Genes involved in cholesterol cascades are linked to brain connectivity in one third of autistic patients

Javier Rasero[1,], Antonio Jimenez-Marin[2,3,*], Ibai Diez[4,5,6], Mazahir T. Hasan[7,8],*

and Jesus M. Cortes[2,8,9,§]

1. Cognitive Axon Laboratory, Department of Psychology, Carnegie Mellon University, Pittsburgh, United States of America.

2. Computational Neuroimaging Lab, Biocruces-Bizkaia Health Research Institute, Barakaldo, Spain.

3. Biomedical Research Doctorate Program, University of the Basque Country (UPV/EHU), Leioa, Spain.

4. Department of Radiology, Division of Nuclear Medicine and Molecular Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, United States of America.

5. Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, United States of America.

6. Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, United States of America.

7. Laboratory of Brain Circuits Therapeutics, Achucarro Basque Center for Neuroscience, Leioa, Spain.

8. IKERBASQUE, The Basque Foundation for Science, Bilbao, Spain.

9. Department of Cell Biology and Histology, University of the Basque Country (UPV/EHU), Leioa, Spain.

* Equal first-author contribution

§ Correspondence: jesus.m.cortes@gmail.com

Title: 90 chars
Abstract: 331 words
Introduction: 709 words
Materials and Methods: 2797 words
Results: 2620 words
Discussion: 1939 words
Figures: 3
Tables: 1
References: 93

Abstract

The large heterogeneity in the symptomatology and severity of autism spectrum disorder (ASD) is a major drawback for the design of effective therapies. Beyond behavioral phenotypes, subtype stratification strategies that can be applied to large populations are needed, these combining different neurobiological characteristics and based on the large-scale organization of the human brain, as well as neurogenetic fingerprints. Here, we make use of ABIDE, the largest publicly available database of functional neuroimaging in ASD, to which we have applied rigorous data harmonization between the different scanning institutions in order to employ analyses based on consensus clustering and to evaluate the patterns of brain connectivity. As a result, we identified three subtypes of ASD, the first of which was characterized by a mixture of hyper- and hypo-connectivity, stronger network segregation and weaker integration, and it represented approximately 13% of all patients. The second subtype was associated with 31% of the patients, and it was characterized by hyperconnectivity but no topological differences with respect to the group of typically developing controls. The third was the most numerous subtype, assigned to 52% of all patients, and it was characterized by hypoconnectivity, decreased network segregation and increased integration. We also defined a neurobiological signature for each of these subtypes, detailing the connectivity and structures most specific to each subtype. Strikingly, at the behavioral level, none of the neuropsychological scores used in the diagnosis of ASD is capable of differentiating any of the subtypes from the other two. Finally, we use the Allen Human Brain Atlas of gene transcription brain maps to show that subtype 2 has an extraordinary enrichment in biological processes related to the synthesis, regulation and transport of cholesterol and other lipoproteins, one of the mechanisms previously attributed to ASD. We also show that this lipid-susceptible ASD subtype could be represented by the dysfunctionality of the network, unlike the other two subtypes that have more structural alterations in the connectome. Thus, our study provide compelling support for prospects of cholesterol-related therapies in this subset of autistic individuals.

Introduction

Autism is a very heterogeneous disorder, producing symptoms that range from impaired social communication and/or interactions, to manifestations of restricted and/or repetitive behavioural patterns, interests and activities (*1–3*). Due to the wide heterogeneity in symptomatology, and as recommended in The Diagnostic and Statistical Manual of Mental Disorders (DSM–5), this disparate condition is referred as autism spectrum disorder (ASD), in which the term "spectrum" emphasizes the variation in the type and severity of symptoms (*4*). We know that ASD results from complex interactions between genetic, epigenetic and environmental factors during development (*5–9*), yet a clear understanding of the factors and mechanisms that produce these precise phenotypic patterns remain largely unknown.

Several mechanisms have been proposed to underlie ASD, including an imbalance in excitation/inhibition (E/I) during development (*10*, *11*). However, the factors driving this disease are not well understood, making therapeutic interventions to restore the E/I balance in ASD a major challenge (*12*). Motivated by the Smith-Lemli-Opitz Syndrome (SLOS), a genetic condition with ASD symptoms characterized by impaired cholesterol biosynthesis (*13*), factors related to the synthesis, metabolism and transport of lipids, cholesterol and other lipoproteins might plays a relatively general role in ASD (*14–16*). Indeed, a therapeutic intervention based on cholesterol supplements has been attempted (*17*), although the final results of these studies and the therapeutic benefits at the behavioural level are not yet clear. In fact, these two mechanisms, E/I imbalance and lipid deficits appear to be related (*18–20*), although a precise causal correspondence between these two phenomena is also unknown.

At the macroscopic level, several brain structures appear to be implicated in ASD, many forming part of the so-called social network of the brain (*21*) and including the primary motor cortex,

fusiform, amygdala, cerebellum, insula, somatosensory and anterior cingulate cortex (*22–24*), the latter playing an important role in the large anatomical heterogeneity observed in ASD (*25*). When considering brain networks, the frontal, default mode and salience have been implicated in ASD (*23*, *26–30*). Moreover, the neuroanatomical structures implicated in ASD are not static but they undergo changes throughout development (*31–33*), and this also happens in social functioning and communication (*34*).

Here, we follow on from previous studies based on consensus clustering applied to brain connectivity matrices (*35*, *36*), and we searched for ASD subtypes by representing the multivariate connectivity pattern of given brain regions, encoded in a vector of dimension equal to the number of regions, and where each component is defined by the connectivity between the given region and any other. Consequently, if two subjects belong to the same subtype, it implies that each brain structure connects to the rest of the brain in a similar way, revealing similarity between the brains of the patients within the same subtype through multivariate aspects of large-scale brain connectivity. We performed our subtyping analyses on 880 autistic patients, based on functional magnetic resonance imaging (fMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) repository (*37*). This data was previously harmonized (*38–41*) to overcome the different sources of variability in the neuroimaging data, such as scanning institution, sex and age.

Linking the connectivity-based ASD subtypes to their neurogenetic signature may also shed some light on the biological mechanisms that differentiate each subtype. As such, we must take into consideration that ASD is an extensive polygenic condition, thought to involve about 944 genes at present according to the SFARI gene human-database (as of July 15th, 2020), see also (*42*). To date, a precise association between the entire transcriptome and brain connectomics in ASD patients has yet to be explored, although an association between transcriptomics and brain morphology was

recently assessed (*43*), indicating that genes that are enriched for synaptic transmission and downregulated in individuals with autism were associated with variations in cortical thickness. Here, we take a step forward and ask if it is possible to perform a correspondence analysis between the large-scale connectivity patterns and the Allen Human Brain Atlas (AHBA) of whole-brain transcriptional data (*44*) for each ASD subtype, following a similar methodology to that used previously (*45–47*). Our hypothesis is that by identifying the genes whose expression maps coincide more closely with the connectivity maps, we might better understand the neurogenetic signature and neurobiological mechanisms associated with each subtype of ASD, a major challenge in this field.

Materials and Methods

Participants

A total of N=1890 subjects from the ABIDE repository participated in this study, of which 880 were ASD patients and 1010 were typically developing controls (TDCs). The data comes from 24 different institutions (see Table S1) and for each Institution, we registered the centre's name, number of subjects contributed, mean age, sex distribution and the number of ASD cases. For each participant, we obtained both anatomical and fMRI data. The acquisition parameters for each scanning institution can be found at http://fcon_1000.projects.nitrc.org/indi/abide/. Moreover, we assessed cognitive performance and disease severity using the Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R) and Intelligence Quotient (IQ) sub-scores: verbal IQ (VIQ), performance IQ (PIQ) and full IQ (FIQ). A brief description of all this neuropsychological information can be found in Table S2.

Neuroimaging pre-processing

A state-of-the-art pre-processing pipeline was adopted using FSL, AFNI (*48*) and MATLAB. We first applied slice-time correction and then, each volume was aligned to the middle volume to correct for head motion artefacts, which was followed by intensity normalization. We next

regressed out 24 motion parameters, as well as the average cerebrospinal fluid (CSF) and the average white matter signal. A band-pass filter was applied between 0.01 and 0.08 Hz, and linear and quadratic trends were removed. All voxels were spatially smoothed with a 6 mm FWHM and finally, FreeSurfer was used for brain segmentation and cortical parcellation. A total of 86 regions were generated, with 68 cortical regions from the Desikan-Killiany Atlas (34 in each hemisphere) and 18 subcortical regions (left/right thalamus, caudate, putamen, pallidum, hippocampus, amygdala, accumbens, ventral DC and cerebellum). The parcellation for each subject was projected to the individual functional data and the mean functional time series of each region was computed. Furthermore, from the 86 regions we reordered those that are part of the social network (for details see Table S3). Finally, a connectivity matrix was obtained for each subject by Fisher z-transforming the Pearson correlation coefficients between the regions' time series.

Data Harmonization

To harmonize our multi-institution data, we used an in-house implementation of Combat (https://pypi.org/project/pycombat), adjusting batch effects by linear mixed modelling and the use of Empirical Bayes methods (*39*). In our case, batch effects may reflect the different set-ups for image acquisition at each institution included (e.g., MRI scanner manufacturer, different antenna and/or software, gradient coils, magnet field strength, etc.). Let Y_{ijk} represent the value of the connectivity entry k for subject j at institution i. Combat adjusts the Y_{ijk} data by estimating the coefficients present in the following linear mixed model:

$$Y_{ijk} = \alpha_k + X\beta_k + \gamma_{ik} + \delta_{ik}\epsilon_{ijk}, \quad \text{(Eq. 1)}$$

where α_k is the fixed intercept, β_k the fixed slopes for the variables in a design matrix X, and γ_{ik} and δ_{ik} the location and scale institution factors modelled as random effects. One of the strong points of Combat is the use of an empirical Bayes (EB) approach to better estimate γ_{ik} and δ_{ik}, an

6

iterative step that is particularly relevant when sample sizes are small. Specifically, it assumes that the two parameters controlling the random effects are sampled from the following prior distribution:

$$\gamma_{ik} \sim N(\gamma_i, \tau_i^2) \qquad \text{(Eq. 2)}$$

$$\delta_{ik} \sim InverseGamma(\lambda_i, \theta_i). \qquad \text{(Eq. 3)}$$

The hyperparameters γ_i, τ_i^2, λ_i and θ_i are empirically estimated using an expectation-maximisation (EM) procedure as described in (39). Thus, the harmonized data $\hat{Y}_{ijk}$ read:

$$\hat{Y}_{ijk} = \frac{Y_{ijk} - \hat{\alpha}_k - X\hat{\beta}_k - \hat{\gamma}_{ik}}{\hat{\delta}_{ik}} + \alpha_k + X\hat{\beta}_k, \quad \text{(Eq. 4)}$$

where $\hat{\alpha}_k$, $\hat{\beta}_k$, $\hat{\gamma}_{ik}$ and $\hat{\delta}_{ik}$ are the fitted coefficients present in Eq. 1. In the original Combat implementation, the design matrix X encodes the effects of interest that we want to preserve during the harmonization process. Potential covariates of interest may include age, sex and group-level variables. However, since one of our main goals is to find ASD subtypes, in addition to batch/institution effects, we may need to remove the additional sources of variability in the data that could strongly affect the connectivity values, thereby hiding the true underlying structure of each ASD subtype. We can easily incorporate this into the Combat framework by modifying Eq. 1 as follows:

$$Y_{ijk} = \alpha_k + X\beta_k^x + C\beta_k^c + \gamma_{ik} + \delta_{ik}\epsilon_{ijk}, \qquad \text{(Eq. 5)}$$

where we have now explicitly separated the effects to keep X from other possible sources of covariation C, such as sex and age, which we want to remove during the harmonization process. Such an extension to Combat has already been proposed and proven to remove additional confounding bias (40). As a result, after fitting Eq. 5, the harmonized dataset was evaluated as follows:

$$\hat{Y}_{ijk} = \frac{Y_{ijk} - \hat{\alpha}_k - X\hat{\beta}_k^x - C\hat{\beta}_k^c - \hat{\gamma}_{ik}}{\hat{\delta}_{ik}} + \alpha_k + X\hat{\beta}_k^x \qquad \text{(Eq. 6)}$$

In our scenario, we will consider the group factor (TDC vs ASD) as the variable of interest after harmonization, and age and sex as the covariate (C) effects to be removed, in addition to the institution effects encoded in γ_{ik} and δ_{ik}.

Consensus clustering

ASD subtyping was investigated using a consensus clustering approach applied to brain connectivity matrices (*35*, *36*). First, for each brain region i we defined a distance matrix for each pair of ASD subjects as:

$$\left(D^i\right) \equiv d_{uv}^i = \sqrt{2(1 - r_{uv})}, \quad \text{(Eq. 7)}$$

where r_{uv} is the Pearson correlation between the connectivity pattern of subjects u and v for region i. The connectivity pattern is a vector of dimension equal to the number of regions, where each component is defined as the amount of connectivity between the given region and any other. Then, each distance matrix D^i is partitioned into k groups of subjects using a k-medoids clustering method (*49*), and the resulting clustering information encoded into an adjacency matrix, whose entries are 1 if a pair of subjects belongs to the same cluster and zero otherwise. Subsequently, a $N \times N$ consensus matrix C was evaluated by averaging this information across the nodes. Hence, the entries of C_{uv} indicate the number of partitions in which subjects u and v are assigned to the same group, divided by the number of partitions. Eventually, the consensus matrix is averaged over the k range in the interval (2-20), so that information about the underlying structure at different resolutions is combined in the final consensus matrix, for more details see (*35*).

ASD subtyping

The consensus matrix C was further used to compute a modularity matrix as follows:

$$B = C - \gamma \cdot P, \quad \text{(Eq. 8)}$$

8

where P is the expected co-assignment matrix, uniform as a consequence of the null ensemble strategy obtained by repeating the permutation of labels 1000 times, and γ is the resolution parameter, which we set to 1 as in the Newman and Girvan scenarios and that corresponds to the maximum modularity value (*50*). Such a modularity matrix B encodes all the information about the interaction between subjects at different levels. As a result, one could now define any distance quantity and apply it to this matrix, subsequently using a k-medoids analysis to assess clustering. Instead, we fed this B matrix into a generalized Louvain method for community detection using modularity matrices (https://github.com/GenLouvain/GenLouvain), which allows an optimal output partition to be obtained without specifying the number of desired clusters.

Separability of the subtypes with respect to TDC

To assess the separability of brain connectivity profiles between each ASD subtype and the TDC group, we utilized the *Multivariate Distance Matrix Regression* (MDMR) method, as described previously (*51*, *52*). First, the distance matrix defined in Eq. 7 was calculated for a fixed brain region i and then, for each brain region i and predictor column x in a design matrix X, the MDMR yields a pseudo-F statistic F_x^i which reads:

$$F_x^i = \frac{tr(H_x G^i)/(m_x)}{tr[(I-H)G^i]/(N-m)}, \qquad \text{(Eq. 9)}$$

where tr indicates the trace operator, N the number of observations, m the degrees of freedom in X, m_x the degrees of freedom of predictor x, H_x the isolated effect of predictor x from the usual matrix $H = X(X^T X)^{-1}X^T$, and G^i the so-called Gower matrix built from the distance matrix D^i (*53*). Like the F-estimator in a standard ANOVA analysis, Eq. 9 assesses the variance explained by a predictor variable with respect to the unexplained variance. However, the pseudo-F statistic is not distributed like the usual Fisher's F-distribution under the null hypothesis and therefore, the related p-values have to be computed by a permutation procedure (*52*). Finally, in order to estimate how

much variability can be attributed to each predictor, a pseudo-R^2 effect size can be computed by dividing the numerator without the degrees of freedom in Eq. 9 by the total sum of squared pairwise distances in the Gower matrix, *i.e.*: $R_x^2 = \frac{tr(H_x G)}{tr(G)}$. Similar to standard linear models, this effect size quantifies the proportion of the total sum of squares that can be explained by the predictors.

Graph topological metrics

The connectivity matrices between each ASD subtype were compared using four metrics obtained from graph theory: Transitivity (*segregation*), global efficiency (*integration*), assortativity (*resilience*), and small-worldness (*communication*). Transitivity is defined as the ratio of triangles to triplets in the network, global efficiency as the average inverse shortest path length, and assortativity as the correlation between the degrees of all nodes at two opposite ends of a link (*54*). Small-world networks are characterized as simultaneously highly integrated and segregated networks (*55*), and they can be detected when $\lambda \sim 1$ and $\gamma > 1$ (*56*), with $\gamma = \frac{C}{C_{random}}$ and $\lambda = \frac{L}{L_{random}}$ representing the ratio of the average clustering coefficient and shortest path lengths of the observed network with respect to randomized networks, respectively (*57*). In our case, randomized networks were achieved by rewiring each node five times while preserving the same number of nodes, edges and degree distribution of the original network. Taking into account these considerations, we evaluated the amount of small-worldness in each connectivity matrix using the ratio $\frac{\gamma}{\lambda}$, a positive quantity greater than 1 for small-world networks, whereby the higher the value the more small-world is the network.

Thus, a common practice to get rid of spurious correlations, that is correlations that they could mask other nodes in the network of greater importance, we binarized the harmonized connectivity matrices by applying a continuous range of proportional thresholds to maintain the same density of

10

stronger (positive) connections across all the individual networks (*58*). Specifically, the density of network links was varied continuously between 5% and 30%. Finally, we evaluated the topological differences between the different pairs of groups (ASD subtypes and TDC) for a particular threshold of 10% using an ordinary least square (OLS) regression, while controlling for age, sex, VIQ, FIQ, PIQ and the overall functional connectivity. This last confounder variable was included to avoid overestimating topological differences in the network organization due to the proportional thresholding of connectivity matrices.

Neurobiological signature

To define the connectivity of structures that are most specific to each subtype, and that can therefore provide a neurobiological signature for each subtype, we made use of a machine learning approach, namely a Support Vector Machine (SVM) with a linear kernel and a Recursive Feature Elimination (RFE) procedure (*59*). As the classification problem that we want to solve is multiclass, given that there are three subtypes, we adopted a strategy of classifying each subtype against the rest, the so-called one-versus-rest strategy, fitting a single classifier per subtype with observations within that subtype as the positive class and the rest of observations as the negative class. RFE is one of the wrapper methods for feature selection. Specifically, one first computes the weights for the full set of features (here, the links in the connectivity matrices). Subsequently, the feature with the lowest (absolute) weight is eliminated, and this process is repeated until a desired number of features are achieved. Since this number is in principle arbitrary, we embedded this process into a 5-fold cross-validation scheme, such that in each feature elimination step, the averaged out-of-sample balanced accuracy is computed. The number of links with the highest cross-validation performance is the one we finally take as valid. To reduce the complexity of each resulting circuit while maintaining a competent performance, we also apply the "1-Standard Error Rule" for feature selection, which chooses the fewest features within a standard error (SE) similar to the highest

accuracy achieved (*60*). Finally, the specific circuit formed by the resulting links is the one we will take as the neurobiological signature of each subtype.

Transcriptomics

To build brain transcription maps, we took advantage of the publicly available data in the AHBA (*44*). The dataset consisted of MRI images, and a total of 58,692 microarray-based transcription profiles of about 20,945 genes sampled from 3,702 different regions across the brains of six humans. To pool all the transcription data into a single brain template, we followed a similar procedure to that employed elsewhere (*61*). First, to re-annotate the probes to genes we made use of the re-annotator toolkit (*62*). Second, we removed those probes with insufficient signal by looking at the sampling proportion (SP), which was calculated for each brain as the ratio between the samples with a signal greater than the background noise divided by the total number of samples. Probes with a SP lower than 70% in any of the six brains were removed from the analysis, thereby ensuring sufficient sampling power in all the brains. After that, we chose the value of the probe for each gene with the maximum differential stability (DS), accounting for the reproducibility of gene expression across brain regions and individuals. This was also calculated using spatial correlations similar to those employed previously (*63*) but using the Desikan-Killiany atlas, the regions of which were eroded with a Gaussian kernel with a full width at half maximum (FWHM) equal to 2 mm, thereby eliminating false-positive sampling sites (i.e.: those that do not belong to the region of interest but to one in the neighborhood). To remove the inter-subject differences, the transcription values for each gene and brain were transformed into Z-scores, and pooled together from the six different brains, obtaining a single map using the MNI coordinates provided in the dataset. Finally, to eliminate the spatial dependencies of the transcription values at the sampling sites (i.e.: to correct for the fact that the nearest sites have better correlated transcription), we obtained a single transcription value for each region in the Desikan-Killiany atlas by calculating the median of all the values belonging to a given region.

Association between subtypes and transcriptomics

We first computed the statistical spatial maps per subtype that were later associated with the transcriptomics data. Specifically, we computed a distance matrix per region of interest (ROI) using the whole-brain connectivity patterns in our harmonized functional matrices. Subsequently, we performed a MDMR for each of these distance matrices to assess the differences between each subtype and the TDC group, while controlling for age, sex and IQ scores (FIQ, VIQ and PIQ). As a result, we obtained a spatial map of F-statistics per subtype. For each gene, we then calculated a similarity index using the Pearson correlation coefficient between F-values and transcriptomics, the two variables represented in vectors with a dimension equal to the number of regions in the atlas. This procedure was done separately for each of the different subtypes.

Identification of relevant genes and gene ontology

Relevant genes were identified by combining the results from two different strategies. The first was the data-driven strategy (DDS), an exploratory-orientated approach to identify the genes with the strongest transcription correlation with the F-statistic in each subtype. In particular, we chose genes with similarity index values corresponding to $|Z| > 2$, *i.e.*: outliers of the correlation distribution in both the negative and positive tails. Genes in the positive tail ($Z > 2$) were designated as *pos-corr* genes (P), whereas those in the negative tail ($Z < -2$) were considered *neg-corr* genes (N). In principle, this is an arbitrary choice that guarantees the selection of genes that are outliers in the two tails, and for which the Gaussian case would correspond to choosing those genes beyond the 95% Confidence Interval (CI)/with a percentile greater the 95% or less than 5%. Therefore, P genes were systematically expressed more strongly than other genes in the brain regions with a more pronounced F, that is, in the regions where connectivity differs most relative to the corresponding TDC, whereas the N genes were expressed much more weakly than the rest. Subsequently, for each phenotype and tail we performed a gene ontology *GO biological process* (*64*) and *Reactome*

13

pathways (*65*) overrepresentation test using PANTHER v15.0 (http://pantherdb.org/), with the entire Homo Sapiens genome as the reference list, and using a Fisher's Exact test with a Bonferroni correction (p < 0.05). To enhance the interpretability, we only reported enrichment ≥2-fold.

The second strategy is the hypothesis-driven strategy (HDS), a more confirmatory-orientated analysis performed on genes from https://gene.sfari.org/database/human-gene/ that belong to Category 1, *i.e.:* non-syndromic genes with a high confidence of being implicated in ASD due to the presence of at least three *de novo* likely-gene-disrupting mutations reported in the literature. This list includes 192 genes and it is also coincident with that published elsewhere (*66*). Thus, for the genes that were also P or N genes, their statistical significance was assessed by surrogate-data testing. The BrainSMASH tool (*67*) was used to build null-distributions by generating 10,000 random maps with the same spatial autocorrelation as that for the F-statistic map.

Results

A total of 880 ASD patients and 1010 TDCs participated in this study, the workflow of which is shown in figure 1. In summary, our analysis was based on the following steps: 1. Data preparation, pre-processing and harmonization to eliminate the variability associated with the differences in scanning at each institution, or the participant's age and sex; 2. ASD Subtyping achieved by applying consensus clustering to brain connectivity matrices; 3. Characterization of connectivity as hyperconnectivity, hypoconnectivity or a mixture of the two; 4. Drawing associations between transcriptomics and the F-statistics maps, accounting for the region separability of each ASD subtype and the TDCs. The results corresponding to each of these steps will be described in detail.

(Figure 1 here)

Data harmonization

14

We ran several exploratory tests to assess the differences in our dataset related to the scanning institution, sex, age and diagnosis (ASD vs TDC). From a demographic point of view, no statistical differences (two-sided t-test $p = 0.82$) were found between the age distributions in the TDC ($\mu = 16.03$, $\sigma = 8.62$) and ASD group ($\mu = 16.12$, $\sigma = 9.17$). By contrast, there was a significant difference in the male:female ratio between these two groups (Fisher's exact test $p = 1.29 \times 10^{-7}$). At the connectivity level, we tested the association among individuals of each link in the functional connectivity matrix with respect to age and found that 2236 edges had a significant association after correcting by false discovery rate (FDR), yet this may be strongly influenced by the large size of our dataset as the effect sizes were generally small (mean r = 0.07, σ = 0.05, max r = 0.319). Similarly, a t-test indicated that 565 connectivity links were significantly different between males and females, and with small effect sizes (mean Cohen d = 0.09, σ = 0.06, max d = 0.32). To assess the differences in the connectivity matrices among the different scanning institutions, we regressed out age, sex and diagnosis through ordinary least squares (OLS), thereafter applying a Lavene's test to each link to assess whether or not they had equal variances across institutions. There were 1498 links that had significantly different variances across institutions after FDR correction. As explained in methods, we ran an updated version of the Combat strategy to remove such sources of variability in the data and the confounding effects that may affect the subsequent subtyping of the ASD group, while retaining the between-group variability of our data. Except for our group variable, link connectivity values across institutions had very homogeneous and comparable distributions after data harmonization (figure S1), and the effect sizes decreased dramatically (figure S2), such that there was no association between the links and any of the confounding variables (institution, age, sex).

Subtyping

15

The harmonized data was used to find subtypes (or *clusters*) in the ASD group. We employed a consensus clustering approach to the brain connectivity matrices to gather together information about the connectivity patterns of all the brain regions at different resolutions. After applying a maximization algorithm to the modularity matrix obtained, three subtypes were found (figure S3): the first comprised of 121 subjects, with a prevalence of 13.75% with respect to the entire group; the second contained 278 subjects (31.59%); and the third subtype 464 subjects (52.72%). The rest of the subjects (17 in total) were considered to be *unclassified*, since the community detection algorithm could not assign them to any specific subtype. None of the resulting subtypes were differentiated by age (one-way ANOVA test, $p = 0.37$) or sex (χ^2 test, $p = 0.45$).

Behavioural assessment

We attempted to characterize each ASD subtype based on the different behavioural scores to clarify whether they had a behavioural signature. None of the tests nor their sub-parts significantly differentiated between the three (Table 1). However, and as expected due to the diagnosis of ASD, the three subtypes had significantly lower scores with respect to the TDC group (figure S4).

Connectivity class

Subsequently, we wondered if the type of connectivity could differentiate each subtype from the TDC and hence, we performed an OLS regression controlling for age, sex and the three measures of intelligence quotient (VIQ, FIQ and PIQ). First, we assessed the difference in the overall connectivity per subject defined as the average of the positive correlation coefficients extracted from the upper triangle in the harmonized connectivity matrices (figure 2a). Interestingly, the contrast between the TDC subjects (baseline) and the subtype 2 individuals was significantly positive ($\beta = 0.09$, $p_{bonf} < 0.01$), which means that there was a tendency for the overall connectivity to be greater in subtype 2 than in the TDC, indicative of overall *hyperconnectivity*. By

contrast, the opposite was true for subtype 3 (β = -0.05, $p_{bonf} < 0.01$), which implies that hypoconnectivity generally existed in this subtype. Finally, we did not find any significant difference between subtype 1 and the TDC group (β = 0.02, $p_{bonf} = 0.21$), indicating a more balanced or *mixed* situation in terms of overall connectivity. The use of the absolute correlation values of the connectivity matrix, or the median and the trimmed means as the overall connectivity metric, preserved all these findings of hypoconnectivity for subtype 3, hyperconnectivity for subtype 2 and a combination of the two classes for subtype 1.

(Figure 2 here)

Over and above the overall connectivity analyses, we asked if we could also find a connectivity class characteristic for each subtype at the link level. As such, we ran the same regression model for each individual link in the connectivity matrices as the response variable (figure S5) and compared this to the TDC group, most of the links gave higher values in subtype 2 individuals, indicative of link hyperconnectivity, and they were lower in subtype 3 patients, reflecting hypoconnectivity. In reference to subtype 1, the link difference distribution was more zero-centred, indicating a mix of hyper and hypoconnectivity at the link level similar to that observed when comparing the overall connectivity.

Graph topological metrics

Having seen that both the overall and the link-wise connectivity followed the same pattern of hyperconnectivity in subtype 2, hypoconnectivity in 3, and a mixture of the two in subtype 1, we asked whether these findings led to topological differences among the ASD subtypes and the TDC group. This is truly critical because two different scenarios could occur. In the first, the different subtypes would have similar topological properties and their connectivity values were simply scaled

by a constant, such that their connectivity footprints were exactly the same. In the second, the topology differs between the subtypes, providing a unique footprint specific to each subtype. In particular, we focused on four metrics that characterized the different global aspects of the topology of the connectivity matrices: Transitivity (*segregation*), global efficiency (*integration*), assortativity (*resilience*), and small-worldness (*communication*).

We found substantial topological differences in the connectivity matrices with a 10% link density threshold between the subtypes (figure 2b). Subtypes 1 and 3 had the highest and lowest values of assortativity and transitivity, respectively. Furthermore, the assortativity of subtype 1 was significantly higher than that of the TDC group ($\beta = 0.08$, $p_{bonf} < 0.01$), as was the transitivity ($\beta = 0.06$, $p_{bonf} < 0.01$), for which subtype 3 had lower values ($\beta = -0.01$, $p_{bonf} = 0.01$). Regarding global efficiency, subtype 3 (the most hypoconnected at the link level) had higher values with respect to each of the other ASD subtypes, and also with respect to the TDC group ($\beta = 0.01$, $p_{bonf} = 0.01$). When evaluating the small-world characteristics of these networks, we found that each ASD subtype and the TDC group exhibited such a property, since $\frac{\gamma}{\lambda} > 1$ held in all cases. Moreover, subtype 3 had the highest values, although it was only significantly greater than that of subtype 2 ($\beta = 0.37$, $p_{bonf} < 0.01$). These results remained qualitatively similar for different thresholds (figure S6).

Neurobiological signature

To determine the connectivity of structures that is most specific to each subtype, i.e.: their neurobiological signature, we applied a SVM multiclass classification with RFE as a feature selection procedure (see Methods). As a result, we obtained a total of 56 specific links from subtype 1, with an 87.39% balanced accuracy (BA) when classifying subtype 1 relative to the combination of subtypes 2 and 3 (figure S7). Similarly, we obtained 75 links for subtype 2 (BA=85.41%) and

18

444 links for subtype 3 (BA=92.55%). To qualitatively visualize the results of this classification, the selected links of each subtype were used to project our data onto a smaller dimension space through a t-distributed Stochastic Neighbour Embedding method (t-SNE hyper-parameters: perplexity=50, early exaggeration=10, learning rate=1000), producing a fairly clear separation of the different subtypes (figure 2c).

Once we determined the set of links that best classified each subtype, we set out to obtain their location in the brain in order to understand the neurobiological signature of each subtype (figure 2d). It should be noted that since the strategy adopted aimed to distinguish one subtype from the other two, the set of links was quite specific to each, even though some individual links could be shared between different subtypes. Moreover, since RFE performs feature removal based on the absolute value of the weights, links with large loadings when a subtype was the positive class could also be selected when the same subtype was part of the negative class, which would count as a false-positive link in the final set. Consequently, we finally only considered those links that were unique to each subtype and that also represented the positive class. Thus, following these considerations we finally obtained 43 specific links to subtype 1, 54 to subtype 2, and 412 to subtype 3.

Subsequently, to obtain brain maps of the links specific to each subtype and therefore, to define their neurobiological signature, we used the absolute loadings of the selected links per subtype to estimate the strength of all the regions of our brain partition, grouping left and right homologous sites of the same region into what we call *bilateralized* regions. Furthermore, if we consider that the most representative regions for each subtype are the two that have the greatest strength, we found that subtype 1 was represented by the caudal anterior connected cortex (ACC) and the precentral gyrus, subtype 2 by the rostral ACC and the paracentral gyrus, and subtype 3 by the frontal pole and the postcentral gyrus (figure 2d). Finally, for each of these regions, we showed their link specificity

19

projected onto a glass brain, reflecting the absolute weights between them and the rest of brain structures. For subtype 1, the largest weights were exhibited by the links between the caudal ACC and the pallidum, and that from the precentral gyrus and posterior cingulate cortex. For subtype 2, the rostral ACC connected largely with the insula, and the paracentral gyrus to the fusiform gyrus. Finally, the connection between the accumbens and the frontal pole, and that from the post-central gyrus to the caudal middle frontal gyrus were those that best represent subtype 3.

Transcriptomics

We first applied DDS to identify which genes in the entire transcriptome had a maximal association with the differences in connectivity between each ASD subtype and TDC, measured through the F-statistics resulting from MDMR. The kernel density distribution of F-values for each subtype was established (figure 2e), yet it is important to emphasize that MDMR allowed a F-value to be obtained for each region of our brain partition. Thus, we found that the connectivity patterns of some regions for one subtype were very different from those in the TDC group, while other regions retained very similar patterns of connectivity. As explained in the methods, the brain maps of the F-values were precisely those correlated with the expression maps of each of the genes (see figure 3c), which resulted in a total of 265 N genes and 384 P genes for subtype 1, 47 N genes and 468 P genes for subtype 2, and 28 N genes and 494 P genes for subtype 3 (figure 3a).

(figure 3 here)

Of all the N and P genes for each subtype, the genes that matched those in the list of hypothesis-driven genes were established (Table S4). For subtype 1, the most significant N genes were RAI1, RERE, ASH1L, FOXP1 and GIGYF (see also figure 3a), whereas there was only one P gene, HECTD4. For subtype 2, none of the genes obtained by DDS that were present in the hypothesis-driven list of genes were significant after surrogate-data testing. Finally, for subtype 3, there were

no significant N genes but a total of six P genes, namely MAP1A, PSMD12, SCN1A, RORB, MKX and HIVEP2. Therefore, none of the significant gene were common to multiple subtypes, yet subtype 1 was mainly driven by N genes, which means that they were systematically expressed more weakly in the brain regions with stronger F-statistics. By contrast, subtype 3 was mainly driven by P genes that were systematically expressed more strongly in association with higher F values.

To understand the neurobiological functions associated with each of the subtypes, we next applied a gene enrichment method to each subtype (figure 3b) using the sets of N and P genes separately (see Table S5 for the list of genes used for the enrichment in each case). In general, subtype 1 had poor enrichment, whereas subtypes 2 and 3 had a poor enrichment for P genes, although there were similarities between the two in relation to transmembrane potassium ion transport, neuron projection, membrane potential regulation and chemical synaptic transmission. Remarkably, the enrichment we found for the N genes of subtype 2 was very high, reaching more than 100-fold enrichment in some cases and corresponding to chylomicron clearance, phospholipid efflux, lipid regulation, and cholesterol esterification, transport and efflux. Thus, the enrichment found for subtype 2, which has a prevalence of 31.6% in our cohort, was strongly associated with phospholipid and cholesterol dynamics, events previously associated with ASD.

We obtained a separate enrichment for each subtype based on the genes in the negative tail (Z < 2), and for those in the positive tail (Z > 2). However, the expression map of some of the genes used for enrichment was poorly correlated with the F statistic, as evident for subtype 1 and the genes given in Table S4. To control for the effect of using genes with weak spatial correlation for enrichment, we repeated the same procedure but for the dual condition of $|Z| > 2$ and observed correlation of $|r| > 0.4$, where $||$ indicates the absolute value. In this way we only used genes with a strong spatial correlation for enrichment (figure S8). For subtype 1, no biological process was enriched

21

from either of the tails, neither the positive nor the negative one. For subtype 2, the enrichment of the positive tail (see figure 3) disappeared, whereas that of the negative tail related to processes involving cholesterol and other lipoproteins was preserved, highlighting the robustness of these results. Finally, for subtype 3, the enrichment of the positive tail (see figure 3) was also preserved.

Enrichment without subtyping

To prove the need to perform subtyping due to the strong enrichment of cholesterol and other lipoproteins in subtype 2, we repeated the same procedure but for the maps of the F-statistic in the entire ASD group (Figure S9). As a result, the association with cholesterol and its derived processes disappeared, and we only found a moderate enrichment for the positive tail of $Z > 2$ with biological processes like axo-dendritic transport, regulation of postsynaptic membrane potential, synaptic transcription, and neuron projection development, to some extent related to the enrichment obtained from the positive tails in subtypes 2 and 3. However, when we compared the list of significant genes that survived surrogate-data testing with those of each subtype (Table S6), we found different genes except for the MAP1A gene, which existed for both subtype 3 and the list of all ASD patients without subtyping (see Table S7 for the list used for this enrichment).

Discussion

The strong heterogeneity in the phenotypes of psychiatric conditions represents a major challenge to the design of protective and restorative therapies, minimizing their impact on symptom severity or their behavioral outcomes in many patients. This occurs in ASD, a condition that is the result of complex interactions between genetic, epigenetic and environmental factors during development. Hundreds of different genes have been associated with ASD and as of July 15th 2020, a total of 944 genes appeared on the SFARI list, 192 of which were attributed a score of 1, maximum evidence of a pathological relationship with ASD. This high genetic complexity is yet another manifestation of the heterogeneity in this condition.

Many studies have set out to assess the heterogeneity in ASD and to better stratify ASD subgroups (dubbed subtypes here), some of which have also made use of neuroimaging as we have here (*68, 69*), and see also (*70*) and references therein). Nevertheless, our approach is unique in several ways. Firstly, our study is based on a large cohort of patients with ASD (N = 880), obtained from the ABIDE initiative, and it combines anatomical and functional neuroimaging data from 24 different institutions. Second, we have used Combat to conduct functional connectivity studies, a highly rigorous method to eliminate the variability between MRI scans across the 24 institutions, one of the largest sources of variability when combining data from multiple institutions (*71*). Third, our analysis of brain connectivity was carried out on a large-scale, where each brain region is represented by its connectivity pattern, a vector that accounts for the connectivity from that region to all the others and that thereby takes into account functional interactions across the entire brain. Therefore, we do not consider any brain region as more dominant or relevant over the others *a priori*. Fourth, we made use of a consensus clustering approach we have developed (*35, 36*), and that has been successfully tested by others (*72*), to establish brain connectivity matrices and to group two subjects in the same subtype if the vectors of connectivity are similar across all the regions analyzed, which in our case involved a total of 86 regions per subject: 68 cortical regions and 18 subcortical ones. Finally, we made use of the AHBA to describe the neurogenetic profiles of each subtype, defined as a vector of the F-statistic, a multivariate distance obtained for each brain region by applying the MDMR technique between each subtype and the TDC group, consisting here of 1010 subjects.

Our main result is that we found an enrichment associated with biological processes related to cholesterol and other lipoproteins for one of the ASD subtypes, with an enrichment above 100-fold in some cases. To the best of our knowledge, such a high enrichment has yet to be reported in any previous studies using AHBA, neither in studies assessing myelination (*73*), hierarchical cortical

organization (*74*), visuomotor integration (*47*), nor in studies of large-scale connectivity (*45*, *75*, *76*). It is exceptionally important to highlight that if enrichment is carried out on the entire autistic population, without previously performing any subtyping, this association with cholesterol is masked, such that the process of subtyping is really critical to find such an enrichment.

Hypocholesterolemia is a phenomenon associated with ASD (*14–16*) and several studies have shown that modifications to cholesterol-derived metabolites like estradiol, dehydroepiandrosterone and glucocorticoids can affect brain connectivity and synaptic plasticity (*77*). In terms of therapy, estradiol treatment of a mouse model of ASD increased protein levels and consequently, it ameliorated ASD-associated social behavioral deficits and decreased repetitive behavior in later life (*18*). In another study, estrogen increased parvalbumin levels to attenuate ASD (*78*) and elsewhere, a decrease in the dietary uptake of cholesterol increased the incidence of ASD (*79*). Our results suggest that cholesterol-based therapies may not be beneficial for all cases of ASD but rather, to some of them, which according to our results would correspond to 31.6% of all ASD individuals.

Accordingly, it will be important to define the mechanism by which the connectivity and expression maps in terms of cholesterol related genes are tightly linked for subtype 2, and yet this association is weaker in subtypes 1 and 3. Although more studies will be required to address this issue in detail, we found a similar pattern of association for the cholesterol-related genes responsible for this enrichment, namely ABCG1, APOE, APOC1 and APOC2, differentiating subtype 2 from the other two subtypes. While these genes are generally expressed more subcortically than cortically in the three subtypes, the cortical F-statistic indicates they are specifically enhanced in areas such as the precuneus or retrosplenial nucleus in subtype 2. This distribution may reflect a compensatory strategy for network dysfunction, which could suggest that ASD in subtype 2 could be more functionally driven than that of the other two subtypes.

24

Moreover, in relation to the connectivity class, we found that in subtype 2 the class associated with the synthesis and dynamics of cholesterol and other lipoproteins dominates hyperconnectivity. Alternatively, when we look at the global topological properties of brain connectivity, such as transitivity (segregation), global efficiency (integration), assortativity (resilience) and small-worldness (communication), we find that subtype 2 is precisely that with the smallest differences from the TDC group. Hence, subtype 2 would appear to maintain a similar organization at the graph level to the TDC group. Perhaps, without affecting the network topology, these alterations in functional connectivity could indicate a more transient and purely functional state of this subtype, compatible with hyperconnectivity acting as a compensatory strategy for network dysfunction.

In the other two subtypes, hypoconnectivity was the class associated with subtype 3 and a combination or mixture of hypo- and hyperconnectivity for subtype 1. However, the changes in the network topology we see between subtypes 1 and 3 are quite contrasting. While for subtype 1 there is greater segregation and less efficiency than in the TDC group, for subtype 3 the opposite tendency is seen, with less segregation and greater efficiency, similar to the earlier observations (*80*, *81*). However, the characteristics of subtype 1 have previously also been shown to arise in ASD (*82*) and indeed, the variations in the topological metrics for subtypes 1 and 3 have already been reported elsewhere (*83*, *84*), yet not those of subtype 2. Considering all these results together, there appears to be an important heterogeneity in the types of connectivity in ASD (hyper, hypo or mixed), as well as in the different metrics that define the topology of the brain connectivity network, as concluded previously (*83*, *84*).

Regarding the neurobiological signature of the ASD subtypes, our methodology allowed us to find certain structures whose connectivity profiles are quite specific to each subtype. In particular, the neurobiological signature for subtype 1 is given by the connectivity of the caudal-ACC to the rest of the brain and that of the precentral gyrus to the rest of the brain. For subtype 2, it is given by the

connectivity of the ACC rostral and that of the paracentral lobule, and for subtype 3 by the connections of the rest of the brain to the frontal pole and the post-central gyrus. Therefore, as elsewhere (*25*), our work proves that the ACC division can explain the strong heterogeneity of the ASD subtypes, in our case differentiating two of the three subtypes. Moreover, we find a smooth transition from the precentral gyrus in subtype 1, with a primary motor function, to the paracentral lobule in subtype 2, with a motor-sensory function, until the postcentral gyrus found in subtype 3, with a primary somato-sensory function. These transitions could in part explain the heterogeneity in the neuroanatomy underlying ASD.

It is also important to remember that our subtypes are invisible, or statistically indistinguishable from the different scores obtained by the behavioral and neuropsychological tests, these tests defining the diagnosis of ASD. Therefore, if we had subtyped according to neuropsychological scores we would have obtained different subtypes, perhaps based on the severity of their symptoms or some other neuropsychological phenotype, for example, having a higher or lower verbal IQ score, or worse socialization. Instead, although the behavioral phenotypes are indistinguishable, we show how connectivity patterns are specific to each subtype and thus, constitute a neurobiological signature imprinted on large-scale functional connectivity. Furthermore, when we look at the associations with the transcriptome, we can also differentiate between subtypes.

Another original aspect of our work is the characterization of each subtype using AHBA, which although it has been used before to characterize morphometric information in ASD (*43*), it has never been used to characterize subtypes based on functional patterns of connectivity. Due to the large heterogeneity in the genetic data associated with this condition, the use of AHBA may shed new light on this aspect of ASD, as it provides information on the transcriptome across the brain in unprecedented detail. Indeed, the use of AHBA allows about 3,702 sampling sites to be covered and activity patterns to be built for about 20,500 genes as a specific signature for each anatomical

region. We have intersected brain maps of gene expression with those corresponding to the F-statistic, which provides a multivariate distance of the connectivity between each subtype and the control group for each brain region. This approach, using AHBA and accounting for the interactions with the entire transcriptome, is also complementary to other techniques, such as genome-wide association studies (GWAS: (*85*) that simultaneously address genotype–phenotype associations from hundreds of thousands to millions of genetic variants in a data-driven manner. Indeed, GWAS has previously been used for ASD subtyping (*86, 87*), yet using behavioral scores as traits and therefore, the subtypes obtained were more closely related to symptom severity and not to functional connectivity.

We identified a different transcription signature for each subtype described here. In subtype 1, with a prevalence of 13.75% among the total ASD cases, we found a negative association with the RAI1, RERE, ASH1L, FOXP1 and GIGFY1 genes, and positive association with the HECTD4 gene. Significantly, some variants of RAI1 have been associated with Potocki-Lupski (*88*) and Smith-Magenis syndromes (*89*), some rare mutations in the FOXP1 gene with intellectual disability and mental retardation (*90*), and some mutations in ASH1L with intellectual disability (*91*). Thus, subtype 1 may perhaps have a stronger overlap with other syndromes or other neurodevelopmental disabilities. Alternatively, when we look at the spatial correlation of each gene, these values are quite low, which may be due to the fact that ABIDE includes individuals with average or above-average IQ (*37*) and thus, this subtype could be under-represented.

In relation to subtype 3, we found a significant association with the MAP1A, PSMD12, SCN1A, RORB, MKK and HIVEP2 genes, with enrichment related to potassium ion transmembrane transport, neuron projection, membrane potential regulation and chemical synaptic transmission, which closely agrees with the enrichment previously reported in ASD (*92*). Moreover, such enrichment was also somehow shared with the positive association in subtype 2. It is interesting to

note the MAP1A gene (with an observed correlation of 0.63) was also present in the enrichment obtained in absence of subtyping, somehow suggesting its dominance in the association with brain connectivity profiles in the autistic brain, and also consistent with its presence in the most prevalent of the three subtypes (52.72%). Moreover, the MAP1A gene has also been implicated in the genetics of schizophrenia (*93*), offering a link between these two conditions.

In summary, our novel approach, which includes data harmonization, multivariate distancing in large scale functional connectivity patterns and transcriptome brain maps, reveals an extremely strong enrichment in ASD for the synthesis and transport of cholesterol and other lipoproteins in approximately 31.6% of autistic patients, suggesting a route for new potential therapeutic strategies.

Acknowledgments

References

1. A. Masi, M. M. DeMayo, N. Glozier, A. J. Guastella, An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. *Neurosci. Bull.* **33**, 183–193 (2017).

2. M.-C. Lai, M. V. Lombardo, S. Baron-Cohen, Autism. *Lancet Lond. Engl.* **383**, 896–910 (2014).

3. H. Hodges, C. Fealko, N. Soares, Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Transl. Pediatr.* **9**, S55–S65 (2020).

4. M. H. Mehling, M. J. Tassé, Severity of Autism Spectrum Disorders: Current Conceptualization, and Transition to DSM-5. *J. Autism Dev. Disord.* **46**, 2000–2016 (2016).

5. C. Lord, M. Elsabbagh, G. Baird, J. Veenstra-Vanderweele, Autism spectrum disorder. *Lancet Lond. Engl.* **392**, 508–520 (2018).

6. V. Eapen, Genetic basis of autism: is there a way forward?: *Curr. Opin. Psychiatry*. **24**, 226–236 (2011).

7. R. Bhandari, J. K. Paliwal, A. Kuhad, in *Personalized Food Intervention and Therapy for Autism Spectrum Disorder Management*, M. M. Essa, M. W. Qoronfleh, Eds. (Springer International Publishing, Cham, 2020; http://link.springer.com/10.1007/978-3-030-30402-7_4), vol. 24 of *Advances in Neurobiology*, pp. 97–141.

8. I. Hertz-Picciotto, R. J. Schmidt, P. Krakowiak, Understanding environmental contributions to autism: Causal concepts and the state of science: Environmental contributions to ASD. *Autism Res.* **11**, 554–586 (2018).

9. T. Momoi, E. Fujita, H. Senoo, M. Momoi, Genetic factors and epigenetic factors for autism: endoplasmic reticulum stress and impaired synaptic function. *Cell Biol. Int.* **34**, 13–19 (2009).

10. J. L. R. Rubenstein, M. M. Merzenich, Model of autism: increased ratio of excitation/inhibition in key neural systems: Model of autism. *Genes Brain Behav.* **2**, 255–267 (2003).

11. S. B. Nelson, V. Valakh, Excitatory/Inhibitory Balance and Circuit Homeostasis in Autism Spectrum Disorders. *Neuron.* **87**, 684–698 (2015).

12. V. S. Sohal, J. L. R. Rubenstein, Excitation-inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatry*. **24**, 1248–1257 (2019).

13. E. Tierney, I. Bukelis, R. E. Thompson, K. Ahmed, A. Aneja, L. Kratz, R. I. Kelley, Abnormalities of cholesterol metabolism in autism spectrum disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **141B**, 666–668 (2006).

14. Z. Hu, Y. Yang, Y. Zhao, H. Yu, X. Ying, D. Zhou, J. Zhong, Z. Zheng, J. Liu, R. Pan, W. Zhang, F. Cheng, S. Duan, APOE hypermethylation is associated with autism spectrum disorder in a Chinese population. *Exp. Ther. Med.* **15**, 4749–4754 (2018).

15. C. Gillberg, E. Fernell, E. Kočovská, H. Minnis, T. Bourgeron, L. Thompson, C. S. Allely, The role of cholesterol metabolism and various steroid abnormalities in autism spectrum disorders: A hypothesis paper: A hypothesis paper. *Autism Res.* **10**, 1022–1044 (2017).

16. A. El-Ansary, S. Chirumbolo, R. S. Bhat, M. Dadar, E. M. Ibrahim, G. Bjørklund, The Role of Lipidomics in Autism Spectrum Disorder. *Mol. Diagn. Ther.* **24**, 31–48 (2020).

17. Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), "Cholesterol in Autism Spectrum Disorder (ASD): Characterization and Treatment" (Clinical trial registration study/NCT00965068, clinicaltrials.gov, 2018), (available at https://clinicaltrials.gov/ct2/show/study/NCT00965068).

18. F. Filice, E. Lauber, K. J. Vörckel, M. Wöhr, B. Schwaller, 17-β estradiol increases parvalbumin levels in Pvalb heterozygous mice and attenuates behavioral phenotypes with relevance to autism core symptoms. *Mol. Autism.* **9**, 15 (2018).

19. V. Luine, M. Frankfurt, Interactions between estradiol, BDNF and dendritic spines in promoting memory. *Neuroscience.* **239**, 34–45 (2013).

20. Y. Lu, G. R. Sareddy, J. Wang, R. Wang, Y. Li, Y. Dong, Q. Zhang, J. Liu, J. C. O'Connor, J. Xu, R. K. Vadlamudi, D. W. Brann, Neuron-Derived Estrogen Regulates Synaptic Plasticity and Memory. *J. Neurosci. Off. J. Soc. Neurosci.* **39**, 2792–2809 (2019).

21. D. P. Kennedy, R. Adolphs, The social brain in psychiatric and neurological disorders. *Trends Cogn. Sci.* **16**, 559–572 (2012).

22. A. Thompson, D. Murphy, F. Dell'Acqua, C. Ecker, G. McAlonan, H. Howells, S. Baron-Cohen, M.-C. Lai, M. V. Lombardo, MRC AIMS Consortium, and Marco Catani, Impaired Communication Between the Motor and Somatosensory Homunculus Is Associated With Poor Manual Dexterity in Autism Spectrum Disorder. *Biol. Psychiatry.* **81**, 211–219 (2017).

23. M. A. Patriquin, T. DeRamus, L. E. Libero, A. Laird, R. K. Kana, Neuroanatomical and neurofunctional markers of social cognition in autism spectrum disorder. *Hum. Brain Mapp.* **37**, 3957–3978 (2016).

24. M. B. Nebel, S. E. Joel, J. Muschelli, A. D. Barber, B. S. Caffo, J. J. Pekar, S. H. Mostofsky, Disruption of functional organization within the primary motor cortex in children with autism. *Hum. Brain Mapp.* **35**, 567–580 (2014).

25. M. L. Simms, T. L. Kemper, C. M. Timbie, M. L. Bauman, G. J. Blatt, The anterior cingulate cortex in autism: heterogeneity of qualitative and quantitative cytoarchitectonic features suggests possible subgroups. *Acta Neuropathol. (Berl.).* **118**, 673–684 (2009).

26. M. Catani, F. Dell'Acqua, S. Budisavljevic, H. Howells, M. Thiebaut de Schotten, S. Froudist-Walsh, L. D'Anna, A. Thompson, S. Sandrone, E. T. Bullmore, J. Suckling, S. Baron-Cohen, M. V. Lombardo, S. J. Wheelwright, B. Chakrabarti, M.-C. Lai, A. N. V. Ruigrok, A. Leemans, C. Ecker, M. A. Consortium, M. C. Craig, D. G. M. Murphy, Frontal networks in adults with autism spectrum disorder. *Brain J. Neurol.* **139**, 616–630 (2016).

27. L. Q. Uddin, K. Supekar, C. J. Lynch, A. Khouzam, J. Phillips, C. Feinstein, S. Ryali, V. Menon, Salience Network–Based Classification and Prediction of Symptom Severity in Children With Autism. *JAMA Psychiatry.* **70**, 869–879 (2013).

28. A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, F. Meneguzzi, Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin.* **17**, 16–23 (2018).

29. J. S. Anderson, J. A. Nielsen, A. L. Froehlich, M. B. DuBray, T. J. Druzgal, A. N. Cariello, J. R. Cooperrider, B. A. Zielinski, C. Ravichandran, P. T. Fletcher, A. L. Alexander, E. D. Bigler, N. Lange, J. E. Lainhart, Functional connectivity magnetic resonance imaging classification of autism. *Brain J. Neurol.* **134**, 3742–3754 (2011).

30. E. P. K. Pua, G. Ball, C. Adamson, S. Bowden, M. L. Seal, Quantifying individual differences in brain morphometry underlying symptom severity in Autism Spectrum Disorders. *Sci. Rep.* **9**, 9898 (2019).

31. J. S. Nomi, L. Q. Uddin, Developmental changes in large-scale network connectivity in autism. *NeuroImage Clin.* **7**, 732–741 (2015).

32. E. Courchesne, R. Carper, N. Akshoomoff, Evidence of brain overgrowth in the first year of life in autism. *JAMA*. **290**, 337–344 (2003).

33. C. He, H. Chen, L. Q. Uddin, A. Erramuzpe, P. Bonifazi, X. Guo, J. Xiao, H. Chen, X. Huang, L. Li, W. Sheng, W. Liao, J. M. Cortes, X. Duan, Structure-Function Connectomics Reveals Aberrant Developmental Trajectory Occurring at Preadolescence in the Autistic Brain. *Cereb. Cortex N. Y. N 1991*. **30**, 5028–5037 (2020).

34. C. Fountain, A. S. Winter, P. S. Bearman, Six Developmental Trajectories Characterize Children With Autism. *Pediatrics*. **129**, e1112–e1120 (2012).

35. J. Rasero, M. Pellicoro, L. Angelini, J. M. Cortes, D. Marinazzo, S. Stramaglia, Consensus clustering approach to group brain connectivity matrices. *Netw. Neurosci.* **1**, 242–253 (2017).

36. J. Rasero, I. Diez, J. M. Cortes, D. Marinazzo, S. Stramaglia, Connectome sorting by consensus clustering increases separability in group neuroimaging studies. *Netw. Neurosci.* **3**, 325–343 (2019).

37. A. Di Martino, C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto, B. Deen, S. Delmonte, I. Dinstein, B. Ertl-Wagner, D. A. Fair, L. Gallagher, D. P. Kennedy, C. L. Keown, C. Keysers, J. E. Lainhart, C. Lord, B. Luna, V. Menon, N. Minshew, C. S. Monk, S. Mueller, R.-A. Müller, M. B. Nebel, J. T. Nigg, K. O'Hearn, K. A. Pelphrey, S. J. Peltier, J. D. Rudie, S. Sunaert, M. Thioux, J. M. Tyszka, L. Q. Uddin, J. S. Verhoeven, N. Wenderoth, J. L. Wiggins, S. H. Mostofsky, M. P. Milham, The Autism Brain Imaging Data Exchange: Towards Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. *Mol. Psychiatry*. **19**, 659–667 (2014).

38. J.-P. Fortin, N. Cullen, Y. I. Sheline, W. D. Taylor, I. Aselcioglu, P. A. Cook, P. Adams, C. Cooper, M. Fava, P. J. McGrath, M. McInnis, M. L. Phillips, M. H. Trivedi, M. M. Weissman, R. T. Shinohara, Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*. **167**, 104–120 (2018).

39. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. **8**, 118–127 (2007).

40. C. Wachinger, A. Rieckmann, S. Pölsterl, Detect and Correct Bias in Multi-Site Neuroimaging Datasets. *ArXiv200205049 Cs Eess* (2020) (available at http://arxiv.org/abs/2002.05049).

41. M. Yu, K. A. Linn, P. A. Cook, M. L. Phillips, M. McInnis, M. Fava, M. H. Trivedi, M. M. Weissman, R. T. Shinohara, Y. I. Sheline, Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum. Brain Mapp.* **39**, 4213–4227 (2018).

42. Human Gene Module. *SFARI Gene*, (available at https://gene.sfari.org/database/human-gene/).

43. R. Romero-Garcia, V. Warrier, E. T. Bullmore, S. Baron-Cohen, R. A. I. Bethlehem, Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Mol. Psychiatry*. **24**, 1053–1064 (2019).

44. M. J. Hawrylycz, E. S. Lein, A. L. Guillozet-Bongaarts, E. H. Shen, L. Ng, J. A. Miller, L. N. van de Lagemaat, K. A. Smith, A. Ebbert, Z. L. Riley, C. Abajian, C. F. Beckmann, A. Bernard, D. Bertagnolli, A. F. Boe, P. M. Cartagena, M. M. Chakravarty, M. Chapin, J. Chong, R. A. Dalley, B. D. Daly, C. Dang, S. Datta, N. Dee, T. A. Dolbeare, V. Faber, D. Feng, D. R. Fowler, J. Goldy, B. W. Gregor, Z. Haradon, D. R. Haynor, J. G. Hohmann, S. Horvath, R. E. Howard, A. Jeromin, J. M. Jochim, M. Kinnunen, C. Lau, E. T. Lazarz, C. Lee, T. A. Lemon, L. Li, Y. Li, J. A. Morris, C. C. Overly, P. D. Parker, S. E. Parry, M. Reding, J. J. Royall, J. Schulkin, P. A. Sequeira, C. R. Slaughterbeck, S. C. Smith, A. J. Sodt, S. M. Sunkin, B. E. Swanson, M. P. Vawter, D. Williams, P. Wohnoutka, H. R. Zielke, D. H. Geschwind, P. R. Hof, S. M. Smith, C. Koch, S. G. N. Grant, A. R. Jones, An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*. **489**, 391–399 (2012).

45. I. Diez, J. Sepulcre, Neurogenetic profiles delineate large-scale connectivity dynamics of the human brain. *Nat. Commun.* **9**, 3876 (2018).

46. J. Sepulcre, M. J. Grothe, F. d'Oleire Uquillas, L. Ortiz-Terán, I. Diez, H.-S. Yang, H. I. L. Jacobs, B. J. Hanseeuw, Q. Li, G. El-Fakhri, R. A. Sperling, K. A. Johnson, Neurogenetic contributions to amyloid beta and tau spreading in the human cortex. *Nat. Med.* **24**, 1910–1918 (2018).

47. E. Bueichekú, M. Aznárez-Sanado, I. Diez, F. d'Oleire Uquillas, L. Ortiz-Terán, A. Y. Qureshi, M. Suñol, S. Basaia, E. Ortiz-Terán, M. A. Pastor, J. Sepulcre, Central neurogenetic signatures of the visuomotor integration system. *Proc. Natl. Acad. Sci.* **117**, 6836–6843 (2020).

48. R. W. Cox, AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res. Int. J.* **29**, 162–173 (1996).

49. L. Kaufman, P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis* (2009).

50. J. Reichardt, S. Bornholdt, Statistical mechanics of community detection. *Phys Rev E*. **74**, 016110 (2006).

51. Z. Shehzad, C. Kelly, P. T. Reiss, R. C. Craddock, J. W. Emerson, K. McMahon, D. A. Copland, F. X. Castellanos, M. P. Milham, A multivariate distance-based analytic framework for connectome-wide association studies. *NeuroImage*. **93**, 74–94 (2014).

52. M. A. Zapala, N. J. Schork, Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proc. Natl. Acad. Sci.* **103**, 19430–19435 (2006).

53. B. H. McArdle, M. J. Anderson, FITTING MULTIVARIATE MODELS TO COMMUNITY DATA: A COMMENT ON DISTANCE-BASED REDUNDANCY ANALYSIS. *Ecology*. **82**, 290–297 (2001).

54. M. Rubinov, O. Sporns, Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage*. **52**, 1059–1069 (2010).

55. D. J. Watts, S. H. Strogatz, Collective dynamics of `small-world' networks. *Nature*. **393**, 440–442 (1998).

56. D. S. Bassett, E. T. Bullmore, Small-World Brain Networks Revisited. *The Neuroscientist*. **23**, 499–516 (2017).

57. M. D. Humphries, K. Gurney, Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence. *PLOS ONE*. **3**, 1–10 (2008).

58. S. Achard, E. Bullmore, Efficiency and Cost of Economical Brain Functional Networks. *PLOS Comput. Biol.* **3**, 1–10 (2007).

59. I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **46**, 389–422 (2002).

60. T. Hastie, *The elements of statistical learning : data mining, inference, and prediction* (Springer, New York, 2009).

61. A. Arnatkevic Iūtė, B. D. Fulcher, A. Fornito, A practical guide to linking brain-wide gene expression and neuroimaging data. *NeuroImage*. **189**, 353–367 (2019).

62. J. Arloth, D. M. Bader, S. Röh, A. Altmann, Re-Annotator: Annotation Pipeline for Microarray Probe Sequences. *PLOS ONE*. **10**, e0139516 (2015).

63. M. Hawrylycz, J. A. Miller, V. Menon, D. Feng, T. Dolbeare, A. L. Guillozet-Bongaarts, A. G. Jegga, B. J. Aronow, C.-K. Lee, A. Bernard, M. F. Glasser, D. L. Dierker, J. Menche, A. Szafer, F. Collman, P. Grange, K. A. Berman, S. Mihalas, Z. Yao, L. Stewart, A.-L. Barabási, J. Schulkin, J. Phillips, L. Ng, C. Dang, D. R. Haynor, A. Jones, D. C. Van Essen, C. Koch, E. Lein, Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015).

64. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

65. A. Fabregat, K. Sidiropoulos, G. Viteri, O. Forner, P. Marin-Garcia, V. Arnau, P. D'Eustachio, L. Stein, H. Hermjakob, Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics*. **18**, 142 (2017).

66. F. K. Satterstrom, J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J.-Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, Autism Sequencing Consortium, iPSYCH-Broad Consortium, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. **180**, 568-584.e23 (2020).

67. J. B. Burt, M. Helmer, M. Shinn, A. Anticevic, J. D. Murray, Generative modeling of brain maps with spatial autocorrelation. *NeuroImage*. **220**, 117038 (2020).

68. A. K. Easson, Z. Fatima, A. R. McIntosh, Functional connectivity-based subtypes of individuals with and without autism spectrum disorder. *Netw. Neurosci.* **3**, 344–362 (2019).

69. S. G. Urchs, A. Tam, P. Orban, C. Moreau, Y. Benhajali, H. D. Nguyen, A. C. Evans, P. Bellec, "Subtypes of functional connectivity associate robustly with ASD diagnosis" (preprint, Neuroscience, 2020), , doi:10.1101/2020.04.14.040576.

70. C. Ecker, D. Murphy, Neuroimaging in autism—from basic science to translational research. *Nat. Rev. Neurol.* **10**, 82–91 (2014).

71. Y. He, L. Byrge, D. P. Kennedy, Nonreplication of functional connectivity differences in autism spectrum disorder across multiple sites and denoising strategies. *Hum. Brain Mapp.* **41**, 1334–1350 (2020).

72. A. Kabbara, M. Khalil, G. O'Neill, K. Dujardin, Y. El Traboulsi, F. Wendling, M. Hassan, Detecting modular brain states in rest and task. *Netw. Neurosci.* **3**, 878–901 (2019).

73. J. Ritchie, S. P. Pantazatos, L. French, Transcriptomic characterization of MRI contrast with focus on the T1-w/T2-w ratio in the cerebral cortex. *NeuroImage*. **174**, 504–517 (2018).

74. J. B. Burt, M. Demirtaş, W. J. Eckner, N. M. Navejar, J. L. Ji, W. J. Martin, A. Bernacchia, A. Anticevic, J. D. Murray, Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).

75. A. Fornito, A. Arnatkevičiūtė, B. D. Fulcher, Bridging the Gap between Connectome and Transcriptome. *Trends Cogn. Sci.* **23**, 34–50 (2019).

76. R. Romero-Garcia, K. J. Whitaker, F. Váša, J. Seidlitz, M. Shinn, P. Fonagy, R. J. Dolan, P. B. Jones, I. M. Goodyer, E. T. Bullmore, P. E. Vértes, Structural covariance networks are coupled to expression of genes enriched in supragranular layers of the human cortex. *NeuroImage*. **171**, 256–267 (2018).

77. A. G. Woods, I. Sokolowska, R. Taurines, M. Gerlach, E. Dudley, J. Thome, C. C. Darie, Potential biomarkers in psychiatry: focus on the cholesterol system. *J. Cell. Mol. Med.* **16**, 1184–1195 (2012).

78. A. Crider, A. Pillai, Estrogen Signaling as a Therapeutic Target in Neurodevelopmental Disorders. *J. Pharmacol. Exp. Ther.* **360**, 48–58 (2017).

79. K. van Elst, H. Bruining, B. Birtoli, C. Terreaux, J. K. Buitelaar, M. J. Kas, Food for thought: dietary changes in essential fatty acid ratios and the increase in autism spectrum disorders. *Neurosci. Biobehav. Rev.* **45**, 369–378 (2014).

80. J. D. Rudie, J. A. Brown, D. Beck-Pancer, L. M. Hernandez, E. L. Dennis, P. M. Thompson, S. Y. Bookheimer, M. Dapretto, Altered functional and structural brain network organization in autism. *NeuroImage Clin.* **2**, 79–94 (2013).

81. T. Itahashi, T. Yamada, H. Watanabe, M. Nakamura, D. Jimbo, S. Shioda, K. Toriizuka, N. Kato, R. Hashimoto, Altered Network Topologies and Hub Organization in Adults with Autism: A Resting-State fMRI Study. *PLoS ONE*. **9**, e94115 (2014).

82. K. Alaerts, F. Geerlings, L. Herremans, S. P. Swinnen, J. Verhoeven, S. Sunaert, N. Wenderoth, Functional Organization of the Action Observation Network in Autism: A Graph Theory Approach. *PLOS ONE*. **10**, e0137020 (2015).

83. J. O. Maximo, R. K. Kana, Aberrant "deep connectivity" in autism: A cortico-subcortical functional connectivity magnetic resonance imaging study: Cortico-subcortical connectivity in ASD. *Autism Res.* **12**, 384–400 (2019).

84. B. E. Yerys, J. D. Herrington, T. D. Satterthwaite, L. Guy, R. T. Schultz, D. S. Bassett, Globally weaker and topologically different: resting-state connectivity in youth with autism. *Mol. Autism.* **8**, 39 (2017).

85. V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, D. Meyre, Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019).

86. M. Asif, H. F. M. C. Martiniano, A. R. Marques, J. X. Santos, J. Vilela, C. Rasga, G. Oliveira, F. M. Couto, A. M. Vicente, Identification of biological mechanisms underlying a multidimensional ASD phenotype using machine learning. *Transl. Psychiatry*. **10**, 1–12 (2020).

87. A. Yousaf, R. Waltes, D. Haslinger, S. M. Klauck, E. Duketis, M. Sachse, A. Voran, M. Biscaldi, M. Schulte-Rüther, S. Cichon, M. Nöthen, J. Ackermann, I. Koch, C. M. Freitag, A. G. Chiocchetti, Quantitative genome-wide association study of six phenotypic subdomains identifies novel genome-wide significant variants in autism spectrum disorder. *Transl. Psychiatry*. **10**, 215 (2020).

88. L. Potocki, J. Neira-Fresneda, B. Yuan, in *GeneReviews®*, M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. Bean, K. Stephens, A. Amemiya, Eds. (University of Washington, Seattle, Seattle (WA), 1993; http://www.ncbi.nlm.nih.gov/books/NBK447920/).

89. M. Falco, S. Amabile, F. Acquaviva, RAI1 gene mutations: mechanisms of Smith-Magenis syndrome. *Appl. Clin. Genet.* **10**, 85–94 (2017).

90. F. F. Hamdan, H. Daoud, D. Rochefort, A. Piton, J. Gauthier, M. Langlois, G. Foomani, S. Dobrzeniecka, M.-O. Krebs, R. Joober, R. G. Lafrenière, J.-C. Lacaille, L. Mottron, P. Drapeau, M. H. Beauchamp, M. S. Phillips, E. Fombonne, G. A. Rouleau, J. L. Michaud, De

novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671–678 (2010).

91. H. A. F. Stessman, B. Xiong, B. P. Coe, T. Wang, K. Hoekzema, M. Fenckova, M. Kvarnung, J. Gerdts, S. Trinh, N. Cosemans, L. Vives, J. Lin, T. N. Turner, G. Santen, C. Ruivenkamp, M. Kriek, A. van Haeringen, E. Aten, K. Friend, J. Liebelt, C. Barnett, E. Haan, M. Shaw, J. Gecz, B.-M. Anderlid, A. Nordgren, A. Lindstrand, C. Schwartz, R. F. Kooy, G. Vandeweyer, C. Helsmoortel, C. Romano, A. Alberti, M. Vinci, E. Avola, S. Giusto, E. Courchesne, T. Pramparo, K. Pierce, S. Nalabolu, D. G. Amaral, I. E. Scheffer, M. B. Delatycki, P. J. Lockhart, F. Hormozdiari, B. Harich, A. Castells-Nobau, K. Xia, H. Peeters, M. Nordenskjöld, A. Schenck, R. A. Bernier, E. E. Eichler, Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).

92. T. Bourgeron, From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).

93. R. A. Myers, F. Casals, J. Gauthier, F. F. Hamdan, J. Keebler, A. R. Boyko, C. D. Bustamante, A. M. Piton, D. Spiegelman, E. Henrion, M. Zilversmit, J. Hussin, J. Quinlan, Y. Yang, R. G. Lafrenière, A. R. Griffing, E. A. Stone, G. A. Rouleau, P. Awadalla, A Population Genetic Approach to Mapping Neurological Disorder Genes Using Deep Resequencing. *PLOS Genet.* **7**, e1001318 (2011).
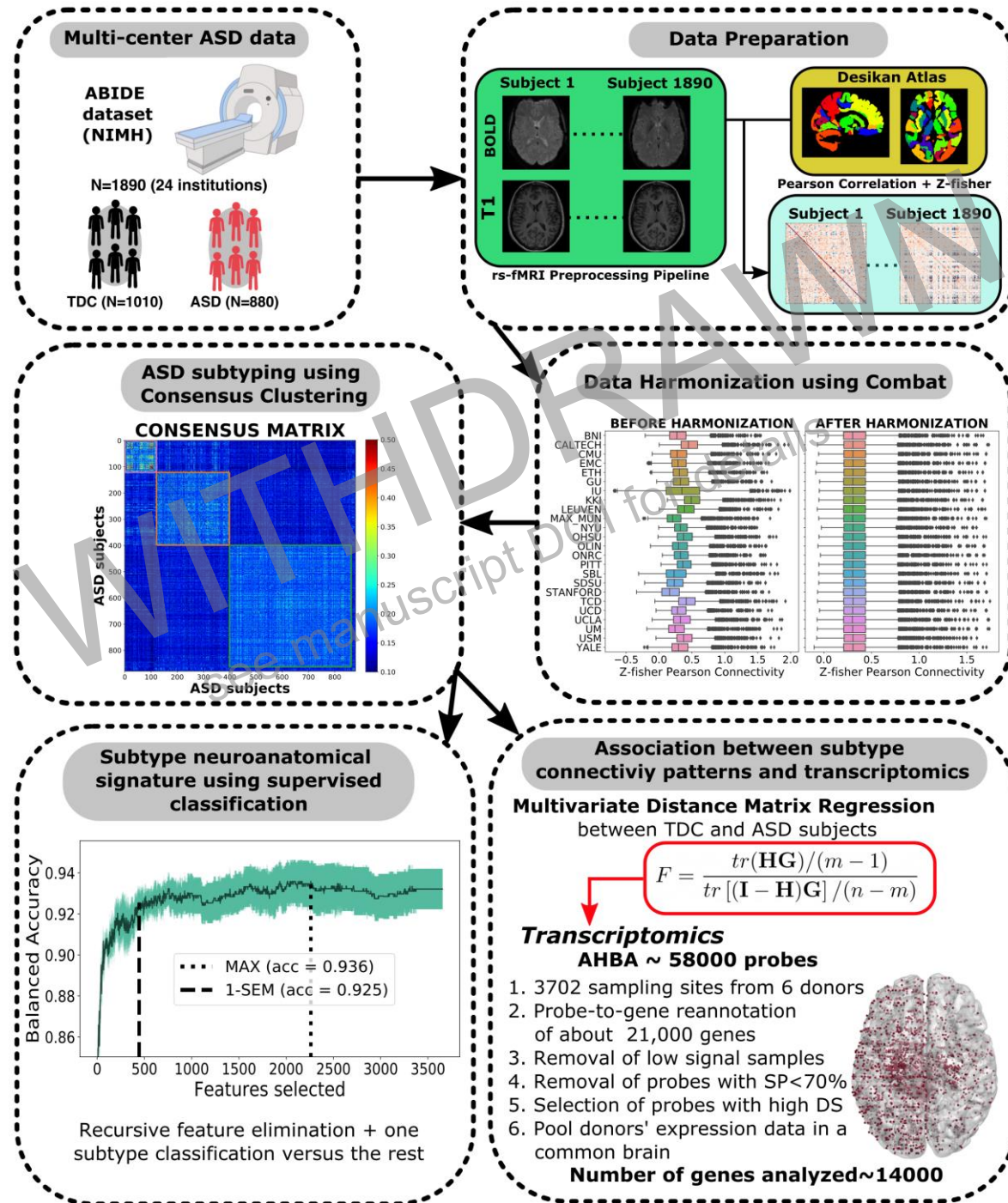
Figure 1. General workflow. Preparation and pre-processing of the ABIDE dataset. Harmonization using the Combat algorithm to remove institution, age and sex effects. ASD subtyping through a consensus clustering approach applied to brain connectivity matrices. Characterization of subtypes from their neurobiological signature and neurogenetic profiles using brain maps of the entire transcriptome.
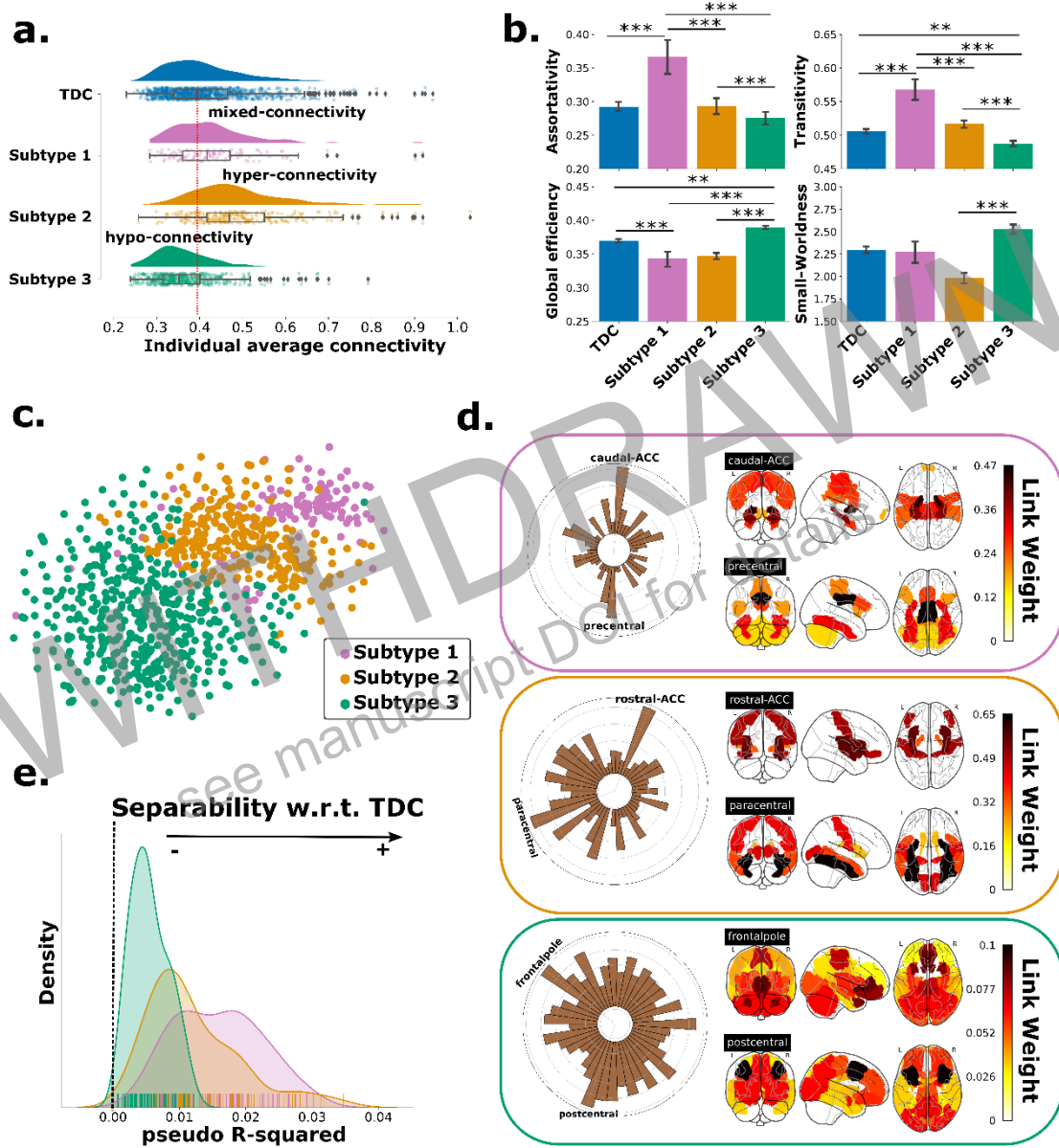
Figure 2. The connectivity class and neurobiological signature of each subtype. (a) Raincloud plots of the individual average connectivity values for the TDC group (blue) and the three ASD subtypes (pink, orange, and green). The mean value of the TDC group is marked as the baseline by a dashed red line. Values greater than the baseline correspond to hyperconnectivity and those below the baseline hypoconnectivity. Note how subtype 2 has hyperconnectivity, subtype 3 hypoconnectivity and subtype 1 a mixture of the two, the latter with a histogram and mean value similar to that of the TDC group. **(b)** The differences between groups in various measures of topology, such as assortativity, transititvity, global efficiency and small-worldness calculated after thresholding individual connectivity matrices to maintain positive link values greater than 10%.

38

Statistical differences were assessed using OLS, controlling for age, sex, and IQ scores (FIQ, VIQ and PIQ) and for average connectivity, and correcting for multiple testing through the Bonferroni-Holm procedure: *p<0.05, **p<0.01, ***p<0.001. When comparing panels a and b, one realizes that the existing hyperconnectivity in subtype 2, like the hypoconnectivity of subtype 3, is not the result of linearly scaling the weight values by a constant, since significant differences appear between some of these topological metrics and therefore, the hyper and hypoconnectivity profiles are much more complex, reflecting non-linear effects whose mechanisms are unknown. **(c)** The t-SNE visualization of ASD individuals using a combination of the most discriminating links per ASD subtype resulting from a Recursive Feature Elimination (RFE) and a linear Support Vector Machine (SVM). The plot shows quite decent clustering. **(d)** *Brown barplots on the left:* For each ASD subtype, the total number of selected links per bilateralized region after recursive feature elimination. *Glass brain plots on the right:* For those two regions with the largest number of selected links, the absolute link weight maps are depicted on the brain. **(e)** Separability of each ASD subtype with respect to the TDC group by illustrating the distribution of the pseudo R-square statistics across regions estimated from a Multivariate Distance Matrix Regression (MDMR).
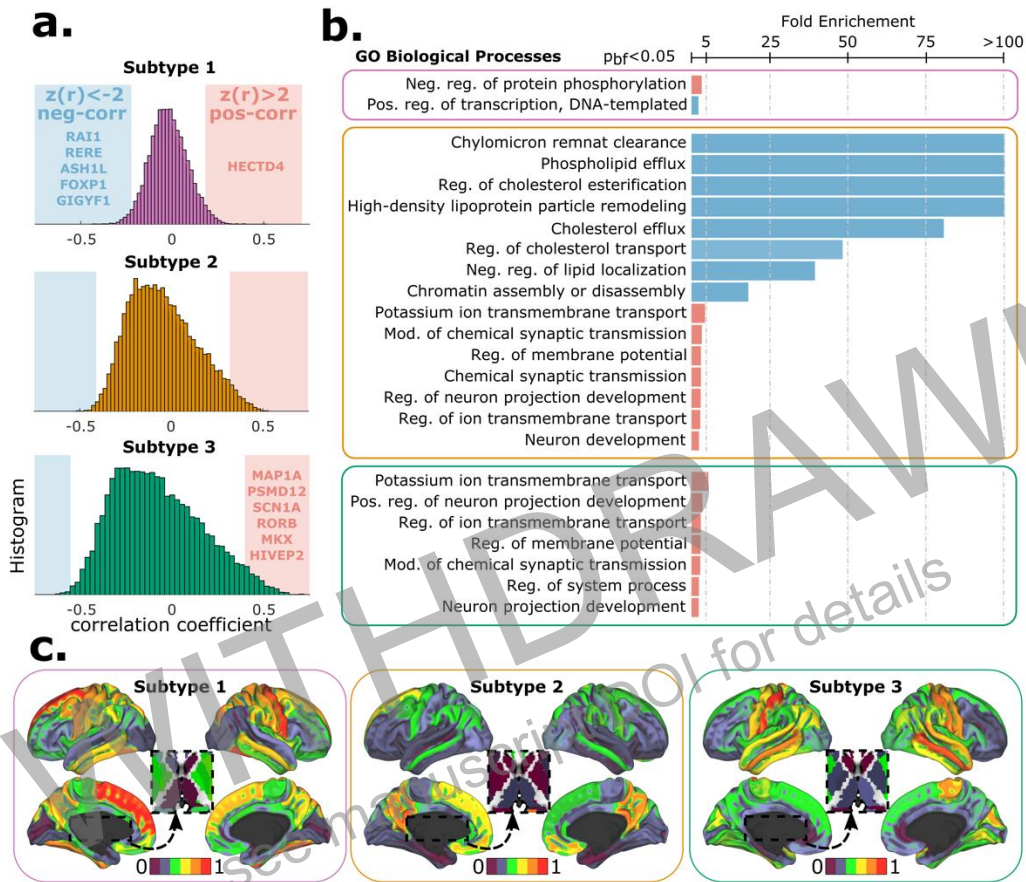
Figure 3. Data-driven strategy to define the association between transcriptome and its separability for each ASD subtype with respect to its TDC. **(a)** Histograms of the spatial correlations between the Z-scores of separability measured by the F-statistics and the transcriptional activity for each gene. This procedure was repeated for each subtype. The tail of the N genes (Z < -2) is marked by a blue rectangle and the tail of the P genes (Z > 2) by a red one. When we compared the lists of N and P genes separately with the list of hypothesis-driven genes, those with a significant association, and that therefore survive surrogate-data testing, are indicated for each subtype next to each tail. For subtype 1, we found the HECTD4 gene in the P-tail, and the RAI1, RERE, ASH1L, FOXP1 and GIGYF1 genes in the N-tail. No gene existed for subtype 2, neither in the N gene tail nor in the P gene tail. For subtype 3, no N gene was evident but as P genes, we obtained MAP1A, PSMD12, SCN1A, RORB, MKX and HIVEP2. **(b)** Gene enrichment for the *GO biological process* tags. The enrichment procedure was performed six times, for each subtype, and for the two groups of N and P genes obtained after applying DDS. The enrichments were low for subtype 1, in both lists N and P-genes, and similar for subtypes 2 and 3 in the list of P genes related to potassium ion transmembrane transport, neuron projection, membrane potential regulation, and chemical synaptic transmission. Remarkably, the enrichment we found for the N genes of subtype 2

40

was extremely high, greater than 100-fold enrichment in some cases, and corresponding to chylomicron clearance, phospholipid efflux, lipid regulation, and cholesterol esterification, transport and efflux. **(c)** Normalized F-statistic for each subtype. Note that for each brain region we have one value of F, therefore the brain maps we are plotting correspond to the normalized F value for each brain region (with a total number of 86 regions per map). In addition, it is also important to emphasize that for each value of F, we compared the values of the connectivity pattern of each region between all the ASD patients within each subtype and the TDC subjects.

Sub-score Name	Subtype 1	Subtype 2	Subtype 3	p_{unc}
FIQ	104.10	106.98	106.54	0.27
VIQ	102.68	106.26	106.87	0.12
PIQ	104.37	106.29	105.94	0.62
ADI_R_SOCIAL_TOTAL_A	18.08	19.46	19.45	0.10
ADI_R_VERBAL_TOTAL_BV	16.18	15.22	15.27	0.23
ADI_RRB_TOTAL_C	6.13	5.82	5.72	0.40
ADI_R_ONSET_TOTAL_D	3.15	3.12	3.08	0.90
ADOS_TOTAL	12.00	11.35	11.52	0.52
ADOS_COMM	3.65	3.55	3.73	0.48
ADOS_SOCIAL	8.35	7.70	7.71	0.21
ADOS_STEREO_BEHAV	2.31	1.91	1.90	0.22
ADOS_GOTHAM_SOCAFFECT	9.46	9.17	9.64	0.52
ADOS_GOTHAM_RRB	3.40	2.95	3.09	0.19
ADOS_GOTHAM_TOTAL	12.78	12.11	12.74	0.39
ADOS_GOTHAM_SEVERITY	7.20	6.90	7.20	0.36

Table 1: Discrimination of the ASD subtypes based on the behavioural scores. The average behavioural scores of the individuals within each ASD subtype and the uncorrected p-values provided by a one-way ANOVA test. None of the scores could differentiate between the three subtypes.

42

Supplementary information
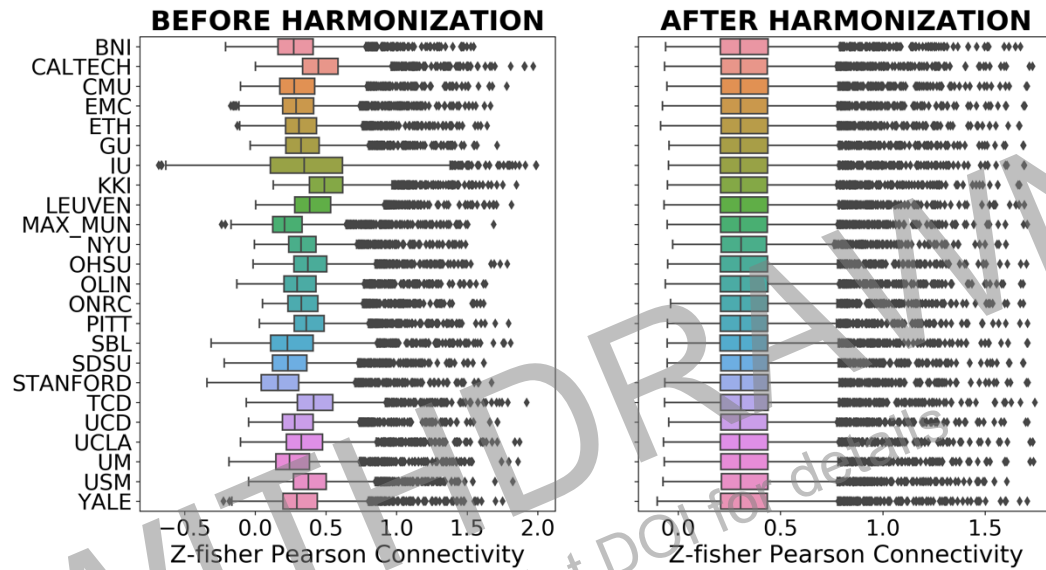
Figure S1. Effect of data harmonization on the connectivity values obtained at the different scanning institutions. Using Combat for data harmonization, we show the original values of the average connectivity matrix before (left) and after (right) harmonization. Different points in the histogram correspond to different links in the population connectivity matrix, calculated here for all participants: ASD and TDC.
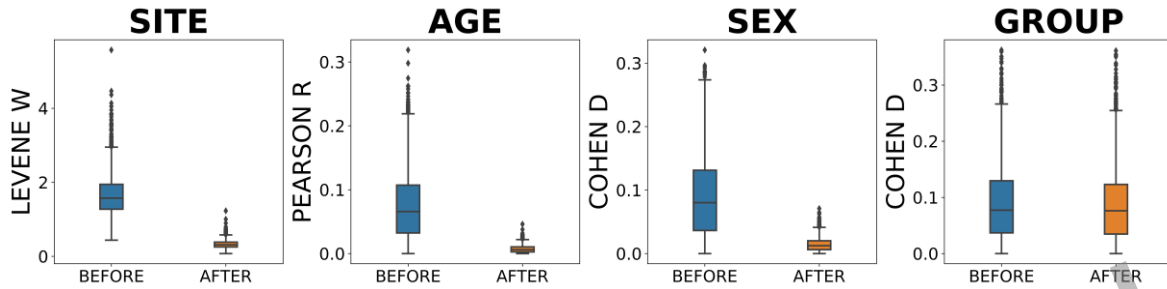
Figure S2. Effects of data harmonization on effect sizes. Each link value in the connectivity matrix for each subject was tested (including the TDC and ASD data), and their statistics were estimated at the sample level before and after Combat harmonization. Regarding the *site* variable, Levene's statistics were calculated, Pearson correlations were used for *age*, and the binary variables of *sex* and *group* (TDC or ASD) were studied with the Cohen D metric.
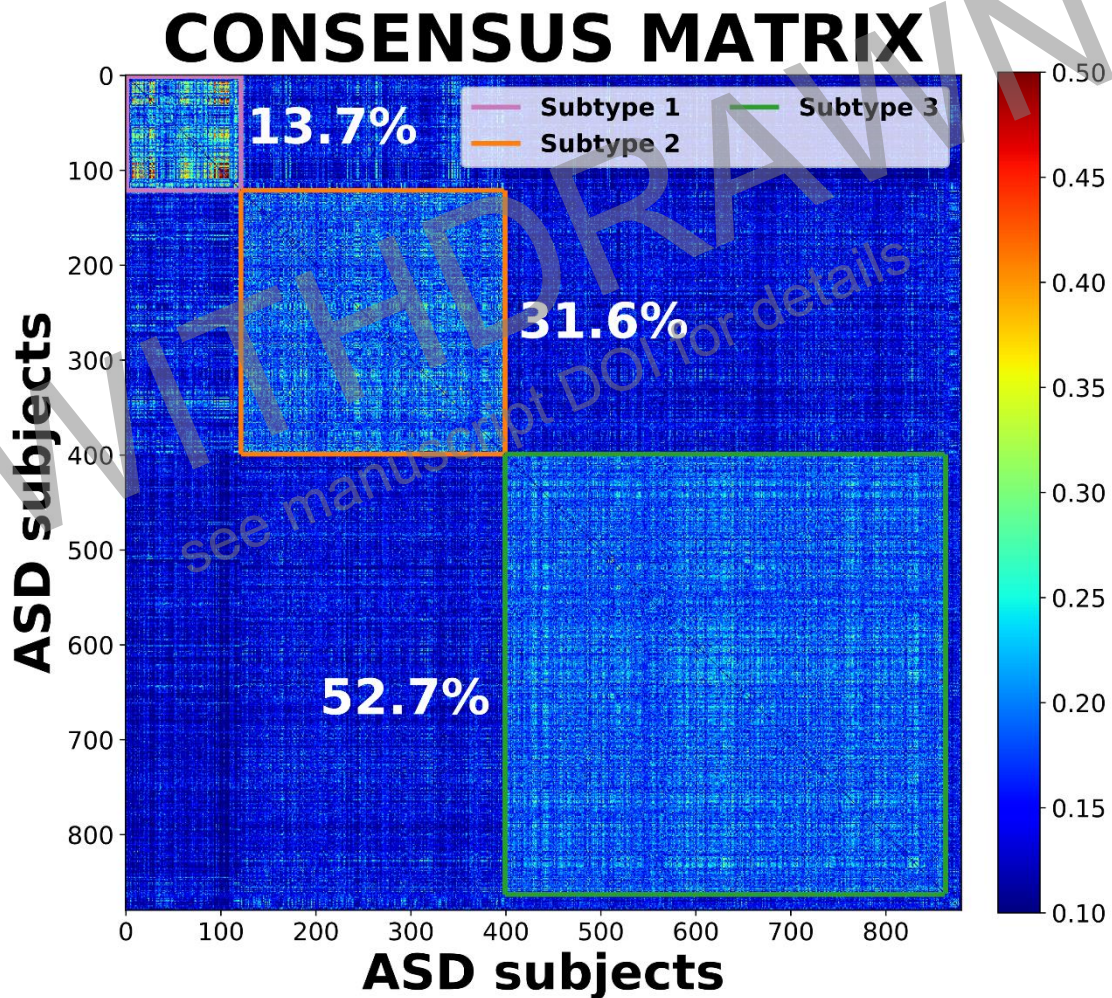
Figure S3. Consensus clustering matrix of the ASD cohort (N=880). After applying a consensus clustering approach to the individual connectivity matrices within the ASD group, a consensus matrix was computed that represents the probability of pairs of individuals belonging to the same cluster, which we have called 'subtype'. Here, the consensus matrix has been rearranged according to the individuals' membership provided by a community detection algorithm and it displays three distinct ASD subtypes.
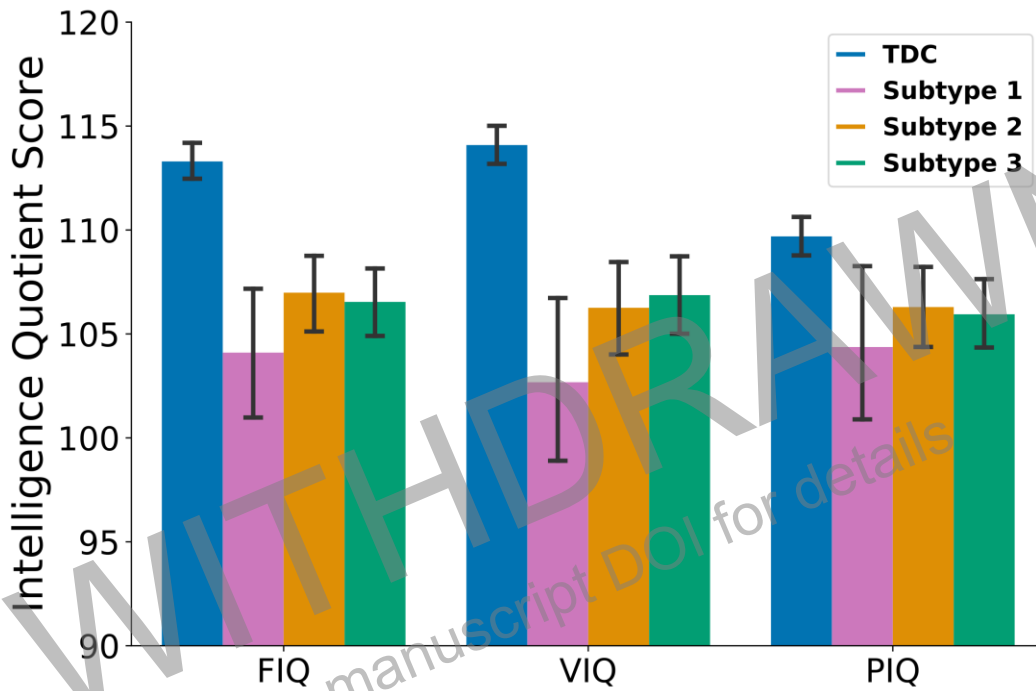
Figure S4. Intelligence Quotient Scores. Mean and the 95% bootstrapped confidence intervals of the Full Intelligence Quotient score (FIQ), Verbal Intelligence Quotient score (VIQ) and Performance Intelligence Quotient score (PIQ) for the TDC group and ASD subtypes. Different subgroups (TDC, subtype 1-3) are represented in a different colour.
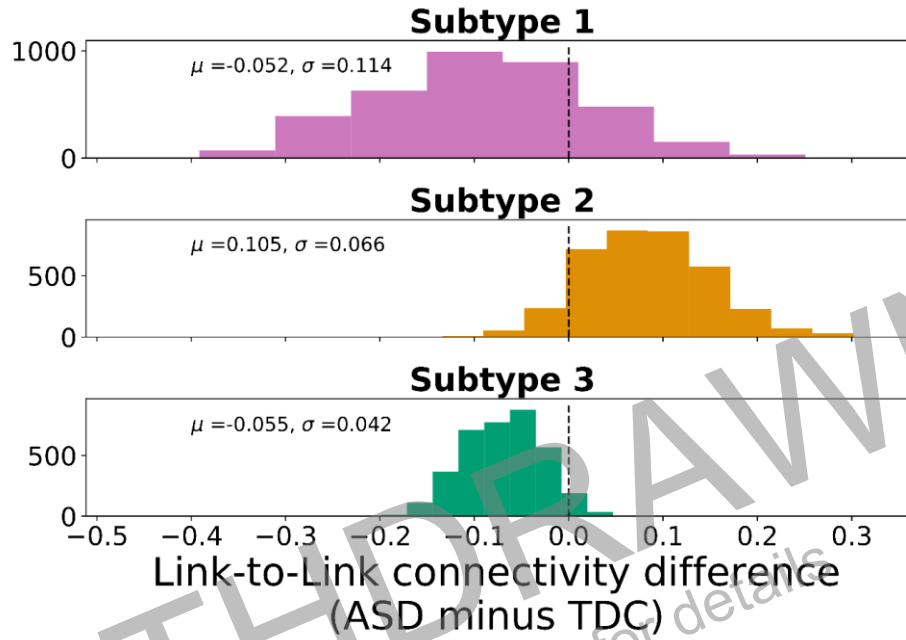
Figure S5. Link connectivity differences between each of the subtypes and the TDC group.
Each connectivity link is regressed onto a categorical binary variable that assigns individuals to the TDC or ASD subtype, controlling for age, sex and IQs (FIQ, VIQ and PIQ).
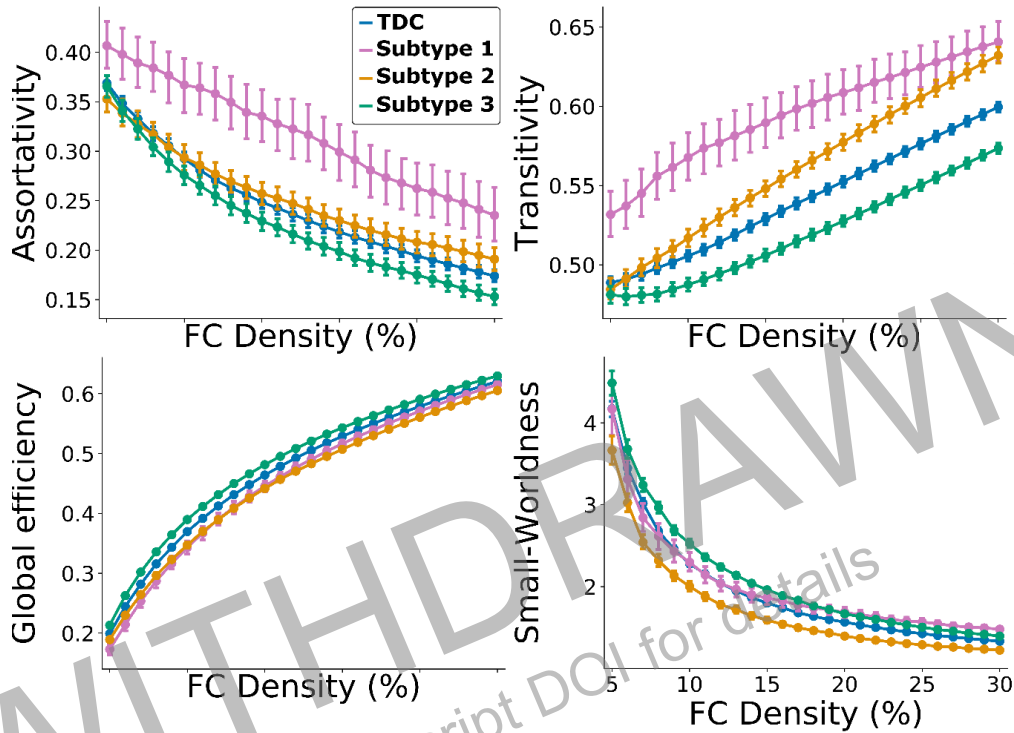
Figure S6. Threshold dependent global network properties of the TDC group and ASD subtypes. Assortativity, transitivity, global efficiency and small-worldness are calculated for the TDC group and ASD subtypes for different thresholds that aim to keep a fixed number of largest positive connections, represented here by the FC density percentages in the individual connectivity matrices. Each point displays the mean and 95% confidence intervals estimated from 1,000 bootstrap samples.
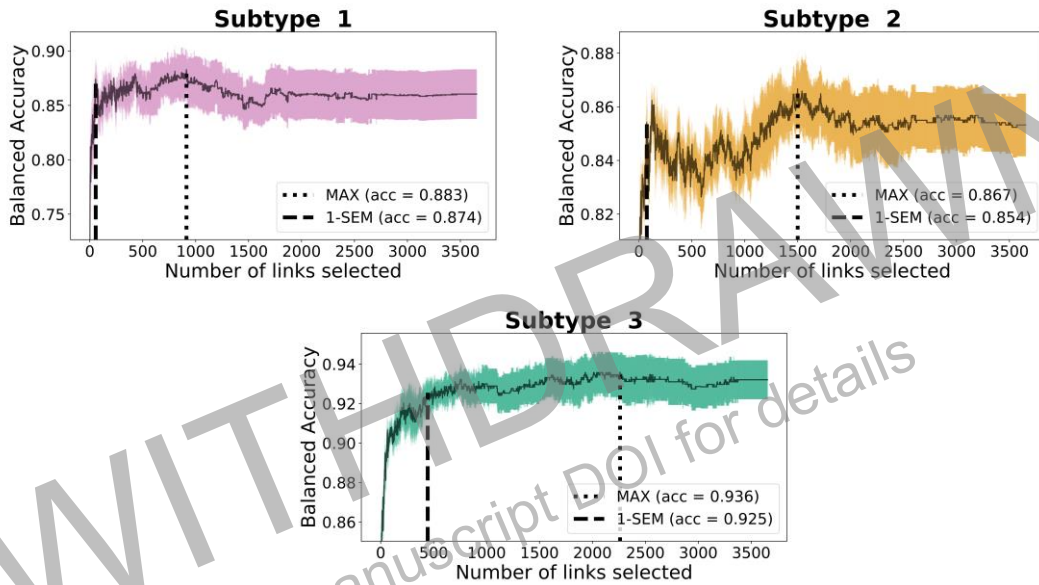
Figure S7. Number of links selected to define the neurobiological signature of each subtype. A Recursive Feature Elimination step with a Support Vector Machine classifier was employed to select the number of links that best discriminate each ASD subtype from the rest of the subtypes. In order to further reduce this number, a one standard error rule was applied by selecting the smallest number of links whose performance is within one standard error with respect to the highest performer.
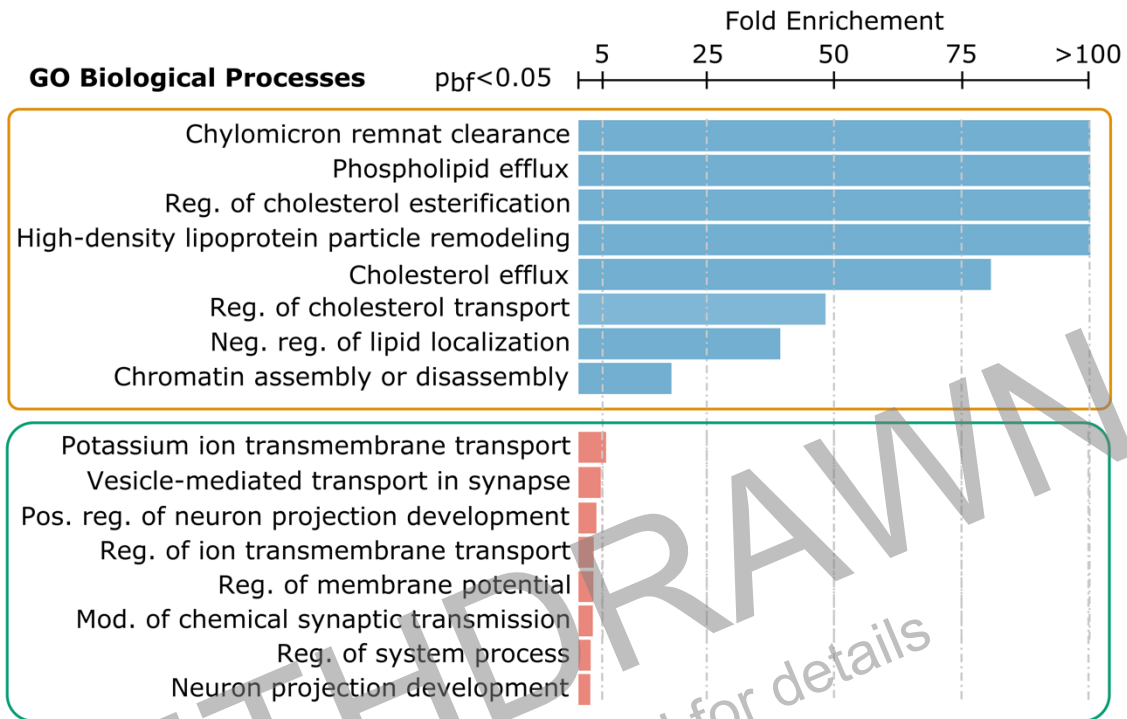
Figure S8. Enrichment for genes such that $|Z| > 2$ and $|r| > 0.4$. For each subtype, two enrichments were performed, one for the positive tail and another for the negative tail. In contrast to the enrichment shown in figure 3, obtained for the genes fulfilling the condition of $|Z| > 2$ where $||$ indicates absolute value, here we also applied the restriction of having an observed correlation such that $|r| > 0.4$. In this enrichment of highly correlated genes, no enrichment was obtained for subtype 1 in either of the tails. By contrast, the enrichment of the positive tail disappeared for subtype 2 but that related to cholesterol and other lipoproteins in the negative tail was preserved. Similarly, the enrichment of the positive tail was preserved for subtype 3.
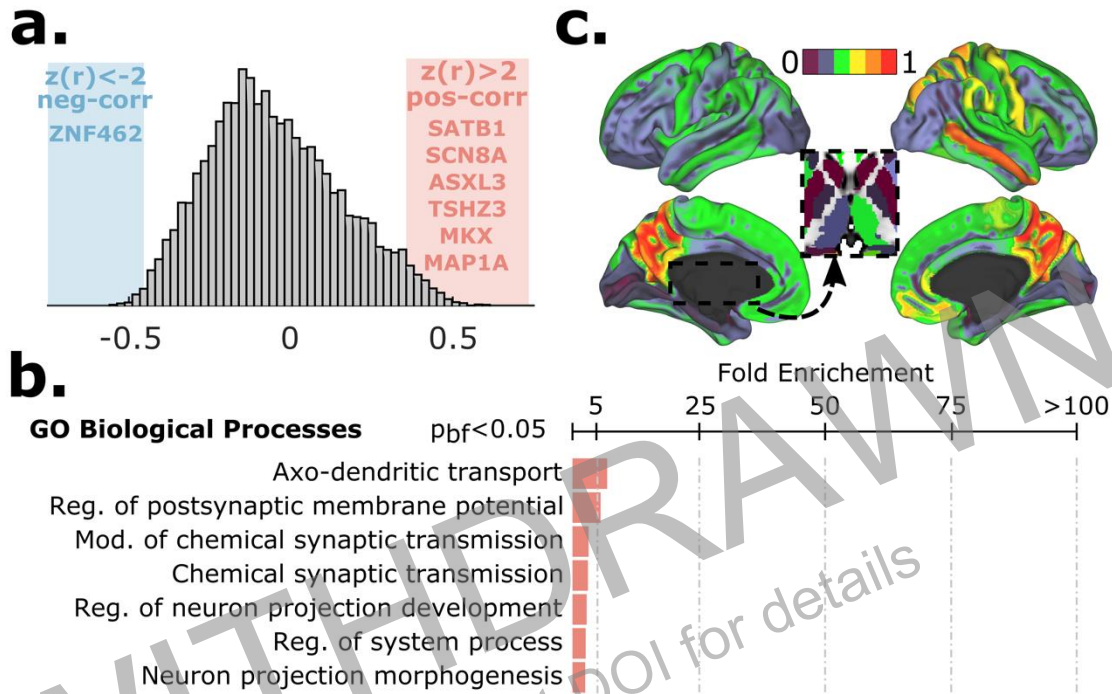
Figure S9. A figure similar to figure 3 but without subtyping. The results of the DDS were used for the entire set of ASD patients (without subtyping) to assess the associations with the transcriptome and their separability with respect to the TDC. **(a)** Histograms of the spatial correlations between the Z-scores of the separability measured through F-statistics and the transcriptional activity for each gene. The tail of the N genes (Z < -2) is indicated with a blue background and that of the P genes (Z > 2) in red. When we compared the lists of the N and P genes separately to the list of the hypothesis-driven genes, those with a significant association that therefore survived surrogate-data testing are indicated next to each tail. For the negative tail, we found the gene ZNF462, and for the positive one, the genes SATB1, SCN8A, ASXL3, TSHZ3, MKX and MAP1A. **(b)** Gene enrichment for the GO tags of biological process was performed twice, one for the groups of N genes and another for the P genes. There is no enrichment in the negative tail, and in the positive tail the enrichment was similar to that of the positive tails of subtypes 2 and 3. Remarkably, the enrichment found for subtype 2 with cholesterol-derived processes no longer existed when subtyping was not performed. **(c):** Normalized brain map of the F-statistic, more similar to that obtained for subtype 3.
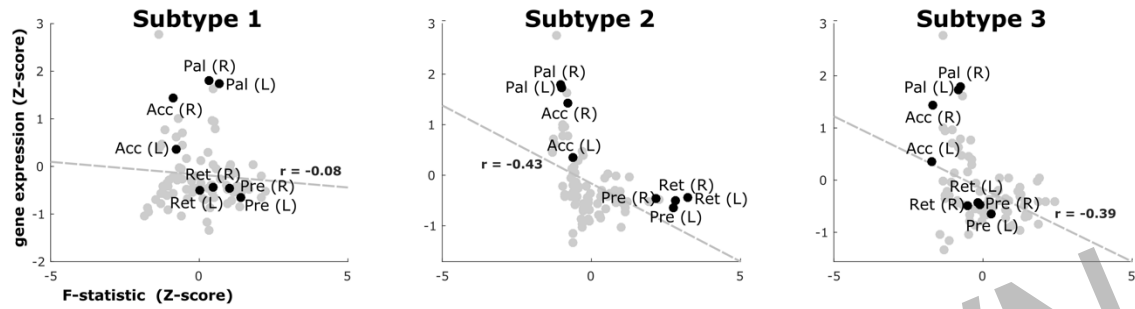
Figure S10. Expression as a function of the pattern of connectivity for the cholesterol-related gene APOC2. We show scatter plots between the F statistic (which explains the separability between the connectivity patterns of each brain region with other regions relative to the TDC group) and the expression. The different points correspond to different regions of the brain, and we highlight the Pal (R, right pallidum), Pal (L, left pallidum), Ret (R, right retrosplenial), Ret (L, left retrosplenial), Pre (R, right precuneus), Pre (L, left precuneus). Similar behavior occurs for APOE and APOC1. The spatial correlation values for each ASD subtype are given by r.

Institution name	Number of subjects contributed	Mean Age ($\pm$ SD)	Sex distribution (Female)	Number of ASD cases
BNI (II)	58	38.52 ± 15.50	0	29
CALTECH (I)	38	28.16 ± 10.64	8	19
CMU (I)	26	26.81 ± 5.69	6	13
EMC (II)	49	8.21 ± 1.06	10	23
ETH (II)	35	22.84 ± 4.50	0	11
GU (II)	101	10.69 ± 1.63	34	47
IU (II)	37	24.43 ± 7.59	9	17
KKI (I and II)	263	10.27 ± 1.29	84	77
LEUVEN (I)	64	17.98 ± 4.96	8	29
MAX_MUN (I)	53	26.89 ± 11.99	7	22
NYU (I and II)	285	13.08 ± 6.62	47	150
OHSU (I and II)	91	10.96 ± 2.04	35	37
OLIN (I)	36	16.81 ± 3.49	5	20
ONRC (II)	18	22.39 ± 3.58	5	7
PITT (I)	57	18.90 ± 6.88	8	30
SBL (I)	30	34.37 ± 8.60	0	15
SDSU (I and II)	93	13.55 ± 2.81	16	47
STANFORD (I)	25	10.11 ± 1.50	6	14
TCD (I and II)	91	16.26 ± 3.56	0	45
UCD (II)	32	14.78 ± 1.839	8	18
UCLA (I and II)	89	12.81 ± 2.40	10	51
UM (I)	131	14.18 ± 3.21	27	57
USM (I and II)	133	21.83 ± 7.7	5	75
YALE (I)	55	12.74 ± 2.89	16	27
ALL	1890	16.07 ± 8.88	353	880

Table S1: Main data characteristics for each Institution participating in our study. BNI = Barrow Neurological Institute; CALTECH = California Institute of Technology; CMU = Carnegie Mellon University; EMC = Erasmus University Medical Center Rotterdam; ETH = ETH Zürich; GU = Georgetown University; IU = Indiana University; KKI = Kennedy Krieger Institute; LEUVEN = University of Leuven; MAX_MUN = Ludwig Maximilians University Munich; NYU =

NYU Langone Medical Center; OHSU = Oregon Health and Science University; OLIN = Olin; Institute of Living at Hartford Hospital; ONRC = Olin Neuropsychiatry Research Center, Institute of Living at Hartford Hospital; PITT = University of Pittsburgh School of Medicine; SBL = Social Brain Lab, Netherlands Institute for Neurosciences; SDSU = San Diego State University; STANFORD = Stanford University; TCD = Trinity Centre for Health Sciences; UCD = University of California Davis; UCLA = University of California Los Angeles; UM = University of Michigan; USM = University of Utah School of Medicine; YALE = Yale Child Study Center. I: ABIDE 1. II: ABIDE 2.

Sub-score Abbreviation	Description	Range of Values	Score vs Severity
FIQ	Full IQ standard score	30-170	Lesser scoring, worse cognitive performance
VIQ	Verbal IQ standard score	31-169	Lesser scoring, worse cognitive performance
PIQ	Performance IQ standard score	31-166	Lesser scoring, worse cognitive performance
ADI_R_SOCIAL_TOTAL_A	Reciprocal Social Interaction subscore (A)	0-30	Greater scoring, more abnormal function
ADI_R_VERBAL_TOTAL_BV	Abnormalities in communication subscore (B)	0-26	Greater scoring, more abnormal function
ADI_RRB_TOTAL_C	Restricted, Repetitive, and Stereotyped Patterns of Behaviour subscore (C)	0-12	Greater scoring, more abnormal function
ADI_R_ONSET_TOTAL_D	Abnormality of Development Evident at or Before 36 Months subscore (D)	0-5	Greater scoring, more abnormal function
ADOS_TOTAL	Classic Total ADOS Score (Communication subscore + Social interaction subscore)	0-22	Greater scoring, more abnormal function
ADOS_COMM	Communication Total subscore of the Classic ADOS	0-8	Greater scoring, more abnormal function
ADOS_SOCIAL	Social Total subscore of Classic ADOS	0-14	Greater scoring, more abnormal function
ADOS_STEREO_BEHAV	Stereotyped Behaviours and Restricted Interests Total subscore of the Classic ADOS		Greater scoring, more abnormal function
ADOS_GOTHAM_SOCAFFECT	Social Affect Total subscore for Gotham Algorithm of the ADOS	0-20	Greater scoring, more abnormal function
ADOS_GOTHAM_RRB	Restricted and Repetitive Behaviours Total subscore for Gotham Algorithm of the ADOS	0-8	Greater scoring, more abnormal function
ADOS_GOTHAM_TOTAL	Social Affect Total + Restricted and Repetitive Behaviours Total	0-28	Greater scoring, more abnormal function
ADOS_GOTHAM_SEVERITY	Individually Calibrated Severity score for Gotham Algorithm of the ADOS	1-20	Greater scoring, more abnormal function

Table S2: Behavioural assessment of ASD patients in different tests, indicating their description, the range of numeric values and their relation with severity.

Social Network region	Corresponding Freesurfer regions
Cerebellum	Cerebellum
Extrastriate Body Area	Lateral Occipital
Ventromedial Prefrontal Cortex	Superior Frontal-Medial Orbitofrontal
Temporal Pole	Temporal Pole
Fusiform Face Area	Fusiform Gyrus
Retrosplenial Cortex Area	Isthmus Cingulate
Posterior Cingulate	Posterior Cingulate
Superior Temporal Gyrus	Superior Temporal
Rostral Anterior Cingulate	Rostral Anterior Cingulate
Dorsomedial Prefrontal Cortex	Superior Frontal
Temporoporietal Junction	Inferior Parietal-Superior Temporal Sulcus
Insula	Insula
Amygdala	Amygdala

Table S3: Social Network regions obtained from the regions in the Desikan-Killiany atlas.

Gene-symbol	p-value	Class	$r_{observed}$
Subtype 1			
RAI1	**0.010**	N	-0.261
RERE	**0.013**	N	-0.226
ASH1L	**0.021**	N	-0.264
FOXP1	**0.040**	N	-0.237
HECTD4	**0.045**	P	0.195
GIGYF1	**0.046**	N	-0.279
DPYSL2	0.058	P	0.202
AUTS2	0.067	N	-0.256
GNAI1	0.068	P	0.188
PHF2	0.085	N	-0.246
AP2S1	0.121	P	0.186
TCF7L2	0.133	N	-0.230
ELAVL3	0.139	P	0.244
Subtype 2			
SATB1	0.066	P	0.470
KCNB1	0.094	P	0.323
SCN8A	0.096	P	0.446
ASXL3	0.100	P	0.335
ZNF462	0.102	N	-0.432
TSHZ3	0.108	P	0.384
MKX	0.122	P	0.391

Gene-symbol	p-value	Class	$r_{observed}$
MAP1A	0.127	P	0.321
RORB	0.138	P	0.349
PSMD12	0.148	P	0.394
Subtype 3			
MAP1A	**0.026**	P	0.633
PSMD12	**0.030**	P	0.539
SCN1A	**0.031**	P	0.465
RORB	**0.038**	P	0.578
MKX	**0.041**	P	0.534
HIVEP2	**0.049**	P	0.415
KCNB1	0.062	P	0.483
SCN8A	0.062	P	0.476
SATB1	0.075	P	0.463
TSHZ3	0.078	P	0.472

Table S4. Statistical significance (p-value) of hypothesis-driven genes using surrogate-data for each ASD subtype. The table indicates the N or P cluster to which each gene belongs, the spatial correlation observed between the map of F-statistics and the expression of each gene. The uncorrected p-values are shown as none of them survived FDR corrections.

Gene-symbol	Class	Z
Subtype 1		
FAM110A	N	-4,09
PLAGL1	N	-3,93
RAD54B	N	-3,82
TRPC3	N	-3,60
THSD4	N	-3,53
RADIL	N	-3,53
NTNG1	N	-3,40
CMYA5	N	-3,28
CMKLR1	N	-3,22
4BRINP3	N	-3,15
OPN1MW	N	-3,09
FAM13B	N	-3,09
BRINP1	N	-3,08
FGF13	N	-3,08
SYT6	N	-3,06
PCDH11Y	N	-3,06
PRDM11	N	-3,06
FSTL1	N	-3,03
PDE3A	N	-2,95
CTSC	N	-2,93
KLHL36	N	-2,84
WDR19	N	-2,83
CD3EAP	N	-2,81
JAK2	N	-2,80
EML4	N	-2,80
CCDC136	N	-2,80
CUX1	N	-2,78
VCX3B	N	-2,77
IL7R	N	-2,77
RFX5	N	-2,76
NCOA3	N	-2,75
WNT3	N	-2,75
ESRRG	N	-2,73
OSBPL5	N	-2,73

UXS1	N	-2,71
ANKRD26	N	-2,69
SRSF5	N	-2,68
VCX2	N	-2,67
DPT	N	-2,64
LRRC49	N	-2,64
ZSWIM4	N	-2,64
GOLGA8N	N	-2,62
SHISAL2A	N	-2,60
TBX1	N	-2,59
RGS16	N	-2,59
PCDHGA2	N	-2,58
N4BP2	N	-2,57
ZIM2	N	-2,57
TIAM1	N	-2,56
IGSF3	N	-2,56
SUGP2	N	-2,55
PDS5B	N	-2,55
EFNA2	N	-2,53
GIGYF1	N	-2,53
PLAU	N	-2,53
ZC3HAV1L	N	-2,53
CARD9	N	-2,53
IP6K2	N	-2,52
MCM3	N	-2,52
RORA	N	-2,52
EEPD1	N	-2,51
NAALAD2	N	-2,51
TTC14	N	-2,50
C5orf24	N	-2,50
FAM20C	N	-2,49
TMEM132E	N	-2,49
PCDHGA12	N	-2,48
TRANK1	N	-2,48
PBX4	N	-2,48
NR2F1	N	-2,48
TCOF1	N	-2,48
GPR161	N	-2,48
CRABP1	N	-2,47
SLC12A5	N	-2,47
PLEKHA5	N	-2,47
SRRM3	N	-2,47
C6orf47	N	-2,47
ANKRD36BP1	N	-2,47
UPP1	N	-2,46
LRRC17	N	-2,45
SPON1	N	-2,45
NEXN	N	-2,43

AMDHD2	N	-2,42
ZNF148	N	-2,42
TMEM216	N	-2,42
MKNK2	N	-2,41
TCTN2	N	-2,41
PLK5	N	-2,40
HIST4H4	N	-2,40
HMGN3	N	-2,39
ZNF830	N	-2,39
ZBTB40	N	-2,39
RBBP6	N	-2,39
ASH1L	N	-2,39
TTC25	N	-2,39
TBL1Y	N	-2,37
ABCG1	N	-2,36
RAI1	N	-2,36
TCEAL2	N	-2,36
GPX7	N	-2,36
LAMA3	N	-2,36
ECD	N	-2,35
BCOR	N	-2,34
DNAJC25	N	-2,34
LOC100506990	N	-2,34
TAP1	N	-2,34
RNF126	N	-2,34
BMP7	N	-2,34
XBP1	N	-2,34
USF2	N	-2,33
FAM95A	N	-2,33
C18orf21	N	-2,32
CLDN5	N	-2,32
RGPD2	N	-2,32
INHBA	N	-2,32
GPCPD1	N	-2,31
ATP10D	N	-2,31
AUTS2	N	-2,31
BACH2	N	-2,31
FSBP	N	-2,31
GPR146	N	-2,31
FAM86C2P	N	-2,30
FLYWCH2	N	-2,30
GOLGA6L10	N	-2,29
PLEKHM1P1	N	-2,29
VCX	N	-2,29
PSPC1	N	-2,29
LRRC23	N	-2,29
CCDC152	N	-2,29
NR1D1	N	-2,28

CCDC65	N	-2,28
IER2	N	-2,27
TBL1X	N	-2,27
TPTEP2-CSNK1E	N	-2,27
NPIPB2	N	-2,27
THSD7A	N	-2,27
EGFR	N	-2,26
ZNF764	N	-2,26
AKAP7	N	-2,26
CPN2	N	-2,26
EIF2A	N	-2,25
NAV2	N	-2,25
UST	N	-2,25
PPARD	N	-2,25
H1F0	N	-2,24
THG1L	N	-2,23
ZBTB43	N	-2,23
SRC	N	-2,23
RPL14	N	-2,23
LINC02018	N	-2,23
LRIG1	N	-2,23
GOLGA6L17P	N	-2,23
ELL	N	-2,23
FEN1	N	-2,23
HDAC7	N	-2,22
ZNF587	N	-2,22
RRP1B	N	-2,22
FAM133CP	N	-2,22
NBPF3	N	-2,21
TENM2	N	-2,21
PHF2	N	-2,21
RBM6	N	-2,21
WWOX	N	-2,21
PATE2	N	-2,20
PCDHGB7	N	-2,20
P3H1	N	-2,20
IGF1R	N	-2,19
TSHZ1	N	-2,19
GYPC	N	-2,19
EFCAB7	N	-2,18
CAMK2B	N	-2,18
PCDH11X	N	-2,18
MON1B	N	-2,17
PRB4	N	-2,17
GTPBP1	N	-2,17
NPNT	N	-2,17
VEGFC	N	-2,17

TBC1D8	N	-2,16
LNX2	N	-2,16
DUSP1	N	-2,16
LINC00431	N	-2,16
KHSRP	N	-2,15
VANGL1	N	-2,15
ZNF131	N	-2,15
MRAP2	N	-2,15
SYT15	N	-2,15
GOLGA8F	N	-2,15
PABPC5	N	-2,14
PLG	N	-2,14
RNU12	N	-2,14
THUMPD1	N	-2,14
LOC646626	N	-2,13
C20orf96	N	-2,12
RNF19B	N	-2,12
FOXP1	N	-2,12
STAG3L4	N	-2,12
RNLS	N	-2,12
HMGN1	N	-2,12
ZNF502	N	-2,11
CDH2	N	-2,11
DPY19L2	N	-2,11
CHFR	N	-2,11
TWNK	N	-2,11
B3GALNT2	N	-2,11
PAWR	N	-2,11
IGF2BP3	N	-2,10
MGC16025	N	-2,10
ENKD1	N	-2,10
MICB	N	-2,10
LRIG3	N	-2,10
ADCY9	N	-2,10
ZNF362	N	-2,09
RCC2	N	-2,08
CLMP	N	-2,08
EXOSC10	N	-2,08
MAP3K5	N	-2,08
HIST1H1C	N	-2,08
SEPT6	N	-2,08
RASL11A	N	-2,07
ZNF689	N	-2,07
HIST1H1D	N	-2,07
RNF43	N	-2,07
CNFN	N	-2,07
LINC01405	N	-2,07
C19orf66	N	-2,06

WITHDRAWN

see manuscript DOI for details

Gene	Type	Value
OLFML2A	N	-2,06
PDAP1	N	-2,06
JADE1	N	-2,06
LRRTM1	N	-2,06
TCF7L2	N	-2,06
THAP10	N	-2,06
DFFB	N	-2,05
PCDHGA3	N	-2,05
ATXN7	N	-2,05
THOC1	N	-2,05
ANTXR2	N	-2,05
RBM15	N	-2,05
PEG3-AS1	N	-2,04
TOP2B	N	-2,04
ZNF710	N	-2,04
ASB13	N	-2,04
ABCD4	N	-2,03
ZNF718	N	-2,03
RIMS2	N	-2,03
ZNF268	N	-2,03
ABCA9	N	-2,03
ZNF165	N	-2,03
OSBP2	N	-2,02
TPM2	N	-2,02
PCDHB18P	N	-2,02
CEP135	N	-2,02
CEP57	N	-2,02
RERE	N	-2,02
WRAP73	N	-2,02
ODF2L	N	-2,02
TXLNG	N	-2,02
CNNM1	N	-2,01
LATS2	N	-2,01
RPL23AP53	N	-2,01
SYNE1	N	-2,01
ZFYVE27	N	-2,01
AQR	N	-2,00
LTV1	N	-2,00
MKNK1	N	-2,00
SNX13	N	-2,00
DUXAP9	P	5,22
CLEC4G	P	4,00
C1QL3	P	3,97
KLK7	P	3,95
SLC22A18	P	3,71
MESP1	P	3,69
GPR20	P	3,68
COL5A2	P	3,38

WITHDRAWN

See manuscript DOI for details

PHLDA2	P	3,36
TSHZ2	P	3,29
OPTN	P	3,21
INPP4A	P	3,18
DPY19L1	P	3,16
TSPAN33	P	3,15
TUBB4A	P	3,11
DOC2A	P	3,09
PRUNE2	P	3,09
RMND5B	P	3,06
ACACA	P	3,06
MIR7-3HG	P	3,00
C1S	P	2,99
TFRC	P	2,98
BCL2L13	P	2,98
DPY19L1P2	P	2,97
PDE1A	P	2,93
CLN8	P	2,90
LINC01963	P	2,90
CTXN3	P	2,89
COG5	P	2,89
NKAIN2	P	2,89
SH3GLB2	P	2,88
AKAP8	P	2,86
AGTPBP1	P	2,85
EIF4G1	P	2,85
LINC00839	P	2,85
KIF1B	P	2,83
NDEL1	P	2,81
SYT17	P	2,80
SNRNP25	P	2,80
NADK2	P	2,80
PID1	P	2,80
RAB27B	P	2,80
EFNB2	P	2,78
WBP2	P	2,78
NIP7	P	2,77
PCYT2	P	2,77
SNCA	P	2,77
TMED9	P	2,77
JUN	P	2,76
GRM1	P	2,73
LRSAM1	P	2,73
GET4	P	2,72
TFPT	P	2,71
GALNS	P	2,71
MYOM2	P	2,71
PRSS1	P	2,69

WITHDRAWN

see manuscript DOI for details

XPO7	P	2,69
TM6SF1	P	2,67
CHCHD6	P	2,66
DMD	P	2,64
GRB14	P	2,63
COMMD9	P	2,63
DACH2	P	2,63
IDH3B	P	2,62
CNTN1	P	2,62
CCN3	P	2,62
POLR2E	P	2,62
SIDT1	P	2,62
RASD1	P	2,62
PKP3	P	2,62
GRIA4	P	2,61
NF2	P	2,60
MYLIP	P	2,59
ELAVL3	P	2,59
FJX1	P	2,58
EHBP1L1	P	2,58
MPP2	P	2,58
ARHGAP25	P	2,57
NWD2	P	2,56
TSTA3	P	2,55
MAGI1	P	2,55
NCOA7	P	2,54
WFDC1	P	2,54
C19orf24	P	2,54
SCPEP1	P	2,54
HTR1E	P	2,54
TYRO3	P	2,53
DCAF11	P	2,53
PITPNB	P	2,53
RBP4	P	2,53
ST3GAL1	P	2,52
UBA1	P	2,52
AVL9	P	2,51
LGALS1	P	2,51
XYLT1	P	2,51
GDF10	P	2,51
MADD	P	2,51
NT5DC3	P	2,51
PDXP	P	2,51
TSR3	P	2,51
SH3GL3	P	2,51
EARS2	P	2,50
ULK2	P	2,50
NUDT11	P	2,50

COL16A1	P	2,49
ZNF350	P	2,49
VKORC1L1	P	2,49
CNTN6	P	2,49
SULF2	P	2,48
CNR1	P	2,48
DSTN	P	2,48
KCNK1	P	2,48
INCENP	P	2,48
VPS26B	P	2,47
MAPK8IP1	P	2,47
CDT1	P	2,46
C16orf95	P	2,46
VRK3	P	2,45
PLXNA1	P	2,45
ADAMTSL3	P	2,44
MTCH1	P	2,44
ARHGDIG	P	2,44
THAP8	P	2,44
LOC100128239	P	2,44
CA10	P	2,44
TPI1	P	2,43
YPEL5	P	2,43
PRKCD	P	2,43
PRSS3	P	2,42
FARSA	P	2,42
AQP11	P	2,42
STX4	P	2,42
LINC01521	P	2,42
FXYD6	P	2,42
DUSP3	P	2,42
AGMAT	P	2,42
IQCJ-SCHIP1	P	2,41
DTNB	P	2,41
PDK3	P	2,41
CDC6	P	2,41
MFGE8	P	2,41
INSIG1	P	2,40
ATP6V0D1	P	2,40
DNASE1L1	P	2,40
LGI4	P	2,40
OXLD1	P	2,39
FADS3	P	2,39
BEND5	P	2,38
RUNDC3A	P	2,38
APBB1	P	2,38
ASL	P	2,37
SPON2	P	2,37

ADRA1B	P	2,37
PLCH1	P	2,36
FRMPD2B	P	2,36
GPATCH4	P	2,36
PIN1	P	2,36
MRPL37	P	2,35
TDRKH	P	2,35
PRSS3P2	P	2,35
LDHA	P	2,34
METTL6	P	2,34
MAPK8IP1P2	P	2,34
CORO1C	P	2,34
MTA3	P	2,34
PNCK	P	2,34
SRCIN1	P	2,34
TNFRSF14	P	2,33
S100A6	P	2,32
PCMT1	P	2,32
ATP6V1A	P	2,32
SLC39A11	P	2,32
KCNMB4	P	2,31
CIB2	P	2,31
RAET1E-AS1	P	2,31
CHP1	P	2,31
SIRPA	P	2,31
STMN2	P	2,30
NICN1	P	2,30
MFSD3	P	2,30
ALAS1	P	2,29
ARFGEF3	P	2,29
KIAA1143	P	2,29
NUDT10	P	2,29
PUSL1	P	2,29
FAM102B	P	2,28
OTUB1	P	2,28
LARGE1	P	2,28
FEZ1	P	2,28
STARD7-AS1	P	2,27
GDAP1L1	P	2,27
SMPD1	P	2,27
TBC1D9B	P	2,27
RITA1	P	2,27
MPND	P	2,27
WDR86	P	2,27
SBF1	P	2,27
NFKBIE	P	2,26
GRM8	P	2,26
NHLRC1	P	2,26

GRK6	P	2,26
STK39	P	2,26
HPCAL4	P	2,26
ZNF233	P	2,26
CUL2	P	2,25
DCTN2	P	2,25
RALBP1	P	2,25
IFITM10	P	2,25
PTPRK	P	2,25
NPTXR	P	2,24
FZR1	P	2,24
FEM1B	P	2,24
HK1	P	2,24
MYRIP	P	2,24
KCNS3	P	2,23
ASS1	P	2,23
NUDT16	P	2,23
RASL11B	P	2,23
CRYM	P	2,23
VAMP7	P	2,23
SLIT3	P	2,23
TPI1P2	P	2,23
ZC3H15	P	2,23
TCFL5	P	2,23
SVOP	P	2,22
WTIP	P	2,22
THAP7	P	2,22
CLPB	P	2,22
GPD1L	P	2,22
PTPN14	P	2,22
NDST2	P	2,22
AP1S1	P	2,21
BMS1P14	P	2,21
SSTR1	P	2,21
TP53TG3B	P	2,21
BAIAP2L2	P	2,21
ADAMTSL2	P	2,21
DCTN3	P	2,21
TCTA	P	2,21
GPD1	P	2,20
FIGNL2	P	2,20
NDUFAF4	P	2,20
TUBB2A	P	2,20
SMIM10L2B	P	2,20
TIMM17A	P	2,20
COLGALT2	P	2,20
SLC25A26	P	2,19
RAN	P	2,19

WITHDRAWN
see manuscript DOI for details

SIGMAR1	P	2,19
MED8	P	2,19
GPR22	P	2,19
ACTR10	P	2,19
CIDEC	P	2,19
RNF214	P	2,18
CTNNBIP1	P	2,18
SMYD3	P	2,18
DYNC1I1	P	2,18
RAB33A	P	2,18
GAL	P	2,17
DPYSL2	P	2,17
WDR26	P	2,16
TEX30	P	2,16
GMFB	P	2,16
RAPH1	P	2,16
PPP1R13B	P	2,16
TPGS2	P	2,16
RASAL1	P	2,15
ZC3H8	P	2,15
TMEFF2	P	2,15
AP1M1	P	2,15
SNX17	P	2,15
PEX11B	P	2,14
BABAM2	P	2,14
LSM11	P	2,14
TUBB4B	P	2,14
CMTR1	P	2,14
TUBA1A	P	2,13
YTHDF2	P	2,13
PDRG1	P	2,13
IARS	P	2,13
GPR150	P	2,12
TRAPPC6B	P	2,12
CERS4	P	2,12
PABPC1L2A	P	2,12
FIBP	P	2,12
HECA	P	2,12
ACTC1	P	2,12
KLC1	P	2,11
BLOC1S6	P	2,11
IGFBP5	P	2,11
ROBO2	P	2,11
IFI6	P	2,11
C1QL2	P	2,11
NAGK	P	2,11
SNX24	P	2,10
TUBA4A	P	2,10

PRMT2	P	2,10
SLIT1	P	2,10
NTMT1	P	2,10
HECTD4	P	2,10
DYRK2	P	2,10
USP20	P	2,10
TAF13	P	2,10
FBXL18	P	2,10
TENT4B	P	2,09
DLG2	P	2,09
OSTF1	P	2,09
KDM8	P	2,09
TIGD3	P	2,08
LINC02361	P	2,08
NTNG2	P	2,08
BAIAP3	P	2,08
RNASEK	P	2,08
DNAJA1	P	2,08
FAM219B	P	2,08
CD52	P	2,08
STBD1	P	2,08
B4GALT2	P	2,08
SELENOW	P	2,07
CMAS	P	2,07
ANKRD50	P	2,07
MPI	P	2,07
ACAT2	P	2,07
SMG9	P	2,07
CRIP2	P	2,07
FXYD6-FXYD2	P	2,06
WSB2	P	2,06
NUDT14	P	2,06
TNFAIP8L1	P	2,06
KHK	P	2,06
PTRH2	P	2,06
COL24A1	P	2,06
GSPT1	P	2,05
RTN4	P	2,05
PC	P	2,05
TRIM36	P	2,05
C2orf69	P	2,05
C11orf49	P	2,05
KCNA3	P	2,05
PDE4DIP	P	2,05
IL27RA	P	2,05
FITM2	P	2,04
DCP2	P	2,04
TBC1D24	P	2,04

WITHDRAWN
see manuscript DOI for details

Gene		Value
C1QL1	P	2,04
CDKN3	P	2,04
SKAP2	P	2,04
MTMR2	P	2,04
LINC00672	P	2,04
GNAI1	P	2,03
ATP6V1C1	P	2,03
TMEM70	P	2,03
ANKRD9	P	2,03
MAPK13	P	2,03
DGKA	P	2,03
CTXN1	P	2,03
UHMK1	P	2,02
RFPL1S	P	2,02
TGFBR3L	P	2,02
GPC6	P	2,02
WARS	P	2,02
RNF187	P	2,02
FRS3	P	2,02
AP2S1	P	2,02
ZNF784	P	2,01
LINC00260	P	2,01
ZNF184	P	2,01
GLRA3	P	2,01
PHB	P	2,01
SEPT9	P	2,01
CLPTM1	P	2,01
ITGA8	P	2,01
TERF2	P	2,01
ZFP64	P	2,01
PPARG	P	2,01
GRIN3A	P	2,01
CHRNA7	P	2,01
TNFRSF21	P	2,01
LOC101927974	P	2,00
COPS4	P	2,00
TUBA1C	P	2,00
KCNN2	P	2,00
TRUB1	P	2,00
MRPL4	P	2,00
Subtype 2		
MCM3	N	-2,54
LIMK2	N	-2,51
APOE	N	-2,36
EMILIN1	N	-2,34
MATN2	N	-2,33
INSYN2	N	-2,26
H2AFJ	N	-2,23

EPCAM	N	-2,19
PINLYP	N	-2,16
ABCG1	N	-2,16
LRP1	N	-2,14
RUVBL1	N	-2,14
PYCARD	N	-2,13
ZNF502	N	-2,13
CPNE3	N	-2,12
ZNF462	N	-2,12
PHKA1	N	-2,11
DPY19L3	N	-2,11
HIST2H2BE	N	-2,11
DNAJC12	N	-2,11
CA12	N	-2,10
ZCCHC12	N	-2,10
MAPK7	N	-2,09
APOC2	N	-2,08
KCNE5	N	-2,08
APOC1	N	-2,07
GPR75	N	-2,07
AQP5	N	-2,07
YBX1	N	-2,06
PCDHGB7	N	-2,05
SLC15A3	N	-2,05
H2BFS	N	-2,05
PGM5P4	N	-2,05
PRKRIP1	N	-2,05
HLA-DPB2	N	-2,04
GLYCTK	N	-2,04
H2AFB2	N	-2,04
ATP1B2	N	-2,03
LRCH2	N	-2,02
PDE3A	N	-2,02
ITGB7	N	-2,02
MTA2	N	-2,01
HIST1H2BE	N	-2,01
BMP2K	N	-2,01
BCCIP	N	-2,00
GPR52	N	-2,00
HTR2C	N	-2,00
COL5A2	P	3,13
DUXAP9	P	3,09
WIPF2	P	2,95
ZBTB16	P	2,92
ECM1	P	2,92
ID2	P	2,91
ADPRHL1	P	2,90
CLEC2L	P	2,87

BAIAP2L2	P	2,86
CDKL1	P	2,82
SATB1	P	2,82
TUBG2	P	2,82
RIMS3	P	2,81
DIRAS1	P	2,81
CABP1	P	2,81
RASD1	P	2,80
OPN3	P	2,79
CHML	P	2,78
MELTF	P	2,77
TPK1	P	2,76
MRTFA	P	2,76
BEND5	P	2,76
TTLL12	P	2,75
TOX	P	2,75
COL23A1	P	2,75
TM6SF1	P	2,74
SOHLH1	P	2,73
FRMPD2B	P	2,72
HAPLN4	P	2,71
MAFB	P	2,71
P2RX6	P	2,70
C9orf129	P	2,70
SCN8A	P	2,69
PKNOX2	P	2,69
PDK3	P	2,69
IPCEF1	P	2,68
TEX30	P	2,68
CLSTN1	P	2,66
MADCAM1	P	2,66
OXR1	P	2,66
FBXW7	P	2,66
STAC2	P	2,66
GPR20	P	2,66
PREP	P	2,65
FNDC4	P	2,65
SGTB	P	2,63
EFNA5	P	2,63
HLF	P	2,63
CNN1	P	2,62
MIR600HG	P	2,62
BRSK1	P	2,62
NUAK1	P	2,62
PLXDC1	P	2,62
MAPK10	P	2,61
NIPAL2	P	2,61
ZNF365	P	2,61

WITHDRAWN

see manuscript DOI for details

BASP1-AS1	P	2,60
HIVEP1	P	2,60
PKDCC	P	2,60
NUDT4P2	P	2,60
PLXNA1	P	2,60
EPHB6	P	2,60
FMN1	P	2,59
DLGAP1-AS4	P	2,59
TNFAIP1	P	2,59
ARHGAP25	P	2,59
HPS1	P	2,59
PNMA5	P	2,58
C11orf87	P	2,57
FAM162B	P	2,56
GUCA1B	P	2,56
TNFRSF14	P	2,56
MYO15A	P	2,56
NR2F1-AS1	P	2,54
SHROOM2	P	2,53
RGS6	P	2,53
ADRA1B	P	2,53
KCNJ8	P	2,53
KLHDC3	P	2,52
FBXO33	P	2,52
LMTK2	P	2,52
ARNTL2	P	2,52
BCL2L2	P	2,52
HTR1F	P	2,52
DBNDD1	P	2,51
VSNL1	P	2,51
DCAF6	P	2,51
JOSD1	P	2,50
PTPA	P	2,50
AACS	P	2,50
ST8SIA5	P	2,50
GPR22	P	2,50
MEF2C	P	2,49
DCLK1	P	2,49
NTN4	P	2,49
ABCF2	P	2,49
KHDRBS3	P	2,49
TOMM40L	P	2,48
MPPED1	P	2,48
MAPT	P	2,48
HCN1	P	2,48
ATRNL1	P	2,47
ST3GAL1	P	2,47
GPR150	P	2,47

WITHDRAWN

see manuscript DOI for details

CRTAC1	P	2,47
STRBP	P	2,47
TIGAR	P	2,47
ID3	P	2,46
ABHD8	P	2,46
MAP3K13	P	2,46
MIR124-2HG	P	2,46
OLFM3	P	2,45
TCERG1L	P	2,45
KCNA2	P	2,45
GABRA1	P	2,44
PHLDB2	P	2,44
PLPP7	P	2,44
COL12A1	P	2,43
ANKRD24	P	2,43
SMYD2	P	2,43
LINC00672	P	2,43
RAB40C	P	2,43
SLC45A4	P	2,43
SORL1	P	2,42
NUDT4	P	2,42
PI4KA	P	2,42
RND1	P	2,42
GPR26	P	2,41
PWARSN	P	2,41
LOC100507507	P	2,41
PSMD12	P	2,41
PDE4D	P	2,40
TRAF3	P	2,40
CHRM3	P	2,40
CRH	P	2,40
SUSD5	P	2,40
AARS	P	2,40
LRRC20	P	2,40
NAT8L	P	2,40
MKX	P	2,39
SPRED2	P	2,39
ISLR	P	2,39
UBE2Q2L	P	2,39
CIDEC	P	2,39
SLC16A7	P	2,38
RFPL1	P	2,38
MYOM2	P	2,38
ADCYAP1	P	2,38
RFPL2	P	2,38
RIMKLA	P	2,38
SSBP3	P	2,38
PLEKHM2	P	2,37

LOC440300	P	2,37
CENPBD1P1	P	2,37
SRRM4	P	2,37
NLK	P	2,37
LDLR	P	2,36
HYAL3	P	2,36
HTR2A	P	2,36
LINC02361	P	2,36
ELAVL4	P	2,36
LMO4	P	2,36
LOC729737	P	2,35
RGS4	P	2,35
KCNIP3	P	2,35
TSHZ3	P	2,35
CAMK2G	P	2,35
KCNMA1	P	2,35
DPP10-AS1	P	2,35
SNTB2	P	2,34
TUBA1B	P	2,34
FAM131B	P	2,34
SYTL2	P	2,34
OLFM1	P	2,34
SPRY4	P	2,34
PKP3	P	2,34
TIAM2	P	2,34
NREP	P	2,33
PHLDA2	P	2,33
KCNC2	P	2,33
UNC5A	P	2,33
NCALD	P	2,33
SV2B	P	2,32
CIAPIN1	P	2,32
FRMPD4	P	2,32
HDAC9	P	2,32
PTPRT	P	2,32
MRPL4	P	2,32
MAGI2	P	2,32
FXYD6	P	2,31
IFT57	P	2,31
TPTE2P6	P	2,31
IGFBP2	P	2,31
EHD3	P	2,31
NFIX	P	2,30
EMX1	P	2,30
APBB1	P	2,30
MFGE8	P	2,29
BICDL2	P	2,29
PAK1	P	2,29

NMT1	P	2,29
CUX2	P	2,29
VPS25	P	2,28
LYNX1	P	2,28
MOSMO	P	2,28
ST3GAL3	P	2,27
PFKP	P	2,27
PSMG3	P	2,27
SPRN	P	2,27
RPH3A	P	2,27
SBF1	P	2,27
CCDC3	P	2,27
TICAM1	P	2,27
DENND3	P	2,27
FHL2	P	2,26
GPR162	P	2,26
MGLL	P	2,26
FAM81A	P	2,26
LINC01963	P	2,26
CIDECP1	P	2,26
DNM1P35	P	2,26
ANKRD9	P	2,26
HK1	P	2,26
THAP7	P	2,25
KIF1A	P	2,25
JAK3	P	2,25
LRRC4	P	2,25
PELI3	P	2,25
TRHDE	P	2,24
CAND2	P	2,24
HECW1	P	2,24
FUT9	P	2,24
CHGA	P	2,24
OPCML	P	2,24
RAB11FIP3	P	2,24
PPM1L	P	2,24
GRK3	P	2,23
KCNAB2	P	2,23
NPM2	P	2,23
NANOS3	P	2,23
FXYD6-FXYD2	P	2,23
HRK	P	2,23
PNMA3	P	2,23
LOC388242	P	2,23
ATP6V1A	P	2,23
FOXK1	P	2,23
C1orf115	P	2,22
RAB11FIP5	P	2,22

PCSK1	P	2,22
VPS39	P	2,22
GAP43	P	2,22
LOC100996385	P	2,22
MCRIP1	P	2,21
TENT4A	P	2,21
STXBP5	P	2,21
FARSA	P	2,21
SERPINF1	P	2,21
MYO5A	P	2,21
C1orf21	P	2,21
ARNTL	P	2,21
PPIE	P	2,21
RPUSD1	P	2,21
LINC01011	P	2,20
CITED2	P	2,20
ATP6V0C	P	2,20
SOBP	P	2,20
COASY	P	2,20
DIRAS2	P	2,20
GRIK1	P	2,20
PRICKLE1	P	2,20
GTDC1	P	2,20
DCBLD1	P	2,20
SLC39A10	P	2,20
EXTL1	P	2,19
BEAN1	P	2,19
SYNDIG1	P	2,19
FAM217B	P	2,19
LINC01106	P	2,19
CHAF1A	P	2,19
SLC30A3	P	2,19
STX1A	P	2,19
PNOC	P	2,19
LOC645202	P	2,19
AKT3	P	2,19
PLEKHG5	P	2,18
CYTH2	P	2,18
ARHGDIG	P	2,18
SHROOM3	P	2,18
KITLG	P	2,18
PRKCE	P	2,18
SLC22A18	P	2,18
CNTNAP2	P	2,18
SLITRK5	P	2,18
ZNF697	P	2,18
OSCAR	P	2,18
IFITM10	P	2,18

WITHDRAWN
see manuscript DOI for details

POU6F1	P	2,17
BAP1	P	2,17
NRN1	P	2,17
IFFO1	P	2,17
ANXA6	P	2,17
STEAP2	P	2,17
STAMBPL1	P	2,17
FAM89B	P	2,16
DGCR5	P	2,16
JAG2	P	2,16
GAST	P	2,16
RAB15	P	2,16
ATP1A1	P	2,16
TPST2	P	2,16
RORB	P	2,16
DAPK3	P	2,16
CHST10	P	2,16
CCK	P	2,15
FLRT2	P	2,15
R3HDM1	P	2,15
ATG9A	P	2,15
RFTN1	P	2,15
PI4KAP2	P	2,15
IGSF21	P	2,14
KIAA1549L	P	2,14
STMN1	P	2,14
CHN1	P	2,14
SPATS2	P	2,13
OSBPL3	P	2,13
LAMB3	P	2,13
CHD5	P	2,13
ADCY2	P	2,13
PPP1R12C	P	2,13
NAGPA	P	2,13
PACSIN1	P	2,13
KIF5A	P	2,13
EPHX4	P	2,12
PGBD4	P	2,12
ATXN7L3	P	2,12
PTGES2	P	2,12
HECA	P	2,12
GSK3A	P	2,12
GDAP1L1	P	2,12
GABBR2	P	2,12
MATK	P	2,11
RASL11B	P	2,11
PNKD	P	2,11
RALBP1	P	2,11

Gene		Value
LRFN2	P	2,11
CREG2	P	2,11
SAMD12	P	2,11
TTC9	P	2,11
NDUFAF4	P	2,11
TAGLN3	P	2,11
DHRS11	P	2,10
HACL1	P	2,10
WDR60	P	2,10
SBNO1	P	2,10
CDH13	P	2,10
ARHGEF3	P	2,10
GSK3B	P	2,10
PDXP	P	2,10
DPY19L1	P	2,10
FAM13A	P	2,10
DNM1P50	P	2,09
LSM11	P	2,09
MOK	P	2,09
HAUS7	P	2,09
BICDL1	P	2,09
ZNF25	P	2,09
KCNT2	P	2,09
CDH10	P	2,09
GLDC	P	2,09
GRIK3	P	2,09
ZNF500	P	2,09
HOPX	P	2,09
ASXL3	P	2,08
ARHGAP39	P	2,08
KIF17	P	2,08
RUNDC3A	P	2,08
HPS6	P	2,08
NELL1	P	2,08
PGP	P	2,08
CDH18	P	2,08
RFPL1S	P	2,08
MCRS1	P	2,08
KCNH7	P	2,08
DNM1	P	2,07
SREBF2	P	2,07
RTN2	P	2,07
EMC1	P	2,07
RTN4RL2	P	2,07
TUBG1	P	2,07
SHD	P	2,07
TUBGCP5	P	2,06
CSPG4P1Y	P	2,06

WITHDRAWN
see manuscript DOI for details

LINC00473	P	2,06
PLEKHA1	P	2,06
MIAT	P	2,06
CLTA	P	2,06
COL24A1	P	2,06
LRRTM4	P	2,06
GSTT2	P	2,05
NLN	P	2,05
TMEM178B	P	2,05
CHRFAM7A	P	2,05
KCNS3	P	2,05
NPAS1	P	2,05
GRIN2A	P	2,05
HR	P	2,05
FCRLB	P	2,05
EPHB2	P	2,05
MGAT5B	P	2,05
SERPINI1	P	2,05
PINK1	P	2,04
MARCH6	P	2,04
SYT1	P	2,04
MAPK11	P	2,04
NECAB3	P	2,04
UBE2M	P	2,04
LRRC8B	P	2,04
TIMM44	P	2,04
PLXND1	P	2,04
R3HDM2	P	2,04
MDH1	P	2,03
MFSD4A	P	2,03
C1orf216	P	2,03
LINC00294	P	2,03
EFHD2	P	2,03
GOLGA7B	P	2,03
PPP2R2D	P	2,03
PNMA6E	P	2,03
SLC24A2	P	2,03
SRCIN1	P	2,03
AMPH	P	2,03
NEFL	P	2,03
RELL2	P	2,03
IGSF8	P	2,03
MICAL2	P	2,03
PIP5K1B	P	2,02
HSD11B1L	P	2,02
PDE4A	P	2,02
CDC42EP3	P	2,02
TMEM81	P	2,02

WITHDRAWN

see manuscript DOI for details

PER1	P	2,02
ZNF641	P	2,02
PSMD11	P	2,02
NR4A1	P	2,02
KCNB1	P	2,02
TNFSF12	P	2,02
TSHZ2	P	2,02
THEMIS2	P	2,02
KCNJ3	P	2,02
MADD	P	2,01
ABCC8	P	2,01
MIGA1	P	2,01
VWCE	P	2,01
UBE2QL1	P	2,01
KRTAP5-6	P	2,01
NECTIN3	P	2,01
PC	P	2,01
MAP1A	P	2,01
CLN6	P	2,01
SLC29A1	P	2,01
RHCG	P	2,01
CYP2E1	P	2,01
SYP	P	2,00
RTN1	P	2,00
HTR5A	P	2,00
SDK1	P	2,00
DPP10	P	2,00
SLC6A17	P	2,00
Subtype 3		
PRRX1	N	-2,32
RYR1	N	-2,31
FABP5P3	N	-2,27
BCL7C	N	-2,23
FZD1	N	-2,19
RIMKLB	N	-2,16
PCDHGC3	N	-2,15
INSM1	N	-2,15
LRCH2	N	-2,14
TENT5A	N	-2,14
LOC101928370	N	-2,14
DPY19L3	N	-2,12
MOB3A	N	-2,12
CADM1	N	-2,10
IL33	N	-2,09
PARD3B	N	-2,07
NUPR1	N	-2,07
BAG3	N	-2,06
CCDC112	N	-2,06

CCDC80	N	-2,05
MAPK7	N	-2,04
PINLYP	N	-2,03
H2AFJ	N	-2,03
ZMAT1	N	-2,03
NKAIN4	N	-2,01
NFIA	N	-2,01
RAVER2	N	-2,01
AK4	N	-2,00
COX7A1	P	3,34
ISLR	P	3,33
ECM1	P	2,99
MAP1A	P	2,99
RIMKLA	P	2,97
FAM217B	P	2,97
CRTAC1	P	2,93
KIF5A	P	2,91
TPTE2P6	P	2,91
BICDL2	P	2,90
DIRAS1	P	2,90
NIPAL2	P	2,89
LRRC20	P	2,88
CACNG7	P	2,85
HTR1F	P	2,81
MAPK10	P	2,81
RFPL2	P	2,80
CUX2	P	2,80
HES6	P	2,79
ATRNL1	P	2,79
LMTK2	P	2,79
SLC24A2	P	2,78
NECAB3	P	2,76
RORB	P	2,76
KCNJ8	P	2,75
LOC100996385	P	2,75
PREP	P	2,75
CABP1	P	2,75
ZNF365	P	2,75
MAFB	P	2,74
ANXA6	P	2,74
PELI3	P	2,74
RND1	P	2,73
SOHLH1	P	2,72
MIR124-2HG	P	2,72
NUAK1	P	2,71
ACTR1B	P	2,71
RASIP1	P	2,70
SCN1B	P	2,70

IFFO1	P	2,69
IPCEF1	P	2,69
KCNA2	P	2,69
PKNOX2	P	2,69
RPH3A	P	2,69
CTNNAL1	P	2,68
NAT8L	P	2,68
SHD	P	2,68
CAMK2G	P	2,67
SHROOM2	P	2,67
DLGAP1-AS4	P	2,67
ARNTL	P	2,66
FMN1	P	2,65
PAK1	P	2,65
KCNIP3	P	2,65
RAB11FIP5	P	2,65
PKDCC	P	2,64
IGFBP2	P	2,64
STAC2	P	2,63
ZBTB16	P	2,63
PLEKHM2	P	2,63
STRBP	P	2,63
SLCO4A1	P	2,62
MEF2C	P	2,62
AARS	P	2,62
STAMBPL1	P	2,62
SEMA7A	P	2,61
MIAT	P	2,61
CHML	P	2,60
EMX1	P	2,60
BCL2L2	P	2,60
SMYD2	P	2,60
OLFM1	P	2,60
RFPL1	P	2,59
PSMD12	P	2,59
RIMS3	P	2,58
ZNF697	P	2,58
RGS6	P	2,58
TUBG2	P	2,58
MKX	P	2,57
JOSD1	P	2,57
ATP2B2	P	2,57
NFIC	P	2,57
HECW1	P	2,56
HAPLN4	P	2,55
ADPRHL1	P	2,55
FBXO33	P	2,54
HTR2A	P	2,54

WITHDRAWN

see manuscript DOI for details

HR	P	2,54
CDKL1	P	2,54
CAMKK1	P	2,54
C9orf129	P	2,53
NLK	P	2,53
LOC388242	P	2,53
SYNDIG1	P	2,53
PHLDB2	P	2,53
CLSTN1	P	2,53
PRICKLE1	P	2,52
KCNH7	P	2,51
FAM131B	P	2,51
LINC01106	P	2,51
NR2F1-AS1	P	2,50
R3HDM1	P	2,50
NPAS1	P	2,50
SNTB2	P	2,49
VSTM2A	P	2,49
FNDC4	P	2,49
MIR600HG	P	2,48
CHGA	P	2,48
RAB40C	P	2,47
LOC729737	P	2,47
SSBP3	P	2,47
EMC1	P	2,47
EHD3	P	2,47
TSNARE1	P	2,47
CHRM3	P	2,47
LYNX1	P	2,47
EFNA5	P	2,46
FBXW7	P	2,45
RFPL1S	P	2,45
CEMIP	P	2,45
NELL1	P	2,45
RSPO2	P	2,45
RELL2	P	2,44
TIGAR	P	2,44
HYAL3	P	2,43
NT5M	P	2,43
LIMK1	P	2,43
HCN1	P	2,43
LRRC4	P	2,43
BRSK1	P	2,43
PLPP7	P	2,42
NEFL	P	2,42
MYO5A	P	2,42
FOXRED1	P	2,41
COASY	P	2,41

DCLK1	P	2,40
GSTT2B	P	2,40
ID2	P	2,40
ATP1A4	P	2,40
HLF	P	2,40
UNC5A	P	2,40
MGLL	P	2,39
SYTL2	P	2,39
LINC01011	P	2,39
PNKD	P	2,39
BAIAP2L2	P	2,39
SAMD12	P	2,39
NICN1	P	2,39
DNM1	P	2,39
SLC16A7	P	2,39
RFTN1	P	2,39
LYSMD4	P	2,39
ADGRL2	P	2,39
CENPBD1P1	P	2,39
EPHB6	P	2,38
RGS4	P	2,38
SUSD5	P	2,38
CNN1	P	2,38
EXTL1	P	2,38
TIAM2	P	2,38
LINC00672	P	2,37
ARNTL2	P	2,37
TRHDE	P	2,37
PLXDC1	P	2,37
MYO15A	P	2,37
KIF1A	P	2,37
KCNB1	P	2,36
RASD1	P	2,36
ARHGAP32	P	2,36
FHL2	P	2,35
CDH12	P	2,35
TMEM268	P	2,35
BASP1-AS1	P	2,35
C11orf87	P	2,35
BEND5	P	2,35
ST8SIA5	P	2,34
MCRS1	P	2,34
UBE2Q2L	P	2,34
DAPK3	P	2,34
ADRA1B	P	2,34
CIDECP1	P	2,34
AKT3	P	2,34
EHD2	P	2,34

GRIN2A	P	2,34
DENND3	P	2,33
SERPINF1	P	2,33
ELAVL4	P	2,33
ATP1A1	P	2,33
ING1	P	2,33
SCN8A	P	2,33
GABRA1	P	2,32
MADCAM1	P	2,32
LOC440300	P	2,32
IFT57	P	2,32
KHDRBS3	P	2,31
SLC45A4	P	2,31
SLC30A3	P	2,31
PFKP	P	2,31
WDR4	P	2,31
TSHZ3	P	2,31
SCAMP5	P	2,31
FLJ42627	P	2,31
SORL1	P	2,31
PCSK1	P	2,31
GNL1	P	2,31
ZNF541	P	2,31
ACSL6	P	2,30
PSMG3	P	2,30
SPRED2	P	2,30
GUCA1B	P	2,30
KLHL8	P	2,30
CHN1	P	2,29
CPLX1	P	2,29
GSTT2	P	2,29
ATP2B3	P	2,29
LRRC75A	P	2,29
FAM162B	P	2,29
AACS	P	2,29
COL12A1	P	2,29
SCN1A	P	2,28
RAB11FIP3	P	2,28
PCDH7	P	2,28
NUDT4P2	P	2,28
SLC24A4	P	2,28
KIAA1107	P	2,28
DNM1P35	P	2,28
PGBD4	P	2,28
VSNL1	P	2,28
NUDT4	P	2,27
MAGI3	P	2,27
SATB1	P	2,27

WITHDRAWN

see manuscript DOI for details

KIAA1217	P	2,27
PDXP	P	2,27
GPR150	P	2,27
NCALD	P	2,27
PIK3R1	P	2,27
MGAT5B	P	2,26
MPPED1	P	2,26
INSYN1	P	2,26
IER5L	P	2,26
MOB2	P	2,26
KCTD16	P	2,26
OXR1	P	2,25
MDH1	P	2,25
PNMA5	P	2,25
BICDL1	P	2,25
CDK14	P	2,25
SBNO1	P	2,25
CIDEC	P	2,25
ZNF48	P	2,25
EXTL2	P	2,25
CITED2	P	2,25
HIVEP1	P	2,25
TOMM40L	P	2,24
ADAM23	P	2,24
LMO4	P	2,24
SGTB	P	2,24
CORO6	P	2,24
CNTNAP5	P	2,24
STMN1	P	2,24
ARHGAP9	P	2,24
CHAF1A	P	2,23
WIPF2	P	2,23
SCRT1	P	2,23
KCNK12	P	2,23
MDGA2	P	2,23
MTDH	P	2,22
ANKRD24	P	2,22
PLXND1	P	2,22
PRKCB	P	2,22
GSK3A	P	2,22
PI4KA	P	2,22
TNFAIP1	P	2,22
LRFN2	P	2,21
MAP3K13	P	2,21
PRSS3	P	2,21
HECA	P	2,21
CCNI	P	2,21
KCNC2	P	2,21

RAB6B	P	2,21
SV2B	P	2,21
MPP1	P	2,20
OPN3	P	2,20
RTN2	P	2,20
ATG16L2	P	2,19
FAM81A	P	2,19
FAM3C	P	2,19
FLRT2	P	2,19
ST3GAL3	P	2,19
NTN4	P	2,19
SULT4A1	P	2,19
KCNS3	P	2,19
PAK5	P	2,19
CTSA	P	2,18
KLHDC3	P	2,18
NPM2	P	2,18
EPHX4	P	2,18
DCAF6	P	2,18
ETS2	P	2,18
FSD1	P	2,18
GLRX2	P	2,17
CHRFAM7A	P	2,17
RAPGEF2	P	2,17
FARSA	P	2,17
KALRN	P	2,17
PHYHIPL	P	2,17
SPATS2	P	2,17
BRINP2	P	2,17
LYRM4	P	2,17
PTPRT	P	2,17
DIRAS2	P	2,17
NEGR1	P	2,16
CDC6	P	2,16
LINC01102	P	2,16
PMS2	P	2,16
ST3GAL1	P	2,16
LYPD5	P	2,16
OSBPL6	P	2,16
ABCC8	P	2,16
DPP10-AS1	P	2,15
NRN1	P	2,15
TICAM1	P	2,15
UBE2QL1	P	2,15
CCDC3	P	2,15
MICAL2	P	2,15
ZNF25	P	2,15
ATP6V0C	P	2,15

GLS2	P	2,15
KCNA5	P	2,14
LRRC8B	P	2,14
LOC645202	P	2,14
DDHD2	P	2,14
LINC01963	P	2,14
MAPT	P	2,14
PTPA	P	2,13
KCNC3	P	2,13
CCK	P	2,13
KCNT1	P	2,13
TSPAN13	P	2,13
NCKIPSD	P	2,13
OSBPL1A	P	2,13
TGFBR3L	P	2,13
FCRLB	P	2,13
POU6F1	P	2,13
PGK1	P	2,12
PI4KAP2	P	2,12
RAB35	P	2,12
TNFSF12	P	2,12
DBNDD1	P	2,12
ZMAT4	P	2,12
MFSD4A	P	2,12
ARHGEF18	P	2,12
TYRO3	P	2,12
NLN	P	2,12
KIAA1549L	P	2,12
CREG2	P	2,12
STX1A	P	2,12
ADAM22	P	2,11
MBD3	P	2,11
GOT1	P	2,11
SRRM4	P	2,11
TMEM241	P	2,11
ABRACL	P	2,11
NMT1	P	2,11
PPM1L	P	2,11
CADPS2	P	2,11
DUXAP9	P	2,11
F7	P	2,10
R3HDM2	P	2,10
FAM78B	P	2,10
MAGI2	P	2,10
PLEKHA1	P	2,10
MOSMO	P	2,10
LAMB3	P	2,10
RAB15	P	2,10

Gene		Value
NAGPA	P	2,10
C1orf216	P	2,09
KCNAB2	P	2,09
LINC00294	P	2,09
PEX5L	P	2,09
P2RX6	P	2,09
CAMKK2	P	2,09
FAM19A2	P	2,09
RTN4RL2	P	2,09
ACOT4	P	2,09
SYT1	P	2,09
MRTFA	P	2,09
VCX2	P	2,09
CSPG4P1Y	P	2,09
CDKN3	P	2,08
PCDHB14	P	2,08
OIP5-AS1	P	2,08
KCNJ9	P	2,08
FBXW5	P	2,08
LYRM7	P	2,08
TPMT	P	2,08
PACSIN1	P	2,08
CIAPIN1	P	2,08
TTC9B	P	2,08
CHST10	P	2,08
RGPD6	P	2,08
RNF157	P	2,08
NAPEPLD	P	2,08
ZADH2	P	2,07
MELTF	P	2,07
HIVEP2	P	2,07
DNM1P50	P	2,07
RPTOR	P	2,07
NTNG2	P	2,07
COPS7A	P	2,07
PRC1	P	2,07
EEF2	P	2,07
FGF9	P	2,06
VCX	P	2,06
FRMPD4	P	2,06
ABHD8	P	2,06
PPME1	P	2,06
MATK	P	2,06
ACTN4	P	2,06
FAM89B	P	2,06
ADCY2	P	2,06
DEPDC5	P	2,06
SEPT8	P	2,06

LOC100507507	P	2,06
KCNMA1	P	2,06
GRIK3	P	2,06
NRGN	P	2,06
HOPX	P	2,06
NANOS3	P	2,06
SCAI	P	2,05
GRASP	P	2,05
GLIS1	P	2,05
APBB1	P	2,05
MYOM2	P	2,05
PRSS1	P	2,05
GDAP1L1	P	2,05
TMEM121	P	2,05
MAP2K2	P	2,05
ELOVL6	P	2,05
HSPA12A	P	2,05
OPCML	P	2,05
DCBLD1	P	2,04
PHLDA2	P	2,04
STARD13	P	2,04
RAD23A	P	2,04
CDS1	P	2,04
APOL2	P	2,04
ZNF641	P	2,04
TPST2	P	2,04
C6orf106	P	2,04
RALB	P	2,04
PHYH	P	2,03
NDRG3	P	2,03
UBL7	P	2,03
CSGALNACT1	P	2,03
PRRT3	P	2,03
PRSS3P2	P	2,03
TTBK1	P	2,03
NUDC	P	2,03
CBX6	P	2,03
SNX8	P	2,02
CA13	P	2,02
SNAI3-AS1	P	2,02
ARC	P	2,02
KLF9	P	2,02
SCYL1	P	2,02
GRK2	P	2,02
KBTBD11	P	2,02
AMPH	P	2,02
CAMK2N1	P	2,02
FLRT3	P	2,02

Gene	Type	Z-value
EFHD2	P	2,02
DPY19L1	P	2,02
LOC407835	P	2,02
LOC644189	P	2,02
MPP7	P	2,01
ANO5	P	2,01
DNAJC18	P	2,01
INCENP	P	2,01
MPND	P	2,01
MAP3K9	P	2,01
NDUFAF4	P	2,01
PNMA3	P	2,01
OLFM3	P	2,01
CLN6	P	2,01
LNX1	P	2,01
SLC39A4	P	2,01
SAE1	P	2,00
NGEF	P	2,00
ARHGEF11	P	2,00
RASGRF2	P	2,00
CYP26B1	P	2,00
NEFM	P	2,00
SOBP	P	2,00

Table S5: List of all the genes used for enrichment in the ASD subtypes. In total, the enrichment method was applied six times, once for each type of N or P gene, and for the three subtypes. The table also indicates the Z-value for each of the genes.

Gene-symbol	p-value	Class	$r_{observed}$
DYNC1H1	0.055	P	0,377
TRIO	0.074	P	0,363
TSHZ3	**0.021**	P	0,360
MKX	**0.026**	P	0,395
SCN8A	**0.039**	P	0,402
SATB1	**0.007**	P	0,407
PCDH19	0.063	P	0,377
ASXL3	**0.047**	P	0,474
MAP1A	**0.024**	P	0,396
PHF2	0.080	N	-0,459
ZNF462	**0.037**	N	-0,461
AUTS2	0.099	N	-0,467

Table S6: Statistical significance (p-value) of hypothesis-driven genes using surrogate-data for the entire ASD group without performing subtyping. The table indicates the N or P cluster to which genes belong, and the spatial correlation observed between the map of F-statistics and the expression of each gene. Uncorrected p-values, as none of them survived FDR corrections.

Gene-symbol	Class	Z
PDE3A	N	-2,49
ANKRD36BP1	N	-2,49
PCDHGA3	N	-2,37
PDE9A	N	-2,35
NBPF3	N	-2,34
IP6K2	N	-2,32
HIST1H1C	N	-2,32
BTN2A2	N	-2,31
TRDMT1	N	-2,27
PCDHGA2	N	-2,26
SNCAIP	N	-2,26
S100PBP	N	-2,25
AVIL	N	-2,25
LRRC37B	N	-2,23
ALMS1	N	-2,22
CFAP53	N	-2,22
TIAM1	N	-2,21
USP3	N	-2,18
PM20D2	N	-2,17
SINHCAF	N	-2,16
AQR	N	-2,16
ATXN7	N	-2,15
PRKRIP1	N	-2,15
MAPK7	N	-2,15
MTA2	N	-2,14
FLYWCH2	N	-2,13
LOC100288203	N	-2,13
RBBP6	N	-2,13
ZNF432	N	-2,13
TRIM4	N	-2,13
ECD	N	-2,13
PLGLB1	N	-2,12
ZNF510	N	-2,12
ZNF532	N	-2,11
PLK5	N	-2,11
ZNF140	N	-2,11
RPL18	N	-2,11

Gene	Type	Value
RSF1	N	-2,11
ZNF141	N	-2,10
SLC22A5	N	-2,10
PLEKHM1P1	N	-2,09
EPHA7	N	-2,09
PYGO1	N	-2,08
CA12	N	-2,08
IGF2BP2	N	-2,08
NPIPB2	N	-2,08
LATS2	N	-2,07
ABCG1	N	-2,07
AUTS2	N	-2,06
DPY19L3	N	-2,06
TPCN1	N	-2,06
ASPH	N	-2,05
TRIT1	N	-2,05
MCL1	N	-2,05
ATP1B2	N	-2,04
SNRNP48	N	-2,04
ZNF462	N	-2,04
MACROD2	N	-2,03
KLHL3	N	-2,03
PHF2	N	-2,03
PABPC1	N	-2,02
ARID3B	N	-2,02
TMEM216	N	-2,02
LRRC37A2	N	-2,02
H2AFJ	N	-2,01
EML5	N	-2,00
EML4	N	-2,00
YBX1	N	-2,00
DUXAP9	P	3,30
ADRA1B	P	3,29
COL5A2	P	3,19
SLC22A18	P	3,13
FRMPD2B	P	3,10
TSHZ2	P	2,93
SERPINF1	P	2,92
TM6SF1	P	2,92

RASD1	P	2,90
GPR150	P	2,89
TUBA1B	P	2,89
PDK3	P	2,89
PLXNA1	P	2,83
DPP10-AS1	P	2,81
BEND5	P	2,80
HDAC9	P	2,79
KCNJ8	P	2,78
DPP10	P	2,78
KCNC2	P	2,77
DBNDD1	P	2,75
PHLDA2	P	2,74
PKP3	P	2,74
CNN1	P	2,73
BAIAP2L2	P	2,72
NELL1	P	2,70
COL12A1	P	2,69
GAST	P	2,69
ARHGAP25	P	2,68
GPR20	P	2,68
CHRFAM7A	P	2,67
ADPRHL1	P	2,67
FXYD6	P	2,65
TPK1	P	2,64
LOC388242	P	2,64
ATP6V1A	P	2,64
EHD3	P	2,63
RGS6	P	2,63
SUSD5	P	2,62
NANOS3	P	2,62
CRH	P	2,61
ASXL3	P	2,61
PKDCC	P	2,61
GAP43	P	2,61
CHRM3	P	2,60
ARNTL2	P	2,59
CDH13	P	2,59
SHISAL1	P	2,58

MYOM2	P	2,57
ISLR	P	2,57
ARHGDIG	P	2,57
NUDT4P2	P	2,56
OXR1	P	2,55
SLC39A10	P	2,55
GUCA1B	P	2,54
FAM162B	P	2,54
TOX	P	2,54
TUBB4B	P	2,54
C11orf87	P	2,53
WIPF2	P	2,53
MRTFA	P	2,53
FARSA	P	2,53
TUBG2	P	2,53
FXYD6-FXYD2	P	2,53
GRIK3	P	2,52
CHML	P	2,52
CLSTN1	P	2,52
PEX5L	P	2,52
NTNG2	P	2,51
COL23A1	P	2,51
CNTNAP2	P	2,51
SMYD2	P	2,51
CREG2	P	2,51
ADCYAP1	P	2,51
TMEM178B	P	2,51
UBE2Q2L	P	2,51
PTPA	P	2,51
ID2	P	2,50
KIF17	P	2,50
DOC2A	P	2,50
IPCEF1	P	2,49
ZNF365	P	2,48
NEFL	P	2,47
TUBA4A	P	2,47
SNX10	P	2,47
OPN3	P	2,46
DIRAS1	P	2,45

KCNH7	P	2,45
AACS	P	2,44
SYTL2	P	2,44
KHDRBS3	P	2,44
MEF2C	P	2,44
MAP1B	P	2,44
ANXA6	P	2,44
PRICKLE1	P	2,44
CCDC3	P	2,44
NUDT4	P	2,43
HTR2A	P	2,43
PDXP	P	2,43
NLK	P	2,43
MAPT	P	2,42
ZBTB16	P	2,42
LINC02361	P	2,42
C1orf216	P	2,42
KCNT2	P	2,42
ACAT2	P	2,42
ATRNL1	P	2,41
C9orf129	P	2,41
CCK	P	2,41
GPR26	P	2,41
HIVEP1	P	2,41
AKT3	P	2,41
CCN3	P	2,41
IFT57	P	2,41
RGS4	P	2,41
ATP6V1C1	P	2,40
LRFN2	P	2,40
ELAVL4	P	2,40
RIMS3	P	2,40
SHROOM2	P	2,40
GRIN3A	P	2,39
KIF1A	P	2,39
ARID5B	P	2,39
NLN	P	2,39
PPFIBP1	P	2,39
TNFRSF14	P	2,38

OLFM3	P	2,38
LRRTM4	P	2,38
NTN4	P	2,38
AGTPBP1	P	2,38
STMN1	P	2,38
MFGE8	P	2,38
CCDC85A	P	2,38
PI4KA	P	2,37
BASP1-AS1	P	2,37
CABP1	P	2,37
RFPL1S	P	2,37
MPND	P	2,37
HRK	P	2,37
OPTN	P	2,36
DNM1P35	P	2,36
TUBG1	P	2,36
ZDHHC22	P	2,36
TUBA4B	P	2,36
RAB27B	P	2,36
PNMA5	P	2,35
MPPED1	P	2,35
NDUFAF4	P	2,35
PREP	P	2,35
CUX2	P	2,35
HYAL3	P	2,35
AP1M1	P	2,35
GSTT2	P	2,34
SERPINI1	P	2,34
MAPK10	P	2,34
BCL2L2	P	2,34
GULP1	P	2,34
RARA-AS1	P	2,33
RND1	P	2,33
KCNIP3	P	2,33
WSB2	P	2,33
TUBA1A	P	2,33
DENND3	P	2,33
FDX2	P	2,32
B4GALT2	P	2,32

NPAS1	P	2,32
GDAP1L1	P	2,32
LINC00672	P	2,31
AMN1	P	2,31
CRTAC1	P	2,31
NUDT16	P	2,31
DLGAP1-AS4	P	2,31
FAM84A	P	2,31
KLHDC3	P	2,31
LMO4	P	2,30
LINC00294	P	2,30
KCNMA1	P	2,30
MAFB	P	2,30
NR2F1-AS1	P	2,30
CIAPIN1	P	2,30
BMS1P14	P	2,30
APBB1	P	2,29
NCALD	P	2,29
SV2B	P	2,29
GRIK1	P	2,29
LOC440300	P	2,28
BRSK1	P	2,28
TEX30	P	2,28
CTNND2	P	2,28
PFKP	P	2,28
SATB1	P	2,28
GNB4	P	2,28
NICN1	P	2,28
ACOT7	P	2,28
LYNX1	P	2,27
MELTF	P	2,27
NPTX2	P	2,27
RTN4RL2	P	2,27
TUBB2A	P	2,27
HPS1	P	2,27
LSM11	P	2,27
AARS	P	2,27
ADCY2	P	2,27
ST3GAL1	P	2,27

ANO5	P	2,26
PNOC	P	2,26
GDA	P	2,26
SLC25A26	P	2,26
LGALS1	P	2,26
LINC01963	P	2,26
AZIN1	P	2,26
SCN8A	P	2,26
GABRA1	P	2,26
TTC9	P	2,26
NUAK1	P	2,25
SLC30A3	P	2,25
OLFM1	P	2,25
TRHDE	P	2,25
VSNL1	P	2,25
AMPH	P	2,25
MTMR2	P	2,25
HLF	P	2,25
LRRC20	P	2,25
RFPL2	P	2,24
PIN1	P	2,24
GPR22	P	2,24
PHLDB2	P	2,24
PAK1	P	2,24
GABRA3	P	2,24
DCAF6	P	2,24
RASGRF2	P	2,24
EXTL1	P	2,24
FDPS	P	2,23
DUSP3	P	2,23
DPY19L1	P	2,23
FAM47E	P	2,23
SPRN	P	2,23
LMTK2	P	2,23
ECM1	P	2,23
RUNDC3A	P	2,22
KITLG	P	2,22
MAP1A	P	2,22
SLC39A4	P	2,22

DCAF11	P	2,22
MKX	P	2,22
SBF1	P	2,22
RAB11FIP5	P	2,22
CCDC68	P	2,22
C1QL3	P	2,22
TM7SF2	P	2,21
MIGA1	P	2,21
PINK1	P	2,21
C11orf74	P	2,21
SNRNP25	P	2,21
TUBA1C	P	2,21
FRMPD4	P	2,20
VKORC1L1	P	2,20
B3GALT2	P	2,20
EFHD2	P	2,20
FAM81A	P	2,20
EPHB6	P	2,20
KIAA1549	P	2,20
GPX1	P	2,20
TOLLIP	P	2,20
DACT3	P	2,20
SLC25A44	P	2,20
TRUB1	P	2,20
EMX1	P	2,19
TGFBR3L	P	2,19
LOC100996385	P	2,19
LAMB3	P	2,19
NECTIN3	P	2,19
ANKRD9	P	2,19
ID3	P	2,19
LYPD6	P	2,18
SOHLH1	P	2,18
FBXW7	P	2,18
BICDL2	P	2,18
SIDT1	P	2,18
IFITM10	P	2,18
RETREG1	P	2,18
ADGRA1	P	2,17

CALML3	P	2,17
ATXN7L3	P	2,17
MTCH1	P	2,17
TTLL12	P	2,17
DNAJA1	P	2,17
NECAB3	P	2,16
KIFAP3	P	2,16
LINC00839	P	2,16
FMN1	P	2,16
TPI1	P	2,15
VXN	P	2,15
PDZRN4	P	2,15
GSK3B	P	2,15
SPRED2	P	2,15
CHRNA4	P	2,15
SNTB2	P	2,15
CDKL1	P	2,14
SAE1	P	2,14
LRRC8B	P	2,14
JAG2	P	2,14
FLOT2	P	2,14
SSTR1	P	2,14
VPS25	P	2,14
RFPL1	P	2,13
KCNS3	P	2,13
MAGI1	P	2,13
NMT1	P	2,13
SYT17	P	2,13
CHD5	P	2,13
TSTA3	P	2,13
PCDH19	P	2,13
SH3GLB2	P	2,13
DYNC1H1	P	2,13
CIDEC	P	2,13
TNFAIP1	P	2,13
BRINP2	P	2,13
FAM83H	P	2,13
FBXL2	P	2,12
TOMM40L	P	2,12

SKAP2	P	2,12
NRN1	P	2,12
FIBP	P	2,12
HSD11B1L	P	2,12
PSMG3	P	2,12
INKA2	P	2,12
TPST2	P	2,12
HTR1F	P	2,12
RFTN1	P	2,12
UBE2M	P	2,11
TNFAIP8L1	P	2,11
GPATCH4	P	2,11
SOBP	P	2,11
CERS6	P	2,11
RPUSD1	P	2,11
GSTT2B	P	2,10
HK1	P	2,10
TOX4	P	2,10
COMMD9	P	2,10
CDT1	P	2,10
ZNF697	P	2,10
HAPLN4	P	2,10
PDE4D	P	2,10
GSK3A	P	2,10
ELOVL6	P	2,10
HECA	P	2,10
CSGALNACT1	P	2,09
KIF3A	P	2,09
ATP6V0C	P	2,09
COASY	P	2,09
RPH3A	P	2,09
THEMIS2	P	2,09
SYT1	P	2,09
PITPNB	P	2,09
FHL2	P	2,09
LDLR	P	2,09
HSPA12A	P	2,09
EHBP1L1	P	2,09
STRIP1	P	2,08

WITHDRAWN

see manuscript DOI for details

Gene		Value
SPRY4	P	2,08
MADD	P	2,08
CA10	P	2,08
GOT1	P	2,08
PGM2L1	P	2,08
SLITRK5	P	2,08
GABBR2	P	2,07
LOC112694756	P	2,07
RTN1	P	2,07
PRSS3	P	2,07
TIGAR	P	2,07
LOC344967	P	2,07
KIAA1143	P	2,07
RAP2A	P	2,07
SLITRK1	P	2,07
P2RX6	P	2,06
DNM1P50	P	2,06
RAD23A	P	2,06
TRIO	P	2,06
PKNOX2	P	2,06
LGI4	P	2,06
RALBP1	P	2,06
NWD2	P	2,06
MADCAM1	P	2,05
CDH10	P	2,05
SH3BGRL2	P	2,05
CSPG4P1Y	P	2,05
CLSTN2	P	2,05
SHC2	P	2,05
SCN3B	P	2,05
KCNMB4	P	2,05
MRPL4	P	2,05
STX1A	P	2,05
EFNB2	P	2,05
GNG3	P	2,04
TSHZ3	P	2,04
SGTB	P	2,04
DCBLD1	P	2,04
RASL11B	P	2,04

WITHDRAWN
see manuscript DOI for details

Gene		Value
LDHC	P	2,04
SNCA	P	2,04
EMC1	P	2,04
MATK	P	2,04
TMEM65	P	2,04
ELAVL2	P	2,04
DNASE1L1	P	2,04
STK32C	P	2,04
RIMKLA	P	2,04
CHP1	P	2,03
LIMK1	P	2,03
CMAS	P	2,03
PITPNM2	P	2,03
HMGCS1	P	2,03
MFSD3	P	2,03
TMEM121	P	2,03
TRIM59-IFT80	P	2,03
GRK3	P	2,03
NIPAL2	P	2,03
FOXRED2	P	2,03
OSBPL3	P	2,03
TPRG1L	P	2,03
LRRC75A	P	2,02
PACSIN1	P	2,02
LDHA	P	2,02
GABARAPL1	P	2,02
KCNAB2	P	2,02
SYNGR3	P	2,02
MCRIP1	P	2,02
CHN1	P	2,02
UBE2Z	P	2,01
ASL	P	2,01
SULT4A1	P	2,01
KLC1	P	2,01
TMEM59L	P	2,01
PRKCE	P	2,01
SRPK2	P	2,01
VTI1B	P	2,01
CRIP2	P	2,01

PLXND1	P	2,00
PDGFA	P	2,00
CIDECP1	P	2,00
COA7	P	2,00
PIN1P1	P	2,00
FUT9	P	2,00
OPCML	P	2,00
HCN1	P	2,00

Table S7: Complete list of genes used for enrichment without ASD subtyping. When defining brain maps of F-statistics for the whole population of ASD without subtyping, we identified two gene clusters, N and P genes, that were used for the enrichment in figure S7. The table also indicates the Z-value for each of the genes.