could lead to some ambiguity with regard to how synaptic and immune dysregulation fit together. In one scenario the synaptic and immune processes could be considered to be independent of each other (i.e. two dysregulated yet non-interacting uncorrelated processes), while in another scenario one might consider that there is some more complex interaction between the two processes, but one that places more emphasis and importance on synaptic than the immune factors. Providing some evidence for the latter scenario, Gupta and colleagues later found that these modules are negatively correlated amongst all data samples combined across both groups, suggesting that there is important interplay between synaptic and immune processes[6]. However, this observation does not clarify the relationship between such modules in each group or critically whether the groups differ in such a relationship. Gupta and colleagues also placed particular emphasis on enrichment in markers for an M2 microglia activation state[6]. M1 microglia activation typically confers pro-inflammatory cytotoxic effects, whereas M2 microglia activation can be neuroprotective (e.g., trophic effects, progenitor cell proliferation)[12, 13, 14, 15]. Given the downregulation of synaptic processes and upregulation of connected immune/inflammation processes, these observations might be more congruent with a role for M1 microglia affecting the synapse. However, given other theoretical ideas

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

regarding pathophysiology potentially emerging in fetal development relating to early brain overgrowth, followed by a later corrective phase of development persisting throughout the lifespan[16], it is also possible that enhanced M2 microglia activation states are important[6] or that a mixture of M1 and M2 microglia states are present but differ throughout development. Thus, there may be a role for both microglia phenotypes in ASD and that such phenotypes could co-exist within the microglia population[17] in the ASD brain.

In contrast to the view that synaptic mechanisms are more primary, it is noteworthy that in animal models wherein immune phenomena linked to ASD are positioned as the main causal mechanism (i.e. maternal immune activation)[18, 19], the result of such non-genetic immune etiology is up-regulation of cell cycle and DNA damage repair and downregulation of multiple neural developmental processes[20] as well as alterations occurring at the synapse[21, 22]. Early neural developmental phenomena occurring near the end of the first trimester of fetal development are also highly perturbed via this immune-mediated non-genetic mechanism[23] and could ultimately trigger cortical developmental defects very similar to those reported in ASD cortex[24, 25]. Therefore, it is clear that the causal directionality of dysregulated mechanisms could go both ways and most importantly, that we must better understand how such dysregulated processes form hierarchical emergent interactions and whether such interaction occurs differently in ASD.

Here we present the first work examining hierarchical disruption of the cortical transcriptome in ASD. Rather than representing independent forms of transcriptome dysregulation, we hypothesize that dysregulated modules work together to form emergent pathophysiological processes that are not visible by looking at single modules in isolation. Therefore, we predict that differentially expressed modules will be highly correlated and that such connectivity may be increased in ASD. Furthermore, given the emerging literature on neuronal/synaptic-immune interactions[20, 21, 22, 23, 24, 26, 27, 28], we also hypothesized that specific dysregulated modules enriched in immune/inflammation and synaptic processes will be aberrantly connected in ASD. We also provide the first look at the full organization of connections between gene modules (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)[29, 30, 31]. It could be that subtle and specific change or global patterns of network reorganization occur within 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.

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

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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).

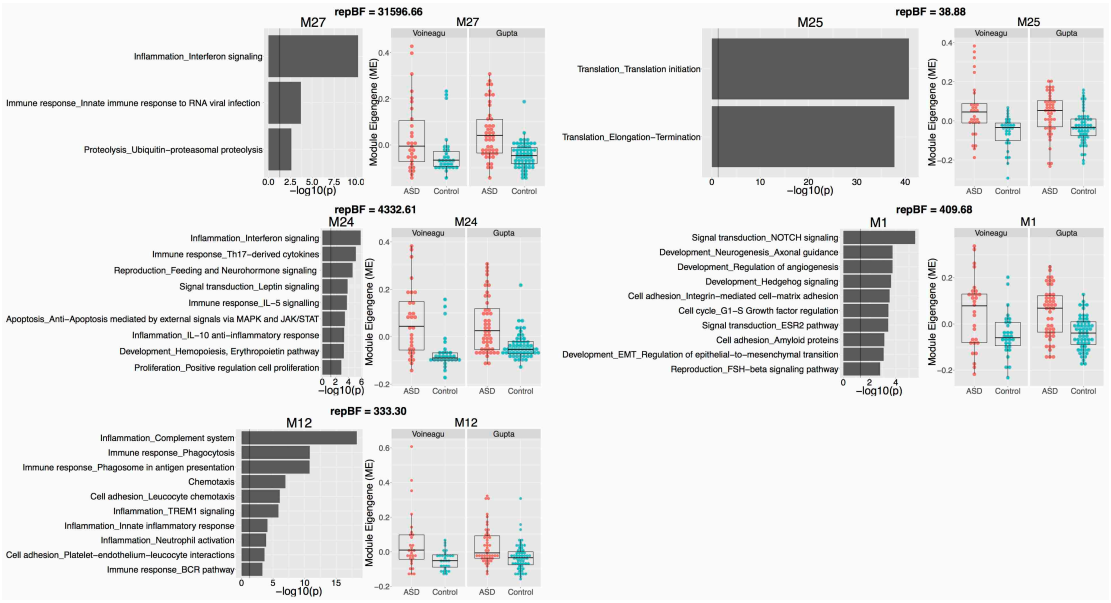

Fig 1: Upregulated gene co-expression modules in ASD. *This figure shows gene co-expression modules that were on-average elevated in ME expression in ASD and in a replicable manner across datasets. Each module has a dot-boxplot whereby each individual is represented by a dot and the central tendency (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)*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 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). The plot in the bottom right corner depicts cell type and cellular compartment enrichments. Enrichment odds ratios (OR) are reported next to each bar.

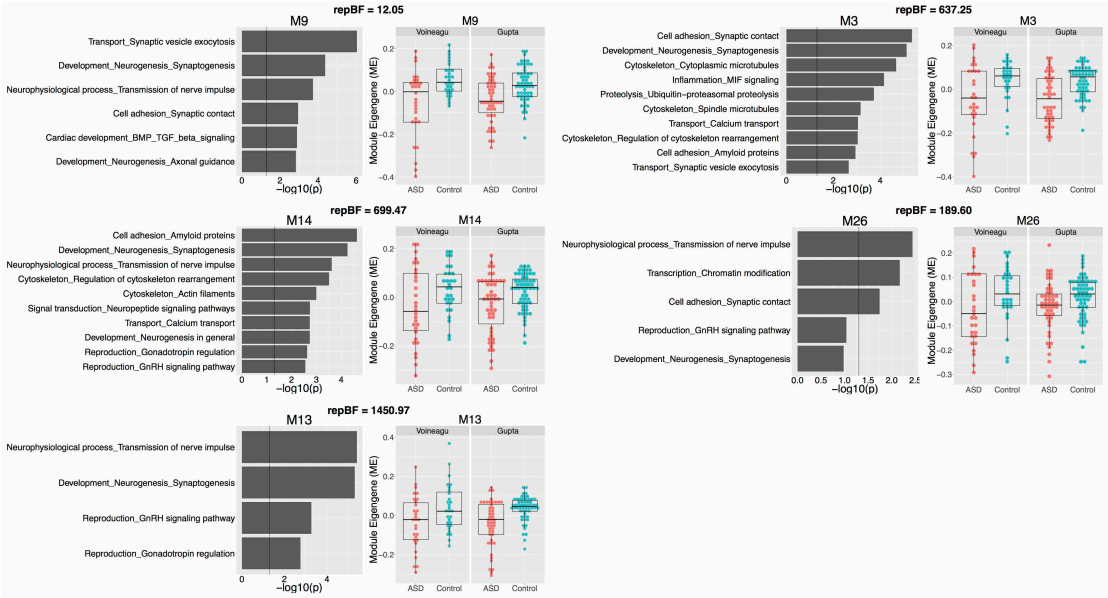

Fig 2: Downregulated gene co-expression modules in ASD. *This figure shows gene co-expression modules that were on-average decreased in ME expression in ASD and in a replicable manner across datasets. Each module has a dot-boxplot whereby each individual is represented by a dot and the central tendency (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).*

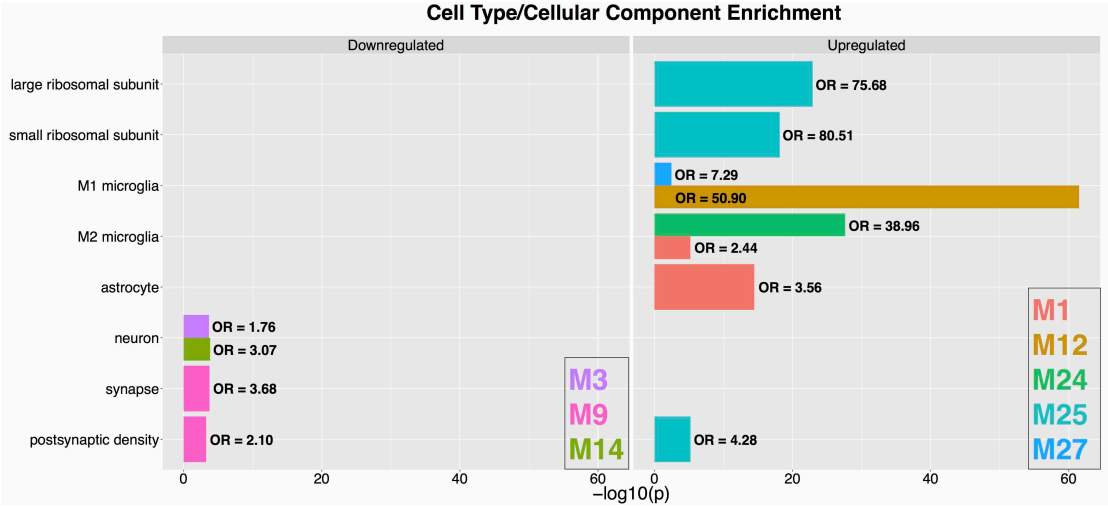

Fig 3: Cell type/cellular compartment enrichments for dysregulated modules. *This figure shows enrichments in a variety of cell types and cellular components for the modules that are replicably dysregulated in ASD. The left panel shows enrichments for downregulated modules, while the right panel shows 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 –log10 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).*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

Differentially Expressed Modules are Highly Connected 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 3C-D). Upregulated modules M25, M12, M24, M27 and M1 are topologically spread out across three distinct meta-modules in Controls (blue, brown and turquoise; Fig 4A). In contrast, these five modules tightly clustered together within the same meta-module in ASD and are at opposite poles of the graph (i.e. high negative correlations) compared to downregulated modules M3, M14, M26, M9, and M13 (Fig 4B). Similarly, downregulated synaptic modules M3, M14, and M26 cluster close together (i.e. high positive correlations) within the Control salmon meta-module, but are far away from (i.e. high negative correlations) other important synaptic modules (M9, M13) located within the Control turquoise meta-module at the other pole of the weighted spring embedded graph (Fig 4A). In ASD this relationship changes as M9 and M13 become positively correlated with M3, M14, and M26 and are brought within closer proximity of each other (Fig 4B).

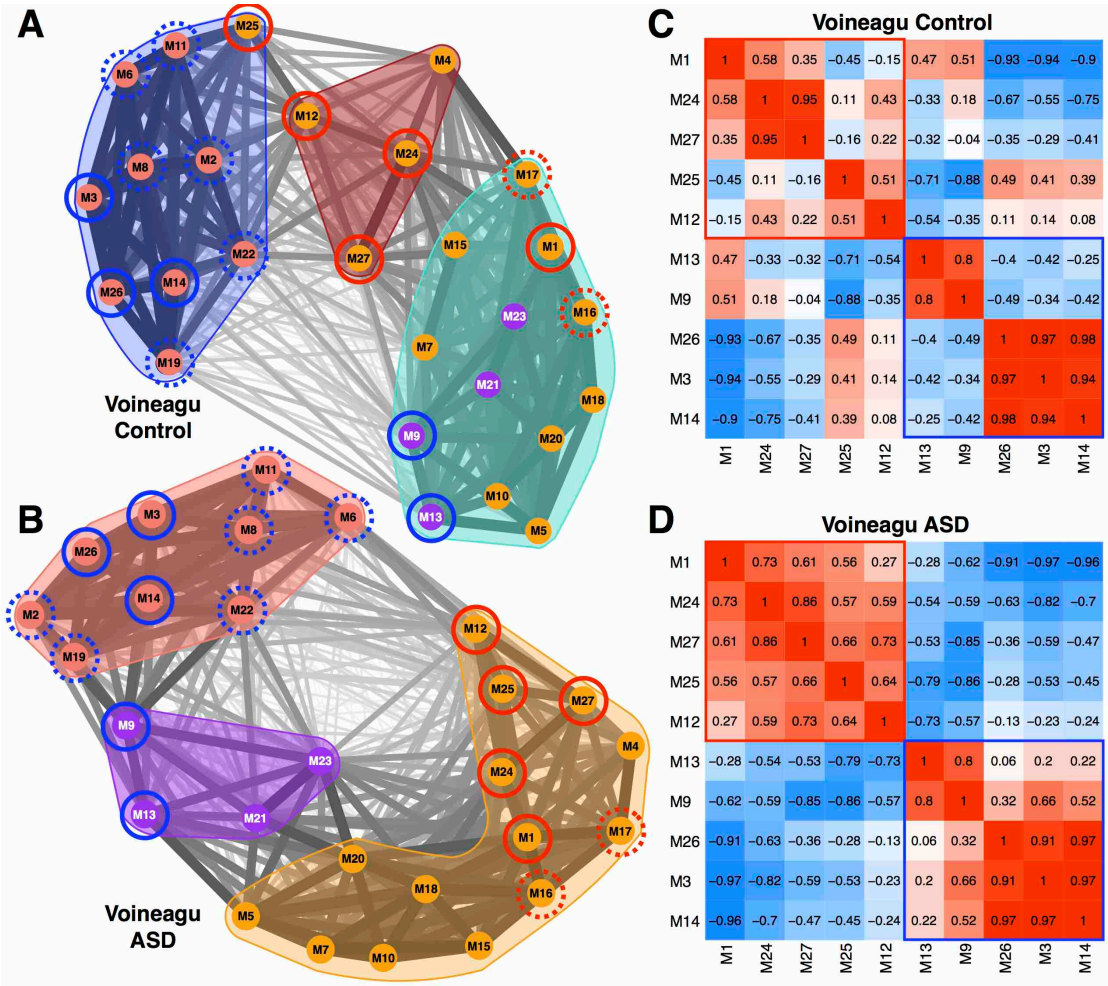

Fig 4: Eigengene networks, meta-modules, and correlations between dysregulated modules within the Voineagu dataset. *This figure shows eigengene networks as weighted graphs in a spring embedded*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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). Panels C and D show the differentially expressed modules only as correlation matrices. The strength of the correlation is shown within each cell and via the coloring (hot colors indicating higher correlations, cooler colors indicating lower correlations). The red outline denotes ASD-upregulated modules while the blue outline denotes ASD-downregulated modules.

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 5C-D). Despite these correlations being already present in Controls and not quantitatively strengthened in ASD, there is further evidence for a qualitative difference in ASD with regards to how these modules cluster within meta-modules. Within Control meta-modules there are mixtures of modules that are either differentially expressed or not and in different directions (i.e. the Control skyblue meta-module contains both downregulated and upregulated modules as well as modules with no differential expression; Fig 5A). However, in ASD, the composition of meta-modules is much more homogeneous with respect to differentially expressed modules of similar directionality and less frequent presence of non-differentially expressed modules (Fig 5B).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

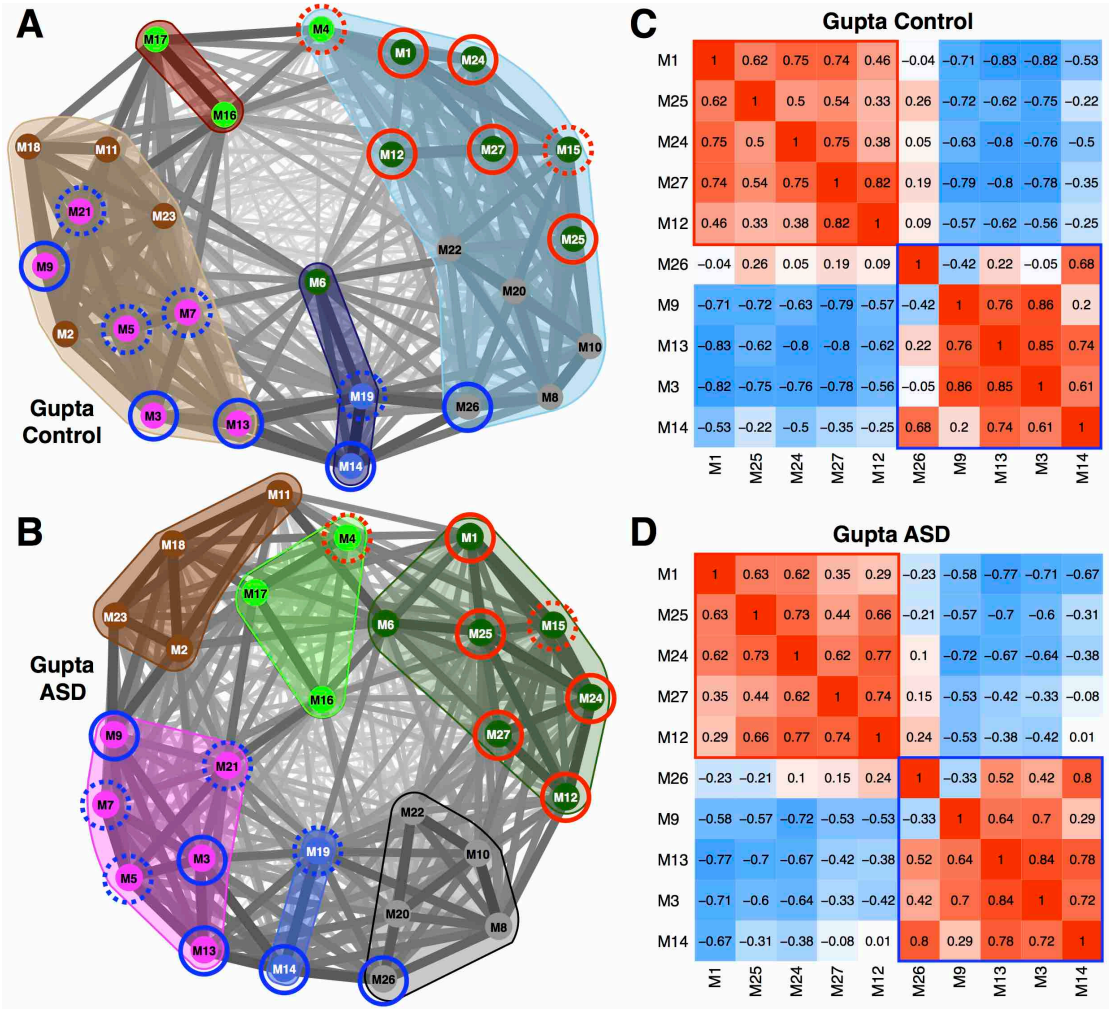

Fig 5: Eigengene networks, meta-modules, and correlations between dysregulated modules within the Gupta dataset. *This figure shows eigengene networks as weighted graphs in a spring embedded layout for the Gupta Control (A) or ASD (B) groups. The spring embedded layout places modules (nodes within the 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). Panels C and D show the differentially expressed modules only as correlation matrices. The strength of the correlation is shown within each cell and via the coloring (hot colors indicating higher correlations, cooler colors indicating lower correlations). The red outline denotes ASD-upregulated modules while the blue outline denotes ASD-downregulated modules.*

We next asked the question of whether these highly connected differentially expressed modules interact at the higher-level of protein-protein interactions. Using the STRING database (STRINGdb v10)[32] we discovered strong evidence that these collections of dysregulated modules possess a high degree of physical and functional associations in protein-protein interaction networks (downregulated modules: observed interactions = 10206, expected interactions = 8040, p = 0; upregulated modules: observed interactions = 15491, expected interactions = 11900, p = 0). This result further bolsters the idea that these collections of highly connected differentially

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

expressed modules form emergent properties. We then asked what biological processes might characterize such emergent properties. 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. Here we find that combining highly connected downregulated modules results in enrichment in a 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[33] 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.*

Disrupted Connectivity Between Modules and Meta-Modules

Fig 4A-B and Fig 5A-B show a number of interesting shifts in topology of eigengene networks, particularly with respect to changes in the composition of modules within meta-modules in ASD and Controls. Thus, we next examined quantitatively whether these topological changes with respect to meta-modular organization are disrupted in ASD. In particular, we assessed disruptions of

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 between a particular module and other modules within or outside of its normative meta-module boundaries. Within the Voineagu dataset we find a specific subset of 4 modules (M25, M9, M21, and M23) that 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-D). M25 is on-average upregulated in expression in ASD and heavily enriched in translation initiation genes (i.e. ribosomal proteins) (Fig 1), while M9, M21, and M23 are all members of a distinct meta-module in ASD (purple) highly enriched in synaptic processes (Table S1) and M9 is also on-average downregulated in expression in ASD (Fig 2). These disruptions in meta-module connectivity are easily visualized in the correlation matrix in Fig 7A-B. For instance, M25's connections within and outside of the normative blue meta-module are reversed in ASD compared to Controls. A similar though less obvious reversal of connectivity strength in ASD can also be seen for M9 and M13.

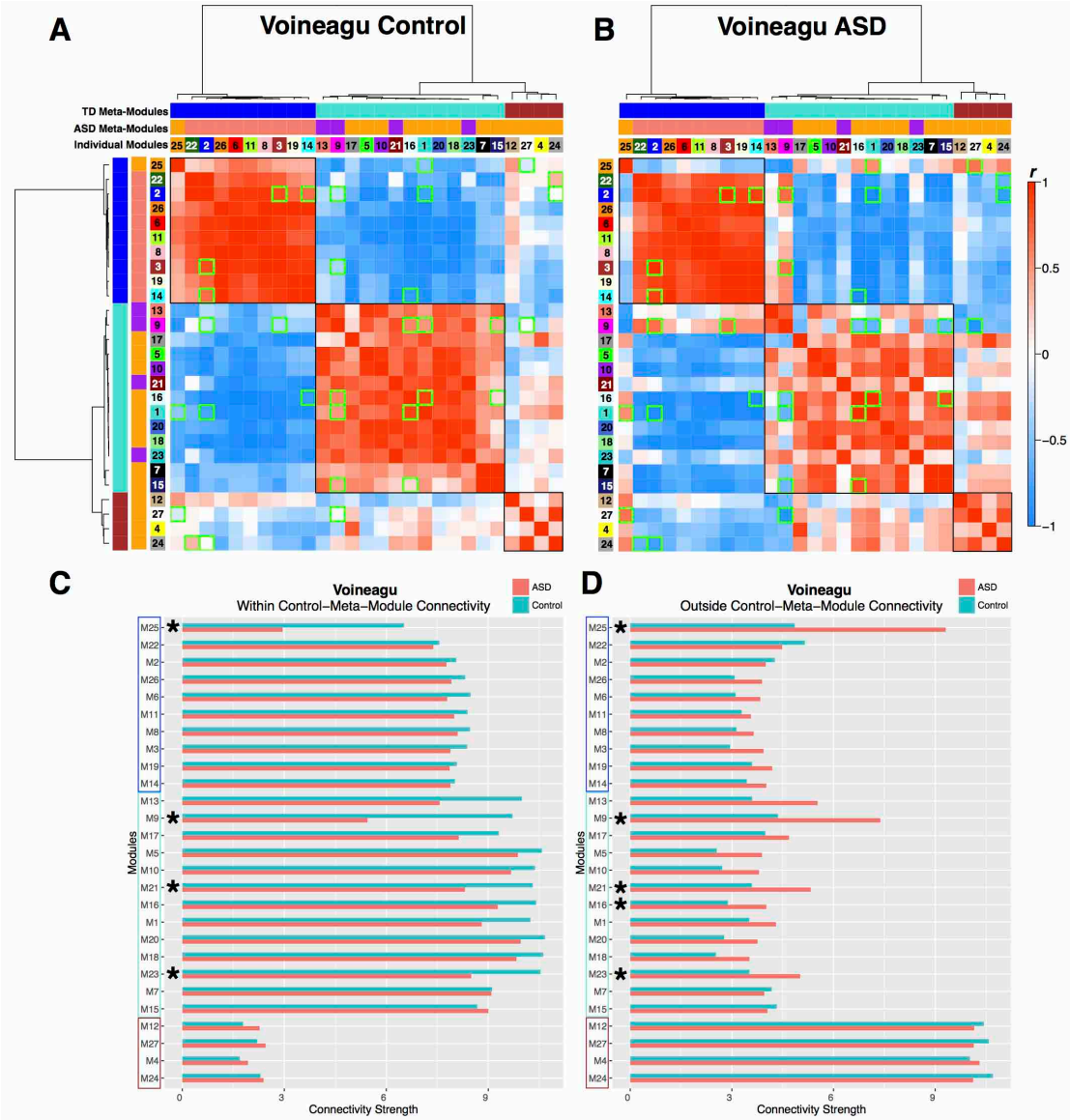

Fig 7: Module and meta-module connectivity differences within the Voineagu dataset. *Panels A and B illustrate eigengene networks as robust ME partial correlation matrices for the Voineagu Control (A) and ASD (B) groups. Red coloring within the matrices indicates increasing positive correlation strength, while blue coloring indicates increasing negative correlation strength; see colorbar for key*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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 below the dendrograms. 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. Panels C and D show 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.

Upon diving specifically into inter-modular connectivity differences we further observed specific between-module connections that are prominently affected in ASD (Fig 7A-B). 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.

Whereas the Voineagu dataset could be characterized by a small subset of modules that differed in within and outside meta-module connectivity, a vast majority of the modules in the Gupta dataset show such connectivity differences. Twenty-four out of all 27 modules showed evidence of decreased within meta-module connectivity and 18/27 modules were enhanced in outside meta-module connectivity in ASD (Fig 8C-D). This more global disorganization of eigengene network topology suggests that many aspects of how modules interact are disrupted in ASD (Fig 8A-B). However, upon diving into inter-module connectivity analyses, we found very few differences given the numerous shifts observed at the level of within and outside meta-module connectivity differences. Most of the specific between-module connectivity differences involved synaptic (M3, M5, M13), proteolysis, protein folding, and cell cycle (M11), and translation (M6) modules (Fig 8A-B). Here we find that this subset

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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 cannot be tied to very pronounced and specific differences within specific subsets of modules.

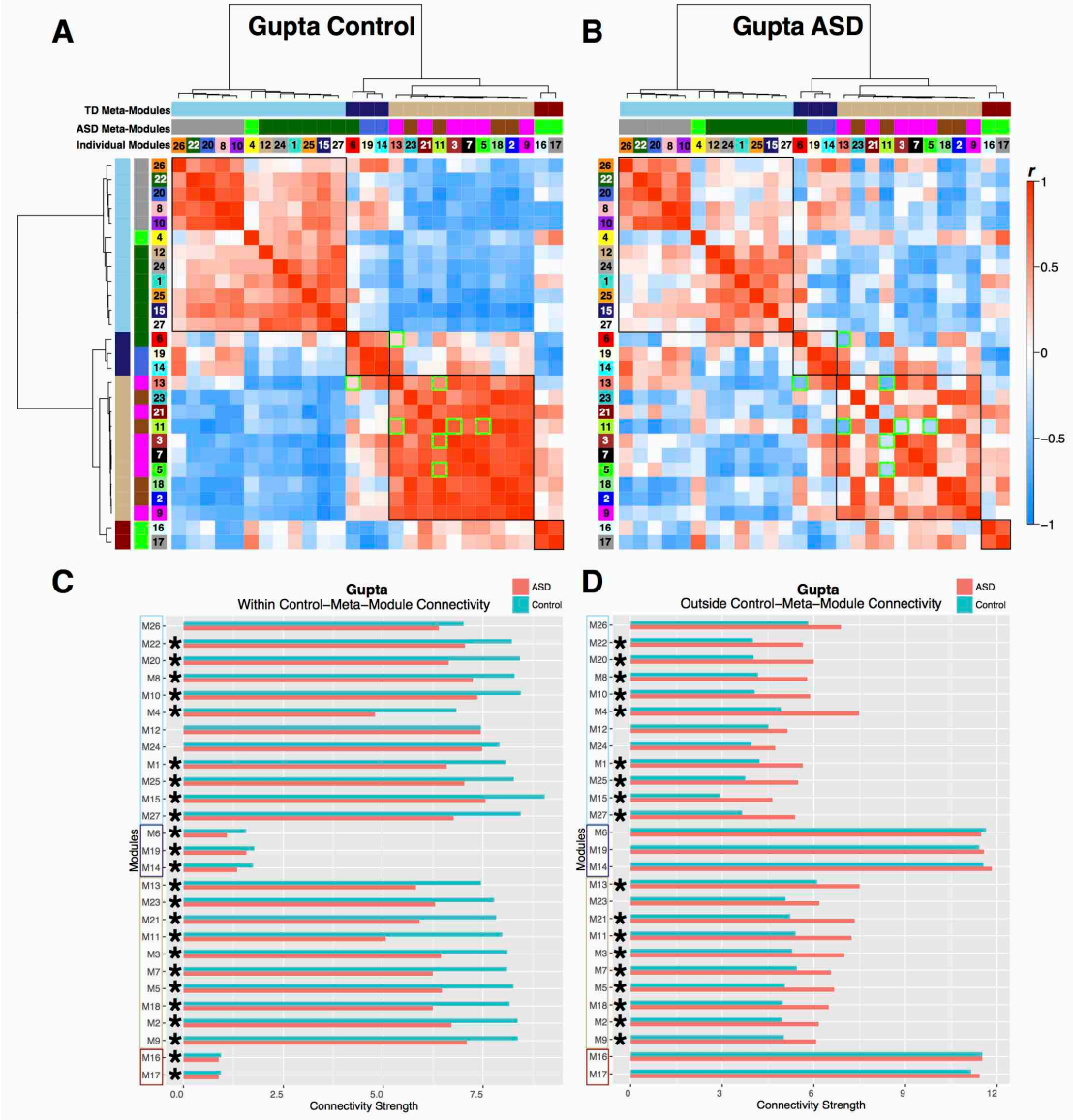

Fig 8: Module and meta-module connectivity differences within the Gupta dataset. *Panels A and B illustrate eigengene networks as robust ME partial correlation matrices for the Gupta Control (A) and ASD (B) groups. Red coloring within the matrices indicates increasing positive correlation strength, while 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 below the dendrograms. 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. Panels C and D show 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.*

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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.

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 individual co-expression modules, we expand on prior work[5, 6] by pinpointing several previously implicated and novel co-expression modules that are replicably dysregulated on-average across the datasets. Prior work has established that synaptic and immune/inflammation co-expression modules are dysregulated in ASD[5, 6, 8], and here re-analysis of this data in a consensus WGCNA framework conceptually replicates these findings. Extending this past work, we have specifically compared identical modules across datasets and show quantitative Bayesian evidence confirming that such findings are indeed replicable signals. Our work here also highlights the importance of several novel modules that are not implicated by prior work. First, unlike prior pinpointing of only one or two 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 is important as there may be subtle distinctions between them, particularly in how they interact with various other biological processes and pathways and also for the different cell types and cellular compartments they are associated with. 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 specific roles of M3 and M9 may be quite different in a systems biological context and the specificity of these roles only becomes apparent by looking at enrichment in different 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. Normatively these modules are negatively correlated in the Voineagu dataset, 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 process 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 hierarchical 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.

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[6], here we find evidence for 2 upregulated interferon signaling modules (M24, M27).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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 upregulated interferon signaling can be linked to both M1 and M2 microglia phenotypes and likely supports the idea that such aberrant processes may have differing impact. 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[27, 28] as well as enhancing pro-inflammatory states of microglia activation in ASD[34, 35, 36]. Recently, the complement system has been noted as a prominent player in the pathophysiology of schizophrenia, particularly for its role in synaptic pruning[37]. 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*[38, 39], as well as the important cap-dependent translation gene *EIF4E*[40, 41, 42, 43]. 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. 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

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

prominent signal in young toddlers with ASD and this signal is present alongside other upregulated immune/inflammation signals, particularly interferon signaling and phagocytosis[3]. 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[44]. Investigating this possible peripheral linkage to brain-relevant dysfunction in living patients using in-vivo techniques like functional and structural neuroimaging[45] 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[46]. 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).

By identifying disruption in the interaction between-modules, this work presents novel insights into how the transcriptome is hierarchically disrupted in ASD. 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. These modules also highly interact at the protein level and their disruption results in emergent phenomenon at a systems level that are not adequately characterized by viewing the modules in isolation. 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[5, 6]. 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[47] and could have key roles in ASD[16, 45, 48, 49]. 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 embedded roles very early in brain development[16, 49, 50].

Alongside this downregulation of important synaptic and neural developmental processes are coordinated upregulated biological phenomena (i.e. immune/inflammation processes, translation, etc) that when put together, could shed significant insights into higher-level systems biological disruptions in the ASD brain. 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, which in combination with several of these other phenomena (e.g., catabolism, translation, protein targeting and localization) suggest a highly coordinated set of processes affecting multiple cell types/compartments, but which may have substantial impact on those downregulated synaptic and neuronal processes

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

– as evidenced by the strong negative correlations between upregulated and downregulated modules. This evidence is generally in agreement with past theoretical ideas[16] 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)[34, 35, 36], 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 phenomena are emergent from 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 can not only produce new insights about systems level phenomena in ASD, but they can also feedback into reductionist frameworks that can produce significant insight into specific mechanisms (e.g., targeting hub genes for many of the dysregulated modules we have implicated; Table S2).

Finally, there are some caveats and limitations to consider in interpreting the results. First, the two datasets are different on a number of levels. Datasets were different with respect to the brain regions sampled, age range, and with regard to how gene expression was measured (i.e. microarray vs. RNAseq). Gene expression within specific subsets of genes can vary widely across brain regions[51]. We have made analytic attempts to reduce the effect of this heterogeneity via including brain region as a covariate but also in our approach to using consensus WGCNA to ensure that the detected modules are indeed present in both datasets despite differences in the brain regions sampled. Given the current scarcity of available post-mortem tissue, it is unlikely that this factor could be eliminated until large enough data exists to design definitive studies that can treat brain region as a separate variable of interest. Age ranges also varied across datasets, with the Gupta dataset possessing a much wider age range of individuals that extends into early childhood. It is well understood that cortical gene expression changes across the lifespan[52] and may interact in interesting ways with ASD diagnosis[8]. These age-related changes are effects that we would ideally one day like to study with large sample sizes. However, again due to the limited post-mortem data currently available, these types of studies are not attainable at large sample sizes that are well powered to identify very subtle effects. In the analyses, we have attempted to covary for age-related variation in order to reduce the impact of this variable on our results. However, because the majority of individuals in the sample are within the adolescent to adult age-range, it is likely that very different results would be observed at much earlier ages and consequently the interpretations we make of such results are restricted primarily to effects that show persistent pathophysiology throughout the life span in ASD[16]. It is worth noting that despite all of these differences, our analyses investigating replicable differential expression of gene modules demonstrates that some replicable signals are still present and robust enough to pop out despite all these factors. The success of this specific aspect of the analysis could be due in part to the fact that consensus WGCNA ensures that the detected modules are present in both datasets, and thus helps in making comparisons between datasets on specific co-expression modules. However, consensus WGCNA will not guarantee that datasets will have similar eigengene network and meta-modular organization, and here we found little evidence of preservation across Control datasets (Fig S2). This issue limits our ability to make some inferences about

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

replication with respect to the eigengene network results and instead these results should be treated as two separate looks at how eigengene network and meta-modular organization could vary in ASD.

In summary, our analyses highlight hierarchical disruption of the cortical transcriptome in ASD. Our approach allows for a better bird's eye view of how multiple pathophysiological processes may operate 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 March 15, 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.

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)[5]. 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/)[6]. 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[53]. 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[54]. A consensus WGCNA analysis was implemented in order to detect consensus modules for cross-dataset comparisons (implemented with the blockwiseConsensusModules function)[29]. 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 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)[55]. 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

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

models included diagnosis as the fixed effect of interest and additionally included age, sex, brain region, 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. Sample collection site in the Gupta dataset was highly correlated with diagnosis (Spearman rho = -0.51) and was not included as a covariate in these specific analyses in order to guard against the possibility that its inclusion would eliminate a large degree of variability associated with diagnosis. 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[56] 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[57] $q<0.05$ 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.

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[33] 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).

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[6]. The exception to this was lists of ribosomal subunit markers. These were obtained 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[58] 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, site). Partial correlation matrices were then converted into adjacency matrices that retain information about

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 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[59] to construct weighted graphs of the ME adjacency matrices for each group. These graphs are depicted using a spring embedded layout algorithm[60] 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[57] $q<0.05$ was used as the threshold for multiple comparisons correction.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 awarded to Prof. Courchesne.

Financial Disclosures

None of the authors have any potential conflicts of interest, financial or otherwise.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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

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

2. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. *Lancet Neurol*, (2015).

3. Pramparo T, *et al.* Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* **72**, 386-394 (2015).

4. Kong SW, *et al.* Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS One* **7**, e49475 (2012).

5. Voineagu I, *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380-384 (2011).

6. Gupta S, *et al.* Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat Commun* **5**, 5748 (2014).

7. Garbett K, *et al.* Immune transcriptome alterations in the temporal cortex of subjects with autism. *Neurobiology of disease* **30**, 303-311 (2008).

8. Chow ML, *et al.* Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. *PLoS Genet* **8**, e1002592 (2012).

9. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. *Neurology* **57**, 1618-1628 (2001).

10. Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. *PLoS One* **7**, e44736 (2012).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

11. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* **16**, 441-458 (2015).

12. Tang Y, Le W. Differential Roles of M1 and M2 Microglia in Neurodegenerative Diseases. *Mol Neurobiol* **53**, 1181-1194 (2016).

13. Cunningham CL, Martinez-Cerdeno V, Noctor SC. Microglia regulate the number of neural precursor cells in the developing cerebral cortex. *J Neurosci* **33**, 4216-4233 (2013).

14. Coull JA, *et al.* BDNF from microglia causes the shift in neuronal anion gradient underlying neuropathic pain. *Nature* **438**, 1017-1021 (2005).

15. Cherry JD, Olschowka JA, O'Banion MK. Neuroinflammation and M2 microglia: the good, the bad, and the inflamed. *J Neuroinflammation* **11**, 98 (2014).

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

17. Martinez FO, Gordon S. The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000Prime Rep* **6**, 13 (2014).

18. Atladottir HO, *et al.* Maternal infection requiring hospitalization during pregnancy and autism spectrum disorders. *J Autism Dev Disord* **40**, 1423-1430 (2010).

19. Lee BK, *et al.* Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. *Brain Behav Immun* **44**, 100-105 (2015).

20. Oskvig DB, Elkahloun AG, Johnson KR, Phillips TM, Herkenham M. 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 (2012).

21. Coiro P, *et al.* Impaired synaptic development in a maternal immune activation mouse model of neurodevelopmental disorders. *Brain Behav Immun* **50**, 249-258 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

22. Elmer BM, Estes ML, Barrow SL, McAllister AK. MHCI requires MEF2 transcription factors to negatively regulate synapse density during development and in disease. *J Neurosci* **33**, 13791-13804 (2013).

23. Le Belle JE*, et al.* 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 (2014).

24. Choi GB*, et al.* The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring. *Science (New York, NY* **351**, 933-939 (2016).

25. Stoner R*, et al.* Patches of disorganization in the neocortex of children with autism. *N Engl J Med* **370**, 1209-1219 (2014).

26. Glynn MW*, et al.* MHCI negatively regulates synapse density during the establishment of cortical connections. *Nature neuroscience* **14**, 442-451 (2011).

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

28. Stevens B*, et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164-1178 (2007).

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

30. Oldham MC*, et al.* Functional organization of the transcriptome in human brain. *Nature neuroscience* **11**, 1271-1282 (2008).

31. Oldham MC, Horvath S, Geschwind DH. 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 (2006).

32. Szklarczyk D*, et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic acids research* **43**, D447-452 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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

34. Morgan JT, *et al.* Microglial activation and increased microglial density observed in the dorsolateral prefrontal cortex in autism. *Biol Psychiatry* **68**, 368-376 (2010).

35. Vargas DL, Nascimbene C, Krishnan C, Zimmerman AW, Pardo CA. Neuroglial activation and neuroinflammation in the brain of patients with autism. *Annals of neurology* **57**, 67-81 (2005).

36. Suzuki K, *et al.* Microglial activation in young adults with autism spectrum disorder. *JAMA Psychiatry* **70**, 49-58 (2013).

37. Sekar A, *et al.* Schizophrenia risk from complex variation of complement component 4. *Nature*, (2016).

38. Kelleher RJ, 3rd, Bear MF. The autistic neuron: troubled translation? *Cell* **135**, 401-406 (2008).

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

40. Neves-Pereira M, *et al.* Deregulation of EIF4E: a novel mechanism for autism. *J Med Genet* **46**, 759-765 (2009).

41. Santini E, *et al.* Exaggerated translation causes synaptic and behavioural aberrations associated with autism. *Nature* **493**, 411-415 (2013).

42. Gkogkas CG, *et al.* Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* **493**, 371-377 (2013).

43. Yang G, Smibert CA, Kaplan DR, Miller FD. An eIF4E1/4E-T complex determines the genesis of neurons from precursors by translationally repressing a proneurogenic transcription program. *Neuron* **84**, 723-739 (2014).

44. Louveau A, *et al.* Structural and functional features of central nervous system lymphatic vessels. *Nature* **523**, 337-341 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

45. Pramparo T, *et al.* Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol Syst Biol* **11**, 841 (2015).

46. Lombardo MV, *et al.* Different functional neural substrates for good and poor language outcome in autism. *Neuron* **86**, 567-577 (2015).

47. Clancy B, Darlington RB, Finlay BL. Translating developmental time across mammalian species. *Neuroscience* **105**, 7-17 (2001).

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

49. Courchesne E, *et al.* Neuron number and size in prefrontal cortex of children with autism. *JAMA : the journal of the American Medical Association* **306**, 2001-2010 (2011).

50. Casanova EL, Casanova MF. Genetics studies indicate that neural induction and early neuronal maturation are disturbed in autism. *Frontiers in cellular neuroscience* **8**, 397 (2014).

51. Hawrylycz M, *et al.* Canonical genetic signatures of the adult human brain. *Nature neuroscience* **18**, 1832-1844 (2015).

52. Miller JA, *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).

53. Miller JA, *et al.* Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* **12**, 322 (2011).

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

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

58. Wager TD, Keller MC, Lacey SC, Jonides J. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage* **26**, 99-113 (2005).

59. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* **48**, 1-18 (2012).

60. Fruchterman TMJ, Reingold EM. Graph drawing by force‐directed placement. *Software: Practice and experience* **21**, 1129-1164 (1991).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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.

Figure Legends

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).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 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 March 15, 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 –log10 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).

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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: Eigengene networks, meta-modules, and correlations between dysregulated modules within the Voineagu dataset. This figure shows 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). Panels C and D show the differentially expressed modules only as correlation matrices. The strength of the correlation is shown within each cell and via the coloring (hot colors indicating higher correlations, cooler colors indicating lower correlations). The red outline denotes ASD-upregulated modules while the blue outline denotes ASD-downregulated modules.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 5: Eigengene networks, meta-modules, and correlations between dysregulated modules within the Gupta dataset. This figure shows 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). Panels C and D show the differentially expressed modules only as correlation matrices. The strength of the correlation is shown within each cell and via the coloring (hot colors indicating higher correlations, cooler colors indicating lower correlations). The red outline denotes ASD-upregulated modules while the blue outline denotes ASD-downregulated modules.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 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[33] 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.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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: Module and meta-module connectivity differences within the Voineagu dataset. Panels A and B illustrate eigengene networks as robust ME partial correlation matrices for the Voineagu Control (A) and ASD (B) groups. 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 below the dendrograms. 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. Panels C and D show 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.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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 8: Module and meta-module connectivity differences within the Gupta dataset. Panels A and B illustrate eigengene networks as robust ME partial correlation matrices for the Gupta Control (A) and ASD (B) groups. 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 below the dendrograms. 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. Panels C and D show 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.

bioRxiv preprint doi: https://doi.org/10.1101/042937; this version posted March 15, 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

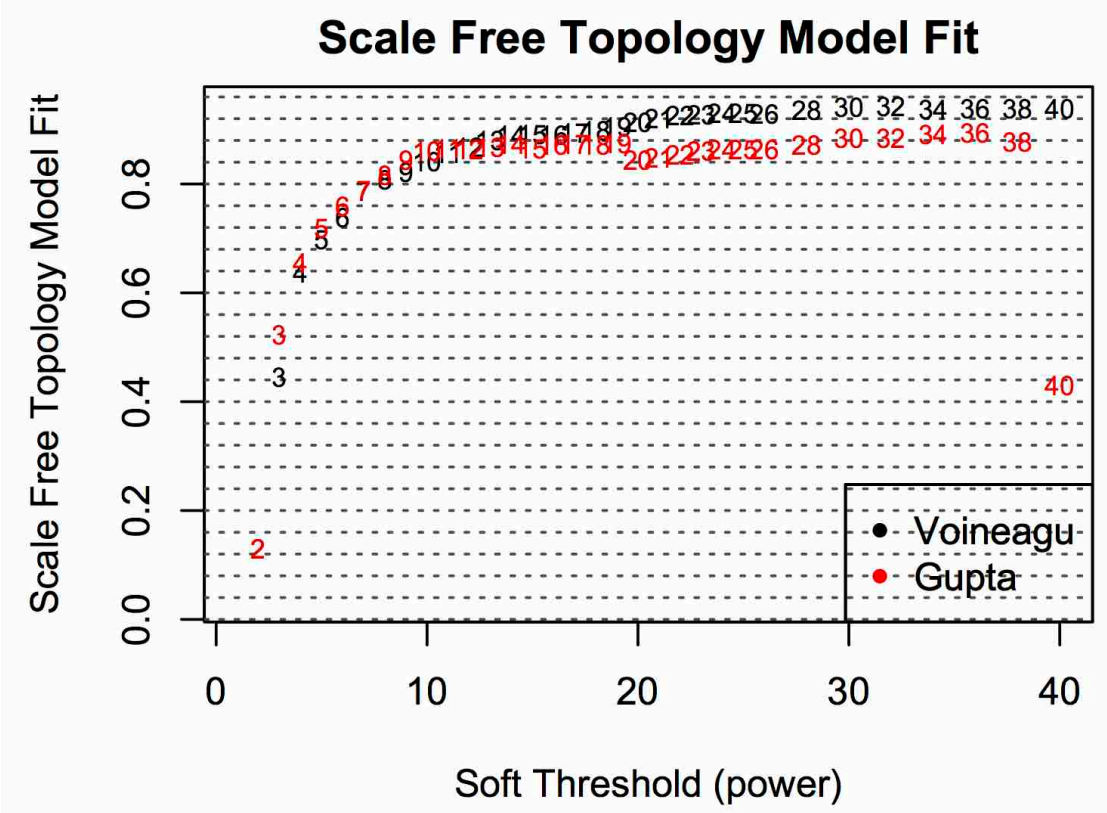

Fig S1: Scale-free topology model fit across a range of soft power thresholds

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 March 15, 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.

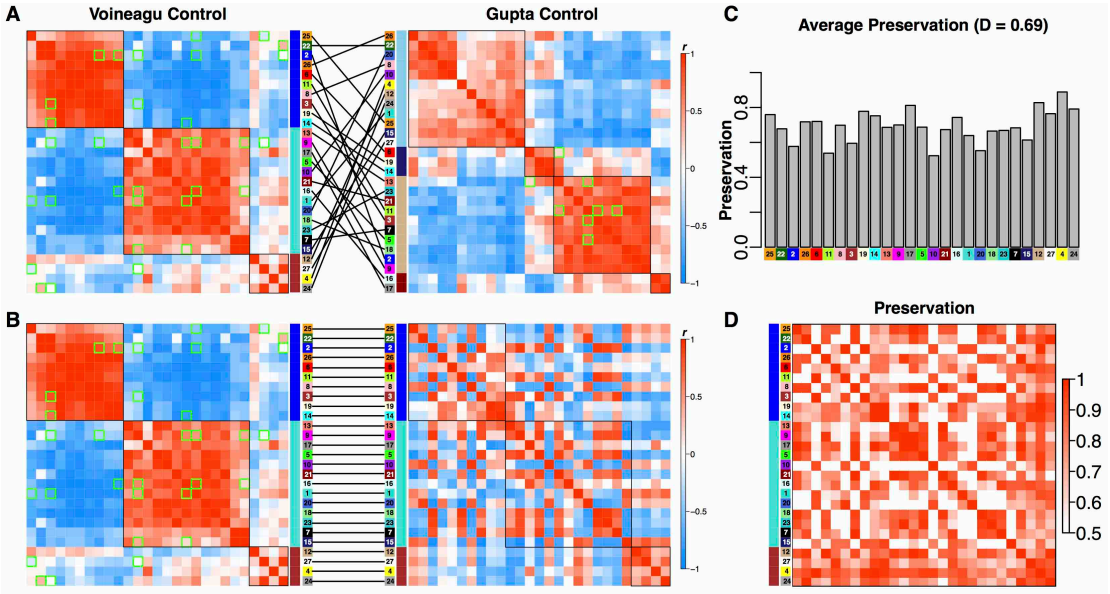

Fig S2: Preservation of eigengene networks in the TD group

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 March 15, 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