NeuroImage xxx (xxxx) 120956

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Sex differences in autism spectrum disorder using class imbalance adjusted functional connectivity

Jong Young Namgung [a, #], Jongmin Mun [b, #], Yeongjun Park [c], Jaeoh Kim [a, d, *], Bo-yong Park [e, f, *]

[a] Department of Data Science, Inha University, Incheon, Republic of Korea
[b] Marshall School of Business, University of Southern California, Los Angeles, United States
[c] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Republic of Korea
[d] Department of Statistics and Data Science, Inha University, Incheon, Republic of Korea
[e] Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea
[f] Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea

ARTICLE INFO

Keywords:
Class imbalance
Gaussian mixture model
Oversampling
Autism spectrum disorder
Functional gradient
Sex difference

ABSTRACT

Autism spectrum disorder (ASD) is an atypical neurodevelopmental condition with a diagnostic ratio largely differing between male and female participants. Due to the sex imbalance in participants with ASD, we lack an understanding of the differences in connectome organization of the brain between male and female participants with ASD. In this study, we matched the sex ratio using a Gaussian mixture model-based oversampling technique and investigated the differences in functional connectivity between male and female participants with ASD using low-dimensional principal gradients. Between-group comparisons of the gradient values revealed significant interaction effects of sex in the sensorimotor, attention, and default mode networks. The sex-related differences in the gradients were highly associated with higher-order cognitive control processes. Transcriptomic association analysis provided potential biological underpinnings, specifying gene enrichment in the cortex, thalamus, and striatum during development. Finally, the principal gradients were differentially associated with symptom severity of ASD between sexes, highlighting significant effects in female participants with ASD. Our work proposed an oversampling method to mitigate sex imbalance in ASD and observed significant sex-related differences in functional connectome organization. The findings may advance our knowledge about the sex heterogeneity in large-scale brain networks in ASD.

1. Introduction

Autism spectrum disorder (ASD) is a common development psychiatric condition characterized by impaired social interaction skills and restricted/repetitive behaviors (Mottron et al., 2006). Notably, male individuals are diagnosed with ASD more frequently than female individuals, with the observed ratio ranging from approximately 4:1 to 9:1 (Werling and Geschwind, 2013). This difference may be because female individuals with ASD often show fewer occurrences of externalizing and repeated/restricted behavioral problems than that of male patients (Baron-Cohen et al., 2003; Baron-Cohen and Wheelwright, 2004, 2003) and are protected from ASD-related gene variants (Levy et al., 2011; Sanders et al., 2011; Sebat et al., 2007). Early neuroimaging studies investigated differences in brain structure and function between male and female individuals with ASD using magnetic resonance imaging (MRI) (Li et al., 2023; Walsh et al., 2021). Some researchers have observed significant interactions between diagnosis and sex based on cortical thickness reductions and altered functional activations (Beacher et al., 2012; Bedford et al., 2019). On the other hand, others have found differences in the neuronal streamline distributions between male and female individuals with ASD (Nordahl et al., 2015). A seed-based analysis unveiled distinct relations of functional connectivity and ASD traits between sexes (Jung et al., 2015). The studies collectively suggest that male and female individuals with ASD show distinct susceptibility to the condition. However, we lack statistical robustness for sex differences in ASD because most study participants were male individuals, leading to the class imbalance problem, which refers to a large imbalance of sample size between two classes in the statistics literature.

A notable study addressing the class imbalance in hypothesis testing uses an undersampling approach, which retains a representative subset

* Corresponding authors.
E-mail addresses: jaeoh.k@inha.ac.kr (J. Kim), boyongpark@korea.ac.kr (B.-y. Park).
# Equal contribution

https://doi.org/10.1016/j.neuroimage.2024.120956
Received 25 March 2024; Received in revised form 17 November 2024; Accepted 22 November 2024
1053-8119/

of samples from the majority class (Chen et al., 2013). By independently repeating the undersampling and test statistic calculation steps and then taking a weighted average of these statistics, we can fully utilize the data and achieve strong statistical power, even under the condition of class imbalance. However, undersampling inevitably introduces additional randomness into the testing procedure, which can affect the Type I error probability. Chen et al., 2013 demonstrated that, despite this randomness, their testing procedure provides asymptotically accurate Type I error control in the regime where the sample size approaches infinity.

It is worth noting that undersampling approaches are part of the broader category of resampling methods, which are well-established in the machine learning community for handling imbalanced data in binary classification tasks (Krawczyk, 2016; López et al., 2013). Another common resampling method used to address data imbalance is oversampling, which generates synthetic samples resembling the original minority class. To ensure these synthetic samples accurately capture the essential characteristics of the minority class, it is crucial to properly estimate its distribution. Thus, oversampling techniques are closely related to generative models, which aim to learn data distributions and generate samples indistinguishable from the original data. Generative models are increasingly applied in the physical sciences, particularly in scenarios where data are scarce or costly. For example, IBM pioneered the use of deep learning-based generative models in physics research (IBM Research, 2021), and many applications of generative models in the physical sciences now use deep learning-based models to approximate a wide range of distributions reliably. In the biological sciences, generative adversarial networks (Maziarka et al., 2020), variational autoencoders (Jin et al., 2018), and reinforcement learning (You et al., 2018) are widely used.

While deep learning-based generative models are useful, they require large amounts of training data to achieve reliable performance, making them less suitable for some settings with limited data sizes. One approach to mitigate this requirement is to exploit the underlying physical structures in the data, such as molecular graph structures (Guo et al., 2022). However, this approach becomes challenging for highly complex data. In contrast, the Gaussian mixture model, a long-standing and flexible distribution estimation and generative modeling algorithm, is known to approximate distributions well even with relatively small sample sizes, as long as the data dimensionality is moderate and the true distribution aligns reasonably with a Gaussian mixture (Mun et al., 2025). In the asymptotic regime, where the sample size approaches infinity, the Gaussian mixture model can approximate all sufficiently smooth densities (Goodfellow et al., 2016), making it suitable for augmenting the minority class sample size. In this context, we extend the approach of Chen et al., 2013 by transitioning from undersampling to oversampling for hypothesis testing under class imbalance using the Gaussian mixture model.

Functional brain organization can be assessed based on the functional connectivity analysis, which defines the inter-regional relations of time series (Bullmore and Sporns, 2009). A recent study introduced dimensionality reduction techniques to identify the principal low-dimensional representations of the connectivity (i.e., gradients) (Margulies et al., 2016). The authors found that the principal gradients exhibited an important cortical hierarchy differentiating low-level sensory and higher-order default mode regions. This gradient approach has been widely adopted in studies to identify imaging markers for psychiatric conditions (Benkarim et al., 2021; Hettwer et al., 2022; Park et al., 2021) and to study structure-function coupling (Valk et al., 2022; Vázquez-Rodríguez et al., 2019). In particular, significant shifts in the gradient patterns were observed in individuals with ASD relative to typically developing controls (Hong et al., 2019). Therefore, we hypothesized that this gradient approach might be appropriate for investigating sex differences in individuals with ASD.

In this study, we investigated differences in the functional connectome organization of the brain between male and female participants with ASD using the connectome gradient approach and class imbalance adjustment methods. Our study may provide insights into the understanding of sex differences in ASD, as well as the oversampling approach for imbalanced datasets.

2. Methods

2.1. Study participants and imaging data

We obtained T1-weighted structural MRI and resting-state functional MRI (rs-fMRI) data of 507 participants with ASD (mean ± SD age = 17.08 ± 8.50 years; 12.22 % female) and 553 typically developing controls (mean ± SD age = 17.10 ± 7.74 years; 17.35 % female) from Autism Brain Imaging Data Exchange initiative (ABIDE I; https://fcon_1000.projects.nitrc.org/indi/abide) (Di Martino et al., 2014). Data were obtained from 20 independent sites, and all participants were diagnosed with ASD using gold standard diagnostic methods, specifically the Autism Diagnostic Observation Schedule (Lord et al., 2000) and/or Autism Diagnostic Interview-Revised (Lord et al., 1994). ABIDE data collection adhered to the local Institutional Review Board guidelines. Following the HIPAA guidelines and 1000 Functional Connectomes Project/INDI protocols, all ABIDE datasets were fully anonymized, with no protected health information included. Detailed demographic information and MRI data acquisition parameters are summarized in **Supplementary Table 1**.

2.2. Data preprocessing

We preprocessed T1-weighted and rs-fMRI data using micapipe (https://github.com/MICA-MNI/micapipe) (Cruces et al., 2022), an integrated software of FreeSurfer, FSL, AFNI, MRtrix3, ANTs, and workbench (Avants et al., 2011; Cox, 1996; Fischl, 2012; Jenkinson et al., 2012; Marcus et al., 2011; Tournier et al., 2019). The T1-weighted MR images underwent de-oblique and reorientation to the left-posterior-inferior orientation. Intensity nonuniformity correction (Tustison et al., 2010), intensity normalization, and non-brain tissue removal were performed. The subcortical structures were segmented using FSL's FIRST (Patenaude et al., 2011), and cortical surfaces were generated using FreeSurfer (Dale et al., 1999; Fischl, 2012; Fischl et al., 2002, 1999a, 1999b; Ségonne et al., 2007). For the rs-fMRI data, a series of preprocessing steps were implemented. The initial five volumes were discarded to address magnetic field saturation, and the images were reoriented. Head motion correction was performed by aligning all volumes to the mean volume across the temporal domain. Nuisance variables were eliminated using FMRIB's independent component analysis-based X-noiseifier (Salimi-Khorshidi et al., 2014), and motion outliers and spikes were regressed out. The processed volumetric data were registered on the Conte69 template surface with 164k vertices using MSMAll (Robinson et al., 2014) and downsampled to a 32k vertex mesh. The time series data were spatially smoothed with a full width at a half maximum of 10 mm.

2.3. Functional connectivity gradients

We constructed the functional connectivity matrix by calculating linear correlations of time series between different brain regions defined using the Schaefer atlas with 300 parcels (Schaefer et al., 2018), and the correlation coefficients were Fisher's r-to-z transformed. The top 10% of elements per row were retained from the connectivity matrix, and an affinity matrix was constructed using a normalized angle kernel. To reduce data dimensionality, we applied a nonlinear dimensionality reduction technique (diffusion map embedding), which captures the intrinsic geometry of complex data by modeling connectivity

patterns (Coifman and Lafon, 2006). This technique is robust to noise and computationally efficient, offering advantages over other nonlinear dimensionality reduction methods (Tenenbaum et al., 2000; Von Luxburg, 2007). Specifically, diffusion map embedding first constructs a graph, where nodes represent data points and edges represent similarities. The graph is then transformed into a transition probability matrix that defines a Markov chain, simulating a diffusion process over the data points. A key concept in diffusion maps is diffusion time, corresponding to the number of steps in this Markov chain, which reveals structures at different scales as time increases. The resulting embedding effectively captures both local and global structures within the data, providing a meaningful low-dimensional representation, particularly suited for capturing nonlinear relationships in complex datasets such as neuroimaging data. The group representative gradients were generated using the averaged functional connectivity matrix across individuals, and individual gradients were aligned to the group template gradients using Procrustes rotation (Langs et al., 2015; Vos de Wael et al., 2020). Given the importance of demonstrating feature reliability (Zuo et al., 2019), we validated the reproducibility of the functional gradients using a test-retest dataset from the Human Connectome Project (HCP) (Van Essen et al., 2013). We obtained rs-fMRI data from 44 participants (mean ± SD age = 30.40 ± 3.29 years; 71.43 % female), which had been preprocessed using the HCP minimal preprocessing pipelines (Glasser et al., 2013). Functional gradients were generated, and linear correlations of the gradients between the test and retest datasets were calculated.

2.4. Oversampling of the gradients

To prevent false discoveries or loss of power stemming from small sample size and class imbalance, we adjusted the sex imbalance in the dataset with the oversampling technique. We first estimated the gradient distribution of the female group using a Gaussian mixture model to avoid unnecessary assumptions on the distribution of gradients in the female group. Subsequently, we generated synthetic samples from the estimated distributions. The Gaussian mixture model is a flexible method that can approximate a large class of continuous distributions. It assumes that the density $f(x)$ of the group as a linear combination of J Gaussian density functions, as follows:

$$f(x) = \sum_{j=1}^{J} \pi_j \mathcal{N}\left(x \middle| \mu_j, \sigma_j^2\right), \quad (1)$$

where $\mathcal{N}\left(x \middle| \mu_j, \sigma_j\right)$ denotes a Gaussian density with mean μ_j and variance σ_j^2, and π_j mixing coefficients satisfying the constraints of $0 \le \pi_j \le 1$ and $\sum_{j=1}^{J} \pi_j = 1$. These mixing coefficients indicate the contribution of each Gaussian density to the total distribution. For a fixed value of J, the parameters $\{\mu_j, \sigma_j,\ \pi_j\}_{j=1}^{J}$ are obtained by maximizing the log-likelihood using the expectation-maximization algorithm (Dempster et al., 1977). The number of components J is determined using a standard measure, namely the Bayesian information criterion:

$$k \ln(n) - 2 \ln\left(\hat{L}\right), \quad (2)$$

where k is the number of parameters, n the sample size, and $\hat{L}$ the maximized likelihood value. After summarizing the region-wise gradient values according to seven or seventeen intrinsic functional networks (Yeo et al., 2011), we conducted oversampling of the female gradients after controlling for age and site with desired imbalance ratios of {0.6, 0.7, 0.8, 0.9, 1.0} for both ASD and control groups. The Gaussian mixture model cannot estimate the distribution when the dimension of the vector is larger than the sample size. Thus, we reduced the data points to seven/seventeen networks as the sample size of female participants was small. To assess the reliability of the oversampling procedure, we compared the distribution of the actual and synthesized gradient values for each brain network. Comparison between the two distributions was quantified by calculating the Wasserstein distance (Vallender, 2006). These oversampling and evaluation processes were equally applied across the functional gradients oversampled with different ratios. To assess site-specific effects, we compared the distribution of the oversampled gradients for two representative sites that contained a relatively large number of female samples (number of female samples [ASD:control = 10:26 at New York University Langone Medical Center; 8:15 at University of Michigan]; **Supplementary Table 1**). For the Gaussian mixture model and synthetic data generation, we used the Mclust package (version 6.0.0) in R (version 4.2.2), which employs the EM algorithm to estimate the distribution with a spherical covariance structure and equal volume option, optimizing efficiency in small sample size settings.

2.5. Sex-condition interaction effects

To evaluate sex-related differences in functional gradients between participants with ASD and typically developing controls, we assessed the interaction effects of sex and group using a two-way analysis of variance (ANOVA) model with interaction effects as follows:

$$Y = \beta_0 + \beta_1 \cdot sex + \beta_2 \cdot group + \beta_3 \cdot sex \cdot group$$

$$contrast = sex * (group = ASD) - sex * (group = control),$$

where Y represents the functional gradient, sex denotes male or female, and $group$ represents either ASD or typically developing controls. In particular, rather than testing whether all interaction effects are zero, our null hypothesis is formulated as a weaker contrast hypothesis, as follows:

$$H_0 : \gamma_{1,1} - \gamma_{1,2} = \gamma_{2,1} - \gamma_{2,2},$$

where each term represents an interaction effect, the first index indicates sex (1: male, 2: female), and the second index represents group (1: treatment, 2: control). This null hypothesis represents a situation where the treatment effect does not differ between male and female participants. We assessed the interaction effects using the BrainStat toolbox (https://github.com/MICA-MNI/BrainStat) (Larivière et al., 2023). The inference was conducted across different desired oversampling ratios from 0.6:1 to 1:1. Multiple comparisons across brain networks were corrected using a false discovery rate (FDR) (Benjamini and Hochberg, 1995). Under our weak null hypothesis, the exchangeability of data points, which is a crucial assumption for the permutation test, did not meet the requirement. Therefore, we did not perform the permutation tests.

2.6. Cognitive decoding and gene enrichment analyses

We added biological underpinnings to the interaction effects of sex and disease state by performing cognitive decoding and gene enrichment analyses. These analyses were performed using the t-statistic map of interaction effects inferred from the oversampled data with an equal ratio (i.e., male:female = 1:1). We first calculated a linear correlation between the interaction and cognition maps obtained from the neuromaps toolbox (https://github.com/netneurolab/neuromaps) (Markello et al., 2022). The cognition map was the first principal component of the meta-analytic maps derived from the NeuroSynth tool (https://neurosynth.org/). We also performed cognitive decoding using the NeuroSynth by calculating correlation coefficients corresponding to the maps of multiple cognitive terms (Rubin et al., 2017; Yarkoni et al., 2011). Additionally, we conducted gene enrichment analysis using the abagen toolbox (https://github.com/rmarkello/abagen) (Arnatkevičiūtė et al., 2019; Markello et al., 2021). The toolbox

processed microarray expression data gathered from six *post-mortem* human brains sourced from the Allen Human Brain Atlas (Hawrylycz et al., 2012). We allocated the expression data to specific parcels, with any missing data being estimated by assigning the expression value of the nearest tissue. Since four of six donors provided expression maps only for the left hemisphere, we applied mirroring techniques to replicate the data to the right hemisphere. The multiple probes were adjusted based on the differential stability method, which calculates the Spearman correlation between expression data for each pair of donors, and retains the probe with the highest correlation. To mitigate potential discrepancies in expression values among the donors, the expression data were normalized using a scaled robust sigmoid function. Finally, we averaged the expression data from all six donors and summarized according to the seven networks (Yeo et al., 2011). We computed linear correlations between the t-statistic interaction map and each gene expression data. We then performed developmental enrichment analysis (http://genetics.wustl.edu/jdlab/csea-tool-2/) (Dougherty et al., 2010) using genes that exhibited an absolute correlation coefficient greater than 0.7 and passed for FDR < 0.05.

2.7. Associations with symptom severity

We performed a regression analysis to associate functional gradients, which showed significant sex-related interaction effects with the symptom severity in ASD, indexed by the Social Responsiveness Scale (SRS) score (Constantino, 2013). As not all participants provided the SRS score, we constructed the regression model using data from 176 participants. We also performed this analysis for male and female participants separately. The performance of the model was assessed using adjusted R^2.

2.8. Reproducibility validation

We evaluated the reproducibility of the sex-condition interaction effects using an independent ABIDE-II dataset. Resting-state functional MRI data were obtained from 448 participants with ASD (mean ± SD age = 14.90 ± 8.84 years; 15.62 % female) and 532 typically developing controls (mean ± SD age = 14.89 ± 9.37 years; 31.02 % female), and the same preprocessing pipeline was applied. Functional connectivity gradients were estimated, sex imbalance was adjusted using a Gaussian mixture model, and sex-condition interaction effects were analyzed.

3. Results

3.1. Oversampling of functional connectivity gradients

We generated the first functional gradient using the dimensionality reduction technique (Coifman and Lafon, 2006; Vos de Wael et al., 2020) and summarized it according to seven brain networks (Thomas Yeo et al., 2011). The gradient clearly differentiated the sensory regions from default mode regions (Fig. 1**a**). To test the feature reliability, we calculated linear correlations of the gradients between the test and retest datasets and observed high similarity (mean ± SD r = 0.92 ± 0.05, $p < 0.001$; **Supplementary Fig. 1**), demonstrating the reproducibility of this feature. To adjust for the sex imbalance problem of the ABIDE dataset (Di Martino et al., 2014), we oversampled functional gradients of female participants using the Gaussian mixture model for autism and control groups, which showed spatial patterns similar to that of the actual data (Fig. 1**a**). We quantitatively compared the distribution of oversampled gradient values with the actual distribution using the Wasserstein distance, which measures the difference between the shapes of two different histograms (Vallender, 2006) (Fig. 1**b**). The distance was overall higher in the ASD group than that in the control group, and the limbic network was higher than the other networks. These results were consistent across different oversampling ratios (**Supplementary Fig. 2**). On performing the oversampling using 17 brain networks (Yeo et al., 2011), the results were consistent (**Supplementary Fig. 3**). The spatial pattern was slightly different across sites, but the trends were consistent overall (**Supplementary Fig. 4**).

3.2. Interaction between sex and condition

We assessed the interaction effect between sex and group using the BrainStat toolbox (https://github.com/MICA-MNI/BrainStat) (Larivière et al., 2023). The somatomotor (t = 2.63, FDR = 0.029), dorsal attention (t = 2.73, FDR = 0.029), and default mode networks

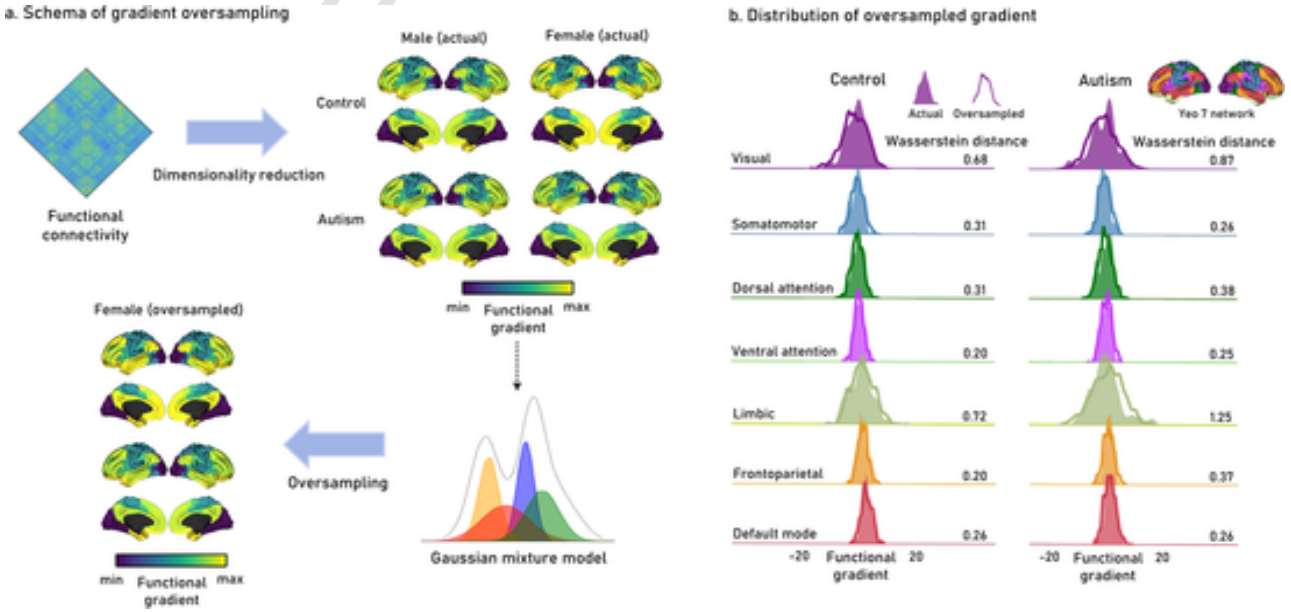

Fig. 1. Oversampling of functional gradients using Gaussian mixture model. (a) Functional gradients generated from the functional connectivity data using dimensionality reduction techniques. We oversampled the gradients of female participants using a Gaussian mixture model. (b) The distributions of actual and oversampled gradient values for each brain network are compared. We assessed the differences between the gradient distributions using the Wasserstein distance.

(t = -5.63, FDR < 0.001) showed significant interaction effects, and the effect sizes increased when the number of male and female participants was equally distributed (Fig. 2**a**). Specifically, the gradient values decreased in female participants with ASD compared to that in controls in the default mode network, while male participants with ASD showed slightly increased values (Fig. 2**b**). Conversely, gradient values of the sensorimotor and attention networks showed opposite patterns (Fig. 2**b**). These patterns were replicated when we assessed the interaction effect using 17 network-based functional gradients (**Supplementary Fig. 5**). Notably, these findings suggest that there exist different patterns of functional connectome organization across sexes in ASD, particularly in the higher-order default mode regions.

3.3. Biological underpinnings

To provide the underlying cognitive associations of sex-related differences across conditions, we associated the interaction effect map with the NeuroSynth-derived cognition map (Markello et al., 2022). We found a significant association (r = 0.771, p = 0.042), indicating that the sex-related differences in functional connectome alterations in ASD relative to typically developing controls might follow the sensory-transmodal cortical hierarchical patterns (Fig. 3**a**). On decoding the interaction map according to cognitive terms, we found high correlations with the higher-order cognitive control-related terms, such as autobiographical, self-referential, theory mind, and moral. Additionally, we performed a transcriptomic association analysis (Arnatkevičiūtė et al., 2019; Markello et al., 2021), and the genes that were strongly correlated with the interaction map (**Supplementary Data**) showed significant enrichments on the cells in the cortex, striatum, and thalamus during development (Fig. 3**b**).

3.4. Clinical score association

We conducted a multiple linear regression analysis to investigate whether the functional gradients were differentially related to symptom severity of ASD between male and female participants (Fig. 4**a**). We found that functional gradients of somatomotor, dorsal attention, and default mode regions were significantly associated with symptom severity only in the female group (adjusted R^2 = 0.270 and p = 0.039; Fig. 4**b**).

3.5. Reproducibility of the findings

We evaluated the reproducibility of our findings using an independent dataset. The spatial patterns of the oversampled functional gradients closely resembled those of the actual data, and the distribution of gradient values was also consistent with the main results (**Supplementary Fig. 6a**). When assessing the interaction effects between sex and condition, we observed largely similar patterns, although the effect size was reduced in the replication dataset (Pearson correlation of the t-statistics between the ABIDE-I and ABIDE-II datasets: r = 0.82, p = 0.024; **Supplementary Fig. 6b**).

4. Discussion

Identification of sex-related brain connectome disorganization in individuals with ASD is crucial for understanding the natural characteristics of ASD. As male individuals are more frequently diagnosed with ASD, the class imbalance problem needs to be solved to identify statistically reliable sex-related differences in ASD. In this study, we investigated the differences in functional connectome organization of the brain between male and female participants with ASD after adjusting the sex imbalance problem using a Gaussian mixture model. The spatial patterns of synthesized functional connectivity were largely similar to the actual data, and the robustness was demonstrated through multiple sensitivity analyses. We assessed the interaction effect between sex and condition and found that sex-related differences in functional connectome organization were associated with cortical hierarchy in patients with ASD. The cognitive decoding and gene enrichment analyses suggested the potential biological underpinnings of the interaction effects.

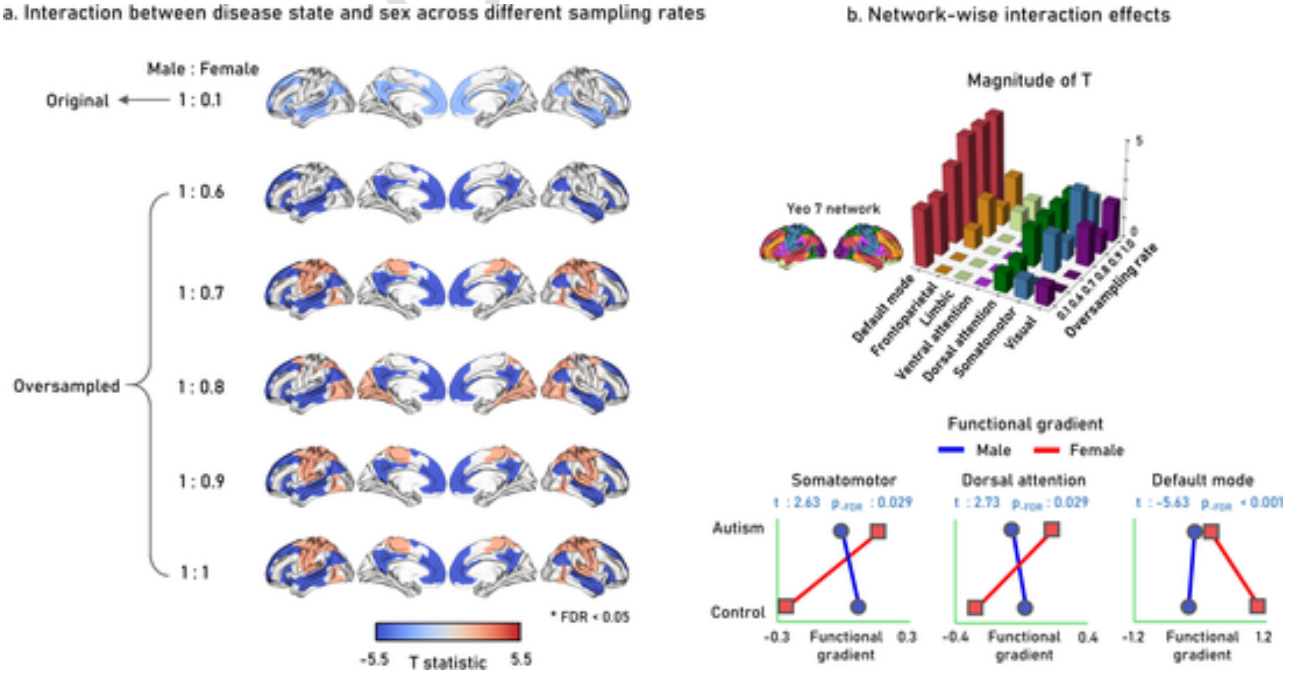

Fig. 2. Interaction effects of functional gradients between sex and condition. (a) T-statistics of the networks showing significant interaction effects are shown on the brain surfaces across different oversampling ratios. (b) The magnitude of the t-statistics of each network is represented with bar plots (top). The interaction plots for the networks that showed significant interaction effects are visualized.

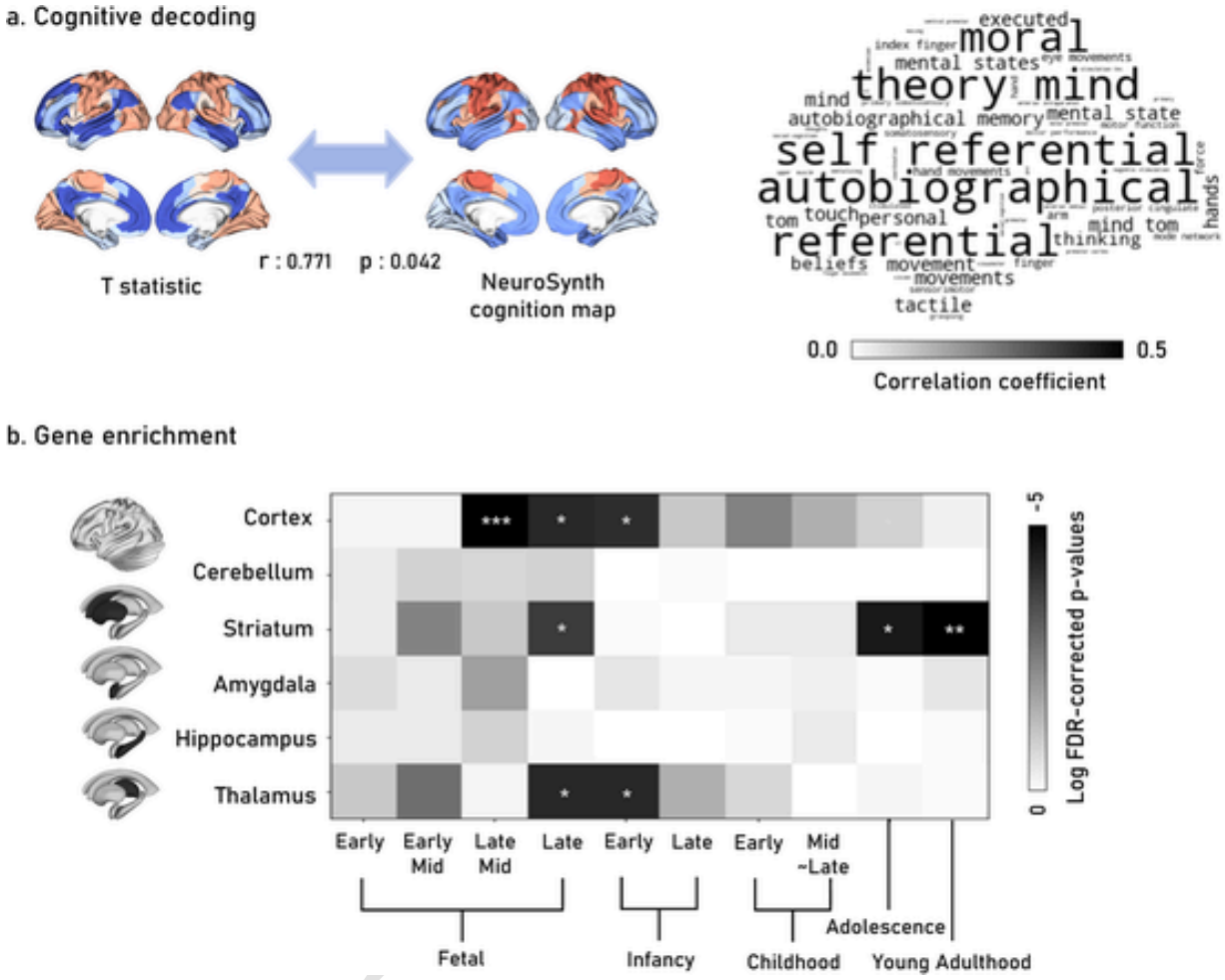

Fig. 3. Cognitive decoding and transcriptomic association analyses. (a) The interaction map is first associated with the first principal component of the NeuroSynth cognition map (left) and then decoded with cognitive terms (right). (b) Genes that were spatially correlated with the interaction map are identified and used for the developmental enrichment analysis. ***FDR < 0.001, ** FDR < 0.01, *FDR < 0.05.

Moreover, symptom severity prediction revealed differential effects between the sexes.

As class imbalance severely impacts the performance of statistical tests, we applied a distribution estimation-based oversampling technique to alleviate the class imbalance while retaining the overall shape of data distribution. This approach worked well for our study since the data of the female group had a well-behaved distribution shape, adequately smooth to be approximated by the Gaussian mixture model. Although the Gaussian mixture model is a flexible method that can approximate large classes of continuous distribution, it cannot approximate distributions with irregular shapes. In those situations, non-parametric distribution estimation methods can be used. However, it should be noted that non-parametric methods require more samples than parametric methods, such as the Gaussian mixture model. This was the rationale for adopting the parametric approach. As with under-sampling-based testing (Chen et al., 2013), oversampling may add randomness to the testing process. Therefore, using an expanded dataset that combines original and synthetic samples may affect the Type I error control. Test statistics typically rely on asymptotic normality, even without oversampling, as real data are rarely normally distributed. Under mild smoothness conditions, Gaussian mixture models with the EM algorithm converge to the true distribution as the sample size increases (Goodfellow et al., 2016). In this asymptotic regime, the expanded dataset approximates the true minority class distribution despite the additional randomness, allowing our test statistics to retain the same asymptotic null distribution and thus achieve asymptotically valid Type I error control.

We adopted the gradient technique, which is useful for assessing the connectome organization of the brain (Margulies et al., 2016). A previous study noted that feature reliability is crucial in neuroscience studies, especially for fMRI, as fMRI-derived features have generally shown lower reliability than morphological features, leading to reduced reproducibility across studies (Zuo et al., 2019). Thus, in this study, we demonstrated the reproducibility of the functional gradients by validating their test-retest reliability. Leveraging the nonlinear dimensionality reduction technique, we generated the principal gradient of functional connectivity, and the gradient showed a clear cortical hierarchy differentiating low-level sensory and higher-order default mode regions. This axis is particularly important for ASD connectopathy because the sensory and default mode networks are the core regions showing structural or functional network anomalies in ASD (Hadjikhani et al., 2004; Hong et al., 2019; Jung et al., 2015; Park et al., 2021). One study observed significant shifts in functional gradient values in the default mode regions, exhibiting reduced cortical hubs in the network (Hong et al., 2019). Our previous study showed that alterations in structural gradients in individuals with ASD were linked to atypical neuronal microcircuit functions of inhibition/excitation balance (Park et al., 2021). Thus, those studies provide a rationale for investigating alterations in connec-

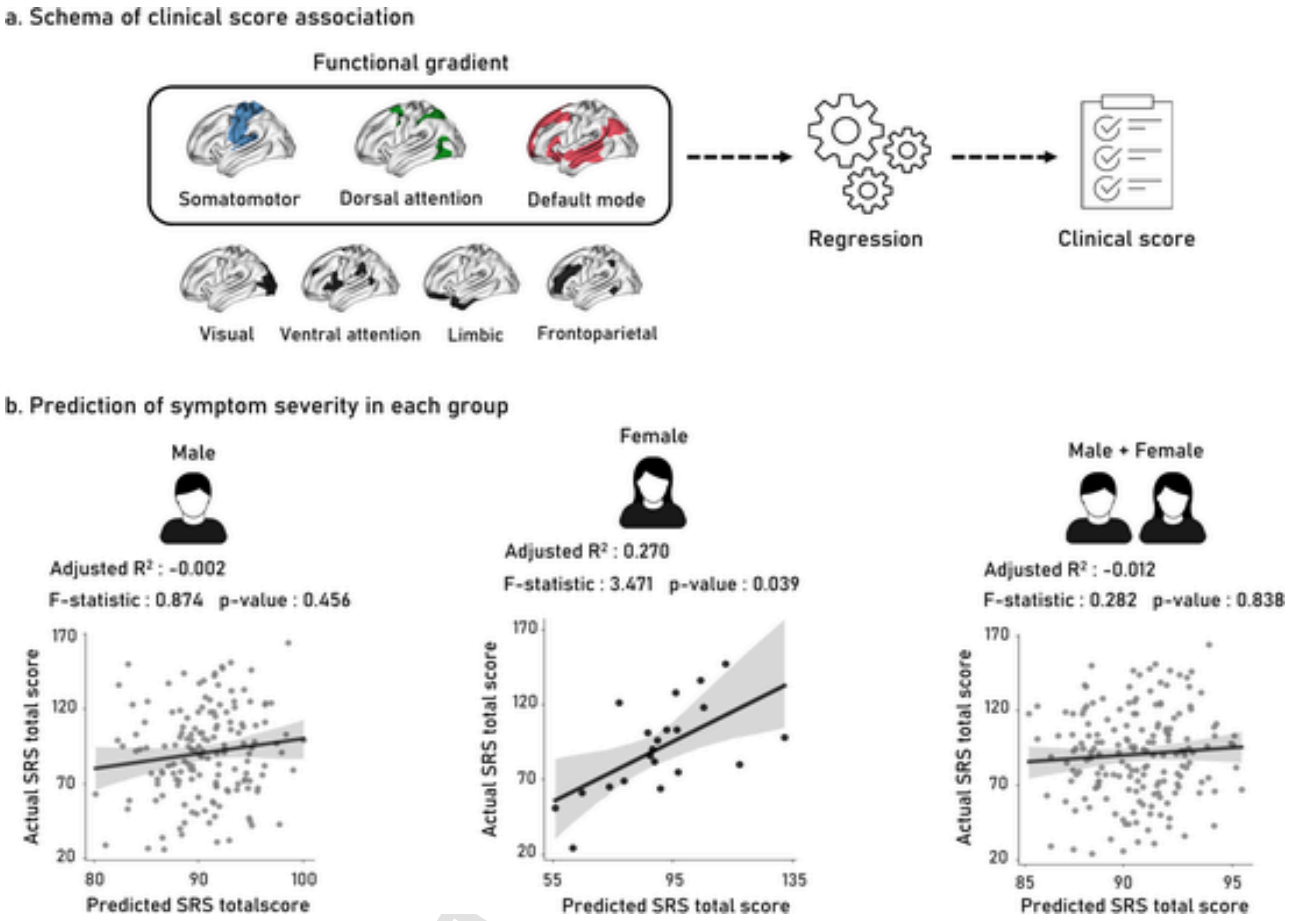

Fig. 4. Association between functional gradients and symptom severity. (a) Schema of the clinical score association that utilizes functional gradients showing significant sex-condition interaction effects (black box). **(b)** We conducted linear regression for male, female, and total participants with ASD, and the actual and predicted scores are reported with scatter plots. *Abbreviation:* SRS, Social Responsiveness Scale.

tome gradients in individuals with ASD. By comparing the functional gradients between male and female individuals with ASD, we observed differences in gradient values in the motor, attention, and default mode networks. These effects were small when the samples were biased toward the male individuals and more evident when the class imbalance became reduced. The results indicate that our oversampling approach could increase the statistical power, which might not be observed in the class imbalanced states. We found that female participants with ASD showed large shifts in the gradient values of the low-level sensory network towards higher-order networks, whereas the gradients of the default mode network were shifted towards low-level sensory areas, making it difficult to delineate the sensory-transmodal hierarchy. Males with ASD showed patterns opposite to females with ASD, but the effects were smaller than those in females. The findings indicate that cortical hierarchical alterations are significantly observed in both male and female participants with ASD, and although they are less frequently observed, the latter may be more susceptible to brain network disorganization. Although the biological mechanisms of change in sex-specific functional brain connectivity require further investigation, our findings indicate underlying connectopathy related to the sensory-transmodal hierarchy in ASD.

Cognitive decoding and gene enrichment analysis provided further insights into sex-related differences in functional connectome distortions in ASD. First, association analysis with the Neurosynth cognition map provided quantitative evidence that the sex-related differences followed cortical hierarchy. We could identify obvious relations of the interaction effects with higher-order cognitive control-related terms, which complement prior work that the default mode network encompasses subsystems responsible for engaging in introspective and adaptive cognitive activities that humans naturally and intentionally participate in during their daily lives (Andrews-Hanna, 2012). Thus, the sex-related differences in the default mode network might be related to an atypical cognitive decoding process. The transcriptomic association analysis revealed that the gene expressions in the striatum, thalamus, and cortical cells were related to the sex-related differences in functional connectivity in ASD. Previous studies showed an increase in the growth rate of striatal structures in ASD (Langen et al., 2014), a reduction in thalamic volume, and dysregulation of thalamocortical networks (Hardan et al., 2006; Tomasi and Volkow, 2019; Tsatsanis et al., 2003). The findings of those studies collectively suggest the heterogeneity of gene expressions in ASD between male and female individuals and provide a direction for future studies: sex heterogeneity should be considered for understanding the multiscale traits of ASD. Moreover, we found a more variable functional network reconfiguration in female participants with ASD, and the clinical score prediction was significant only in the female group. These results indicate that female individuals with ASD might have a stronger relationship between behavior symptoms and network dysfunction.

In summary, we suggested an oversampling method to mitigate sex imbalance in the ASD dataset and observed significant sex-related differences in functional gradients in individuals with ASD. Cognitive decoding and transcriptomic analyses provided potential links between macroscale network disorganization and microscale gene expressions, which may advance our insights into sex heterogeneity in ASD. Finally,

clinical score association analysis indicated the importance of female sex in understanding the epidemiology of ASD. Our systematic analyses investigating sex-related differences in functional brain networks may provide insights for understanding the heterogeneity of ASD.

Data availability

The imaging and phenotypic data were provided, in part, by the Autism Brain Imaging Data Exchange initiative (ABIDE I; https://fcon_1000.projects.nitrc.org/indi/abide/).

Code availability

The codes for gradient generation are available at https://github.com/MICA-MNI/BrainSpace; codes for statistical analyses are available at https://github.com/MICA-MNI/BrainStat/tree/master.

Funding

This work was supported by the Institute for Information and Communications Technology Planning and Evaluation (IITP) funded by the Korea Government (MSIT) (No. 2022-0-00448/RS-2022-II220448, Deep Total Recall: Continual Learning for Human-Like Recall of Artificial Neural Networks; RS-2021-II212068, Artificial Intelligence Innovation Hub), Institute for Basic Science (IBS-R015-D1), and the National Research Foundation of Korea (NRF-2022R1A5A7033499 and RS-2022-00165709).

Author contributions

J.Y.N., J.M., J.K., and B.P. designed the study, analyzed data, and wrote the manuscript. Y.J.P. aided the experiment and reviewed the manuscript. J.K. and B.P. are the corresponding authors of this study and are responsible for the integrity of the data analysis. All authors reviewed and approved the final version of the paper to be published.

CRediT authorship contribution statement

Jong Young Namgung: Writing – original draft, Visualization, Validation, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Jongmin Mun:** Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Yeongjun Park:** Writing – review & editing, Methodology, Data curation. **Jaeoh Kim:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Bo-yong Park:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Bo-Yong Park is an Associate Editor for NeuroImage but was not involved in the handling or review process of this manuscript. All other authors declare no conflicts of interest.

Data availability

I have shared the link to my data/code at the Attach File step

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120956.

References

Andrews-Hanna, J.R., 2012. The brain's default network and its adaptive role in internal mentation. Neuroscientist. 18, 251. https://doi.org/10.1177/1073858411403316.
Arnatkevičiūtė, A., Fulcher, B.D., Fornito, A., 2019. A practical guide to linking brain-wide gene expression and neuroimaging data. Neuroimage 189, 353–367. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.011.
Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C., 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54, 2033–2044. https://doi.org/10.1016/J.NEUROIMAGE.2010.09.025.
Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., Wheelwright, S., 2003. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. Philos. Trans. R. Soc. Lond. B Biol. Sci. 358, 361–374. https://doi.org/10.1098/RSTB.2002.1206.
Baron-Cohen, S., Wheelwright, S., 2004. The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. J. Autism. Dev. Disord. 34, 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00.
Baron-Cohen, S., Wheelwright, S., 2003. The friendship questionnaire: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. J. Autism. Dev. Disord. 33, 509–517. https://doi.org/10.1023/A:1025879411971.
Beacher, F.D.C.C., Radulescu, E., Minati, L., Baron-Cohen, S., Lombardo, M.V., Lai, M.C., Walker, A., Howard, D., Gray, M.A., Harrison, N.A., Critchley, H.D., 2012. Sex differences and autism: brain function during verbal fluency and mental rotation. PLoS. One 7, e38355. https://doi.org/10.1371/JOURNAL.PONE.0038355.
Bedford, S.A., Park, M.T.M., Devenyi, G.A., Tullo, S., Germann, J., Patel, R., Anagnostou, E., Baron-Cohen, S., Bullmore, E.T., Chura, L.R., Craig, M.C., Ecker, C., Floris, D.L., Holt, R.J., Lenroot, R., Lerch, J.P., Lombardo, M.V., Murphy, D.G.M., Raznahan, A., Ruigrok, A.N.V., Smith, E., Spencer, M.D., Suckling, J., Taylor, M.J., Thurm, A., Lai, M.C., Chakravarty, M.M., 2019. Large-scale analyses of the relationship between sex, age and intelligence quotient heterogeneity and cortical morphometry in autism spectrum disorder. Mol. Psychiatry 25, 614–628. https://doi.org/10.1038/s41380-019-0420-6. . 2019 253.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc.: Ser. B (Methodol.) 57, 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.
Benkarim, O., Paquola, C., Park, B.yong, Hong, S.J., Royer, J., Vos de Wael, R., Lariviere, S., Valk, S., Bzdok, D., Mottron, L., Bernhardt, B.C., 2021. Connectivity alterations in autism reflect functional idiosyncrasy. Commun. Biol. 4, 1–15. https://doi.org/10.1038/s42003-021-02572-6. . 2021 41.
Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198. https://doi.org/10.1038/nrn2575. . 2009 10:3.
Chen, L., Dou, W.W., Qiao, Z., 2013. Ensemble subsampling for imbalanced multivariate two-sample tests.
Coifman, R.R., Lafon, S., 2006. Diffusion maps. Appl. Comput. Harmon. Anal. 21, 5–30. https://doi.org/10.1016/J.ACHA.2006.04.006.
Constantino, J.N., 2013. Social responsiveness scale. Encyclopedia of autism spectrum disorders, 2919–2929. https://doi.org/10.1007/978-1-4419-1698-3_296
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. https://doi.org/10.1006/cbmr.1996.0014.
Cruces, R.R., Royer, J., Herholz, P., Larivière, S., Vos de Wael, R., Paquola, C., Benkarim, O., Park, B.yong, Degré-Pelletier, J., Nelson, M.C., DeKraker, J., Leppert, I.R., Tardif, C., Poline, J.B., Concha, L., Bernhardt, B.C., 2022. Micapipe: a pipeline for multimodal neuroimaging and connectome analysis. Neuroimage 263, 119612. https://doi.org/10.1016/J.NEUROIMAGE.2022.119612.
Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis: I. segmentation and surface reconstruction. Neuroimage 9, 179–194. https://doi.org/10.1006/NIMG.1998.0395.
Dempster, A.P., Laird, N M, Rubin, D B, 1977. Maximum Likelihood from Incomplete Data via the EM algorithm. J. R. Stat. Soc. Ser. B (Methodol.) 39, 1–38.
Di Martino, A., Yan, C.G., Li, Q., Denio, E., Castellanos, F.X., Alaerts, K., Anderson, J.S., Assaf, M., Bookheimer, S.Y., Dapretto, M., Deen, B., Delmonte, S., Dinstein, I., Ertl-Wagner, B., Fair, D.A., Gallagher, L., Kennedy, D.P., Keown, C.L., Keysers, C., Lainhart, J.E., Lord, C., Luna, B., Menon, V., Minshew, N.J., Monk, C.S., Mueller, S., Müller, R.A., Nebel, M.B., Nigg, J.T., O'Hearn, K., Pelphrey, K.A., Peltier, S.J., Rudie, J.D., Sunaert, S., Thioux, M., Tyszka, J.M., Uddin, L.Q., Verhoeven, J.S., Wenderoth, N., Wiggins, J.L., Mostofsky, S.H., Milham, M.P., 2014. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol. Psychiatry 19, 659–667. https://doi.org/10.1038/MP.2013.78.
Dougherty, J.D., Schmidt, E.F., Nakajima, M., Heintz, N., 2010. Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic. Acids. Res. 38, 4218–4230. https://doi.org/10.1093/NAR/GKQ130.
Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781. https://doi.org/10.1016/J.NEUROIMAGE.2012.01.021.
Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe,

A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355. https://doi.org/10.1016/S0896-6273(02)00569-X.

Fischl, B., Sereno, M.I., Dale, A.M., 1999a. Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage 9, 195–207. https://doi.org/10.1006/NIMG.1998.0396.

Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M., 1999b. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., Van Essen, D.C., Jenkinson, M., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Guo, M., Thost, V., Li, B., Das, P., Chen, J., Matusik, W., 2022. Data-efficient graph grammar learning for molecular generation. ArXiv abs/2203.08031.

Hadjikhani, N., Chabris, C.F., Joseph, R.M., Clark, J., McGrath, L., Aharon, I., Feczko, E., Tager-Flusberg, H., Harris, G.J., 2004. Early visual cortex organization in autism: an fMRI study. Neuroreport 15.

Hardan, A.Y., Girgis, R.R., Adams, J., Gilbert, A.R., Keshavan, M.S., Minshew, N.J., 2006. Abnormal brain size effect on the thalamus in autism. Psychiatry Res. NeuroimAging 147, 145–151. https://doi.org/10.1016/J.PSCYCHRESNS.2005.12.009.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., Van De Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., Abajian, C., Beckmann, C.F., Bernard, A., Bertagnolli, D., Boe, A.F., Cartagena, P.M., Mallar Chakravarty, M., Chapin, M., Chong, J., Dalley, R.A., Daly, B.D., Dang, C., Datta, S., Dee, N., Dolbeare, T.A., Faber, V., Feng, D., Fowler, D.R., Goldy, J., Gregor, B.W., Haradon, Z., Haynor, D.R., Hohmann, J.G., Horvath, S., Howard, R.E., Jeromin, A., Jochim, J.M., Kinnunen, M., Lau, C., Lazarz, E.T., Lee, C., Lemon, T.A., Li, L., Li, Y., Morris, J.A., Overly, C.C., Parker, P.D., Parry, S.E., Reding, M., Royall, J.J., Schulkin, J., Sequeira, P.A., Slaughterbeck, C.R., Smith, S.C., Sodt, A.J., Sunkin, S.M., Swanson, B.E., Vawter, M.P., Williams, D., Wohnoutka, P., Ronald Zielke, H., Geschwind, D.H., Hof, P.R., Smith, S.M., Koch, C., Grant, S.G.N., Jones, A.R., 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391–399. https://doi.org/10.1038/nature11405. . 2012 489:7416.

Hettwer, M.D., Larivière, S., Park, B.Y., van den Heuvel, O.A., Schmaal, L., Andreassen, O.A., Ching, C.R.K., Hoogman, M., Buitelaar, J., van Rooij, D., Veltman, D.J., Stein, D.J., Franke, B., van Erp, T.G.M., van Rooij, D., van den Heuvel, O.A., van Erp, T.G.M., Jahanshad, N., Thompson, P.M., Thomopoulos, S.I., Bethlehem, R.A.I., Bernhardt, B.C., Eickhoff, S.B., Valk, S.L., 2022. Coordinated cortical thickness alterations across six neurodevelopmental and psychiatric disorders. Nat. Commun. 13, 1–14. https://doi.org/10.1038/s41467-022-34367-6. . 2022 13:1.

Hong, S.J., de Wael, R.V., Bethlehem, R.A.I., Lariviere, S., Paquola, C., Valk, S.L., Milham, M.P., Di Martino, A., Margulies, D.S., Smallwood, J., Bernhardt, B.C., 2019. Atypical functional connectome hierarchy in autism. Nat. Commun. 10, 1–13. https://doi.org/10.1038/s41467-019-08944-1. . 2019 10:1.

Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. Neuroimage 62, 782–790. https://doi.org/10.1016/J.NEUROIMAGE.2011.09.015.

Jin, W., Barzilay, R., Jaakkola, T., 2018. Junction tree variational autoencoder for molecular graph generation. In: Dy, J., Krause, A. (Eds.), Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, pp. 2323–2332.

Jung, M., Mody, M., Saito, D.N., Tomoda, A., Okazawa, H., Wada, Y., Kosaka, H., 2015. Sex differences in the default mode network with regard to autism spectrum traits: a resting state fMRI study. PLoS. One 10, e0143126. https://doi.org/10.1371/JOURNAL.PONE.0143126.

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Prog. Artif. Intell. 5, 221–232. https://doi.org/10.1007/S13748-016-0094-0/TABLES/1.

Langen, M., Bos, D., Noordermeer, S.D.S., Nederveen, H., Van Engeland, H., Durston, S., 2014. Changes in the development of striatum are involved in repetitive behavior in autism. Biol. Psychiatry 76, 405–411. https://doi.org/10.1016/j.biopsych.2013.08.013.

Langs, G., Golland, P., Ghosh, S.S., 2015. Predicting activation across individuals with resting-state functional connectivity based multi-atlas label fusion. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9350, pp. 313–320. https://doi.org/10.1007/978-3-319-24571-3_38/COVER.

Larivière, S., Bayrak, Ş., Vos de Wael, R., Benkarim, O., Herholz, P., Rodriguez-Cruces, R., Paquola, C., Hong, S.J., Misic, B., Evans, A.C., Valk, S.L., Bernhardt, B.C., 2023. BrainStat: a toolbox for brain-wide statistics and multimodal feature associations. Neuroimage 266, 119807. https://doi.org/10.1016/J.NEUROIMAGE.2022.119807.

Levy, D., Ronemus, M., Yamrom, B., Lee, Y.ha, Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., Buja, A., Krieger, A., Yoon, S., Troge, J., Rodgers, L., Iossifov, I., Wigler, M., 2011. Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70, 886–897. https://doi.org/10.1016/J.NEURON.2011.05.015.

Li, Y., Li, R., Wang, N., Gu, J., Gao, J., 2023. Gender effects on autism spectrum disorder: a multi-site resting-state functional magnetic resonance imaging study of transcriptome-neuroimaging. Front. Neurosci. 17, 1203690. https://doi.org/10.3389/FNINS.2023.1203690/BIBTEX.

López, V., Fernández, A., García, S., Palade, V., Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. Inf. Sci. (N. Y.) 250, 113–141. https://doi.org/10.1016/J.INS.2013.07.007.

Lord, C., Risi, S., Lambrecht, L., Cook, E.H., Leventhal, B.L., Dilavore, P.C., Pickles, A., Rutter, M., 2000. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. J. Autism. Dev. Disord. 30, 205–223. https://doi.org/10.1023/A:1005592401947/METRICS.

Lord, C., Rutter, M., Le Couteur, A., 1994. Autism diagnostic interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J. Autism. Dev. Disord. 24, 659–685. https://doi.org/10.1007/BF02172145/METRICS.

Marcus, D.S., Harwell, J., Olsen, T., Hodge, M., Glasser, M.F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S.W., Van Essen, D.C., 2011. Informatics and data mining tools and strategies for the human connectome project. Front. Neuroinform. 5. https://doi.org/10.3389/FNINF.2011.00004.

Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., Jefferies, E., Smallwood, J., 2016. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. U.S.A. 113, 12574–12579. https://doi.org/10.1073/PNAS.1608282113/SUPPL_FILE/PNAS.201608282SI.PDF.

Markello, R.D., Arnatkevičiūtė, A., Poline, J.B., Fulcher, B.D., Fornito, A., Misic, B., 2021. Standardizing workflows in imaging transcriptomics with the Abagen toolbox. Elife 10. https://doi.org/10.7554/ELIFE.72129.

Markello, R.D., Hansen, J.Y., Liu, Z.Q., Bazinet, V., Shafiei, G., Suárez, L.E., Blostein, N., Seidlitz, J., Baillet, S., Satterthwaite, T.D., Chakravarty, M.M., Raznahan, A., Misic, B., 2022. Neuromaps: structural and functional interpretation of brain maps. Nat. Methods 19, 1472–1479. https://doi.org/10.1038/s41592-022-01625-w. . 2022 19:11.

Maziarka, Ł., Pocha, A., Kaczmarczyk, J., Rataj, K., Danel, T., Warchoł, M., 2020. Mol-CycleGAN: a generative model for molecular optimization. J. Cheminform. 12, 2. https://doi.org/10.1186/s13321-019-0404-1.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., Burack, J., 2006. Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. J. Autism. Dev. Disord. 36, 27–43. https://doi.org/10.1007/S10803-005-0040-7/METRICS.

Mun, J., Bang, S., Kim, J., 2025. Weighted support vector machine for extremely imbalanced data. Comput. Stat. Data Anal. 203, 108078. https://doi.org/10.1016/j.csda.2024.108078.

Nordahl, C.W., Iosif, A.M., Young, G.S., Perry, L.M., Dougherty, R., Lee, A., Li, D., Buonocore, M.H., Simon, T., Rogers, S., Wandell, B., Amaral, D.G., 2015. Sex differences in the corpus callosum in preschool-aged children with autism spectrum disorder. Mol. Autism. 6, 1–11. https://doi.org/10.1186/S13229-015-0005-4/FIGURES/4.

Park, B.yong, Hong, S.J., Valk, S.L., Paquola, C., Benkarim, O., Bethlehem, R.A.I., Di Martino, A., Milham, M.P., Gozzi, A., Yeo, B.T.T., Smallwood, J., Bernhardt, B.C., 2021. Differences in subcortico-cortical interactions identified from connectome and microcircuit models in autism. Nat. Commun. 12, 1–15. https://doi.org/10.1038/s41467-021-21732-0. . 2021 12:1.

Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56, 907–922. https://doi.org/10.1016/J.NEUROIMAGE.2011.02.046.

Robinson, E.C., Jbabdi, S., Glasser, M.F., Andersson, J., Burgess, G.C., Harms, M.P., Smith, S.M., Van Essen, D.C., Jenkinson, M., 2014. MSM: a new flexible framework for multimodal surface matching. Neuroimage 100, 414–426. https://doi.org/10.1016/J.NEUROIMAGE.2014.05.069.

Rubin, T.N., Koyejo, O., Gorgolewski, K.J., Jones, M.N., Poldrack, R.A., Yarkoni, T., 2017. Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. PLoS. Comput. Biol. 13, e1005649. https://doi.org/10.1371/JOURNAL.PCBI.1005649.

Salimi-Khorshidi, G., Douaud, G., Beckmann, C.F., Glasser, M.F., Griffanti, L., Smith, S.M., 2014. Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. Neuroimage 90, 449–468. https://doi.org/10.1016/J.NEUROIMAGE.2013.11.046.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., Mason, C.E., Bilguvar, K., Celestino-Soper, P.B.S., Choi, M., Crawford, E.L., Davis, L., Davis Wright, N.R., Dhodapkar, R.M., DiCola, M., DiLullo, N.M., Fernandez, T.V., Fielding-Singh, V., Fishman, D.O., Frahm, S., Garagaloyan, R., Goh, G.S., Kammela, S., Klei, L., Lowe, J.K., Lund, S.C., McGrew, A.D., Meyer, K.A., Moffat, W.J., Murdoch, J.D., O'Roak, B.J., Ober, G.T., Pottenger, R.S., Raubeson, M.J., Song, Y., Wang, Q., Yaspan, B.L., Yu, T.W., Yurkiewicz, I.R., Beaudet, A.L., Cantor, R.M., Curland, M., Grice, D.E., Günel, M., Lifton, R.P., Mane, S.M., Martin, D.M., Shaw, C.A., Sheldon, M., Tischfield, J.A., Walsh, C.A., Morrow, E.M., Ledbetter, D.H., Fombonne, E., Lord, C., Martin, C.L., Brooks, A.I., Sutcliffe, J.S., Cook, E.H., Geschwind, D., Roeder, K., Devlin, B., State, M.W., 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70, 863–885. https://doi.org/10.1016/J.NEURON.2011.05.002.

Schaefer, A., Kong, R., Gordon, E.M., Laumann, T.O., Zuo, X.-N., Holmes, A.J., Eickhoff, S.B., Yeo, B.T.T., 2018. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. Cereb. Cortex 28, 3095–3114. https://doi.org/10.1093/cercor/bhx179.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimäki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., Wigler, M., 2007. Strong association of de novo copy number mutations with autism. Science (1979) 316, 445–449. https://doi.org/10.1126/SCIENCE.1138659.

Ségonne, F., Pacheco, J., Fischl, B., 2007. Geometrically accurate topology-correction of cortical surfaces using nonseparating loops. IEEe Trans. Med. ImAging 26, 518–529. https://doi.org/10.1109/TMI.2006.887364.

Tenenbaum, J.B., De Silva, V., Langford, J.C., 2000. A global geometric framework for nonlinear dimensionality reduction. Science (1979) 290, 2319–2323. https://doi.org/10.1126/SCIENCE.290.5500.2319.

Thomas Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., Fisch, B., Liu, H., Buckner, R.L., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125. https://doi.org/10.1152/JN.00338.2011.

Tomasi, D., Volkow, N.D., 2019. Reduced local and increased long-range functional connectivity of the thalamus in autism spectrum disorder. Cereb. Cortex 29, 573–585. https://doi.org/10.1093/CERCOR/BHX340.

Tournier, J.D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.H., Connelly, A., 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. Neuroimage 202, 116137. https://doi.org/10.1016/J.NEUROIMAGE.2019.116137.

Tsatsanis, K.D., Rourke, B.P., Klin, A., Volkmar, F.R., Cicchetti, D., Schultz, R.T., 2003. Reduced thalamic volume in high-functioning individuals with autism. Biol. Psychiatry 53, 121–129. https://doi.org/10.1016/S0006-3223(02)01530-5.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEe Trans. Med. ImAging 29, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908.

Valk, S.L., Xu, T., Paquola, C., Park, B.yong, Bethlehem, R.A.I., Vos de Wael, R., Royer, J., Masouleh, S.K., Bayrak, Ş., Kochunov, P., Yeo, B.T.T., Margulies, D., Smallwood, J., Eickhoff, S.B., Bernhardt, B.C., 2022. Genetic and phylogenetic uncoupling of structure and function in human transmodal cortex. Nat. Commun. 13, 1–17. https://doi.org/10.1038/s41467-022-29886-1. . 2022 13:1.

Vallender, S.S., 2006. Calculation of the Wasserstein distance between probability distributions on the line. 18, 784–786. https://doi.org/10.1137/1118101

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. Neuroimage 80, 62–79. https://doi.org/10.1016/j.neuroimage.2013.05.041.

Vázquez-Rodríguez, B., Suárez, L.E., Markello, R.D., Shafiei, G., Paquola, C., Hagmann, P., Van Den Heuvel, M.P., Bernhardt, B.C., Spreng, R.N., Misic, B., 2019. Gradients of structure–function tethering across neocortex. Proc. Natl. Acad. Sci. u S. a 116, 21219–21227. https://doi.org/10.1073/PNAS.1903403116/SUPPL_FILE/PNAS.1903403116.SAPP.PDF.

Von Luxburg, U., 2007. A tutorial on spectral clustering. Stat. Comput. 17, 395–416. https://doi.org/10.1007/S11222-007-9033-Z/METRICS.

Vos de Wael, R., Benkarim, O., Paquola, C., Lariviere, S., Royer, J., Tavakol, S., Xu, T., Hong, S.J., Langs, G., Valk, S., Misic, B., Milham, M., Margulies, D., Smallwood, J., Bernhardt, B.C., 2020. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. Commun. Biol. 3, 1–10. https://doi.org/10.1038/s42003-020-0794-7. . 2020 3:1.

Walsh, M.J.M., Wallace, G.L., Gallegos, S.M., Braden, B.B., 2021. Brain-based sex differences in autism spectrum disorder across the lifespan: A systematic review of structural MRI, fMRI, and DTI findings. Neuroimage Clin. 31. https://doi.org/10.1016/J.NICL.2021.102719.

Werling, D.M., Geschwind, D.H., 2013. Sex differences in autism spectrum disorders. Curr. Opin. Neurol. 26, 146. https://doi.org/10.1097/WCO.0B013E32835EE548.

Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C., Wager, T.D., 2011. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods 8, 665–670. https://doi.org/10.1038/nmeth.1635. . 2011 8:8.

You, J., Liu, B., Ying, R., Pande, V., Leskovec, J., 2018. Graph convolutional policy network for goal-directed molecular graph generation. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18. Red Hook, NY, USA. Curran Associates Inc., pp. 6412–6422.

Zuo, X.-N., Xu, T., Milham, M.P., 2019. Harnessing reliability for neuroscience research. Nat. Hum. Behav. 3, 768–771. https://doi.org/10.1038/s41562-019-0655-x.