**ABSTRACT**

**BACKGROUND:** Heterogeneity in autism spectrum disorder (ASD) has hindered the development of biomarkers, thus motivating subtyping efforts. Most subtyping studies divide individuals with ASD into nonoverlapping (categorical) subgroups. However, continuous interindividual variation in ASD suggests that there is a need for a dimensional approach.

**METHODS:** A Bayesian model was employed to decompose resting-state functional connectivity (RSFC) of individuals with ASD into multiple abnormal RSFC patterns, i.e., categorical subtypes, henceforth referred to as “factors.” Importantly, the model allowed each individual to express one or more factors to varying degrees (dimensional subtyping). The model was applied to 306 individuals with ASD (5.2–57 years of age) from two multisite repositories. Post hoc analyses associated factors with symptoms and demographics.

**RESULTS:** Analyses yielded three factors with dissociable whole-brain hypo- and hyper–RSFC patterns. Most participants expressed multiple (categorical) factors, suggestive of a mosaic of subtypes within individuals. All factors shared abnormal RSFC involving the default mode network, but the directionality (hypo- or hyper–RSFC) differed across factors. Factor 1 was associated with core ASD symptoms. Factors 1 and 2 were associated with distinct comorbid symptoms. Older male participants preferentially expressed factor 3. Factors were robust across control analyses and were not associated with IQ or head motion.

**CONCLUSIONS:** There exist at least three ASD factors with dissociable whole-brain RSFC patterns, behaviors, and demographics. Heterogeneous default mode network hypo- and hyper–RSFC across the factors might explain previously reported inconsistencies. The factors differentiated between core ASD and comorbid symptoms—a less appreciated domain of heterogeneity in ASD. These factors are coexpressed in individuals with ASD with different degrees, thus reconciling categorical and dimensional perspectives of ASD heterogeneity.

**Keywords:** ASD heterogeneity, Bayesian modeling, Behavioral deficits, Default mode network, Phenotypes, Resting-state functional connectivity

https://doi.org/10.1016/j.biopsych.2019.11.009

High heterogeneity exists among individuals with autism spectrum disorder (ASD), encompassing core ASD symptoms (1), cognitive skills (2), comorbid conditions (3), brain atypicalities (4,5), and genetics (6). Consequently, investigators have made significant efforts to define ASD subtypes. Most studies have focused on the variability of behavioral or cognitive characteristics (7–10). Studies focusing on brain features are emerging (11–14). Here, we propose a Bayesian framework to decompose whole-brain resting-state functional connectivity (RSFC) patterns in individuals with ASD into multiple hypo- and hyper–RSFC patterns, which we will refer to as “factors” (Figure 1A). This approach allows an individual to express one or more factors (categorical subtype) to varying degrees (continuous), potentially reconciling dimensional (13–16) and categorical (11,12,17) models of ASD heterogeneity.

This approach is motivated by two important considerations. First, most previous ASD subtyping studies assumed that each participant belonged to a single (categorical) subtype. By contrast, the term “spectrum” in ASD suggests that there is continuous variation across individuals (18). This variation is observed at varying degrees across multiple symptom domains (19,20). In parallel, evidence from genetics and neurobiology suggests that autism results from the combination of multiple factors underlying distinct pathways (21,22). Thus, ASD interindividual variability may reflect different degrees of expression of such factors and related mechanisms (6,23). Together, these observations motivate a mosaic approach to ASD subtyping that incorporates categorical and dimensional features (Figure 1A). Our model allows each individual with ASD to express more than one latent...
factor. For example, the hypo- and hyper–RSFC pattern of one individual with ASD might be explained by 90% factor 1 and 10% factor 2, while the hypo- and hyper–RSFC pattern of another individual with ASD might be explained by 40% factor 1 and 60% factor 2.

Second, early resting-state fMRI (rs-fMRI) investigations supported models of ASD as a dysconnection syndrome (24–27). Although these early studies focused on a priori regions or networks of interest in small to moderately sized samples, more recent whole-brain investigations of larger samples have suggested that multiple functional networks subserving the wide range of processes impaired in ASD are affected (26). Importantly, recent studies have reconciled previously inconsistent findings of either hypo- or hyperconnectivity in ASD by showing that both patterns coexist and each affects distinct functional circuits (28–30). Nevertheless, these studies rely on traditional case-control analyses, which may miss less-frequently expressed RSFC patterns because of ASD heterogeneity or sampling biases. Thus, in our study we sought to provide a detailed characterization of the nature and spatial extent of functional dysconnections in ASD while accounting for significant heterogeneity among individuals with ASD.

To address these challenges and estimate latent ASD factors with distinct patterns of whole-brain hypo- and hyperconnectivity, we combined 2 multisite rs-fMRI data repositories—Autism Brain Imaging Data Exchange second release (ABIDE-II) (31) and the Gender Explorations of Neurogenetics and Development to Advance Autism Research (GENDAAR) (32). Post hoc analyses were performed to examine common and distinct abnormal RSFC across factors. Furthermore, associations between the latent factors and multiple phenotypic information were examined using multivariate analyses to capture the complexity of ASD.

METHODS AND MATERIALS

Our analyses proceeded in 4 steps (Figure 1B). First, to identify latent ASD factors, we applied a Bayesian model (Figure 1A) to a combined dataset comprising ABIDE-II (31) and GENDAAR
(32). We used this combined dataset to maximize sample size for MRI and non–brain-imaging phenotypic data. Second, we examined the associations between latent factors and phenotypes of participants with ASD (i.e., demographic or behavioral symptoms) in the ABIDE-II + GENDAAR combined sample. Third, control analyses were performed to ensure robustness of results. Last, we utilized another independent dataset, ABIDE first release (ABIDE-I) (28), to explore the drawbacks of case-control analyses, which do not account for ASD heterogeneity. Code for this work is publicly available at https://github.com/ThomasYeoLab/CBIG/tree/master/stable_projects/disorder_subtypes/Tang2020_ASDFactors.

**Participants**

MRI data from the ABIDE and GENDAAR repositories were analyzed. All MRI data underwent preprocessing and quality control (see MRI Preprocessing). The resulting sample comprised 242 participants with ASD and 276 neurotypical (NT) participants from ABIDE-II, which were combined with 64 participants with ASD and 72 NT participants from GENDAAR for primary analyses. An independent sample of 166 participants with ASD and 150 NT participants from ABIDE-I was used for secondary analyses. Age, sex, and head motion were matched between ASD and NT participants within each site. Participants’ characteristics are summarized in Table 1 and Supplemental Table S1.

**MRI Preprocessing**

The neuroimaging data were processed using a pipeline that had been reported previously (33–35). See Supplemental Methods for details. Here, we briefly outline the procedure. rs-fMRI data underwent slice time correction, motion correction, and spatial smoothing prior to application of spatial independent component analysis (ICA) (62) to extract resting-state functional connectivity (RSFC) data from the whole brain. The resulting RSFC maps were regressed against behavioral data. To reduce any potential bias from using multiple ICA pipelines, we utilized the same three ICA pipelines to extract RSFC data from the whole brain. The resulting hemodynamic time courses were extracted and submitted to a linear mixed-effects model to identify latent factor loadings, which were then imported to the Gaussian Mixture Model Analysis (GMM) (36) to identify five latent factors associated with behavior (63). We next examined the associations between latent factors and phenotypes of participants with ASD (i.e., demographic or behavioral symptoms) in the ABIDE-II and GENDAAR combined sample. In the primary analysis, we used this combined dataset to maximize sample size (32). For the secondary analysis, we used an independent sample of 166 participants with ASD and 150 NT participants from ABIDE-I.

**Table 1. Characteristics and Behavioral Data of Participants From the ABIDE-II + GENDAAR Combined Sample**

<table>
<thead>
<tr>
<th>Demographics</th>
<th>ASD, n = 306</th>
<th>NT, n = 348</th>
<th>p Value, 2-Tailed^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, mean (± SD) [range]</td>
<td>14.94 (± 8.59) [5.22–57.00]</td>
<td>14.89 (± 8.20) [5.89–62.00]</td>
<td>.94</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>70 (22.88)</td>
<td>94 (27.01)</td>
<td>.26</td>
</tr>
<tr>
<td>Full-scale IQ^b, mean (± SD) [range]</td>
<td>106 (± 16.90) [68.00–149.00]</td>
<td>115 (± 13.37) [79.00–149.00]</td>
<td>4.43 × 10^−13^c</td>
</tr>
<tr>
<td>Head Motion, Mean (± SD) [Range]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean FD before censoring</td>
<td>0.10 (± 0.06) [0.03–0.56]</td>
<td>0.09 (± 0.07) [0.02–1.09]</td>
<td>.13</td>
</tr>
<tr>
<td>Mean FD after censoring</td>
<td>0.05 (± 0.02) [0.02–0.17]</td>
<td>0.05 (± 0.01) [0.02–0.10]</td>
<td>.49</td>
</tr>
<tr>
<td>Current Medication (by Target)^d, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GABA</td>
<td>2 (0.97)</td>
<td>0 (0.00)</td>
<td>.45</td>
</tr>
<tr>
<td>Glutamate</td>
<td>3 (1.45)</td>
<td>0 (0.00)</td>
<td>.20</td>
</tr>
<tr>
<td>Serotonin</td>
<td>22 (10.63)</td>
<td>1 (0.45)</td>
<td>8.12 × 10^−6^</td>
</tr>
<tr>
<td>Epinephrine/adrenaline</td>
<td>1 (0.48)</td>
<td>0 (0.00)</td>
<td>.97</td>
</tr>
<tr>
<td>Norepinephrine/noradrenaline</td>
<td>50 (24.15)</td>
<td>3 (1.35)</td>
<td>2.13 × 10^−12^</td>
</tr>
<tr>
<td>Dopamine</td>
<td>43 (20.77)</td>
<td>2 (0.90)</td>
<td>5.58 × 10^−11^</td>
</tr>
<tr>
<td>Others</td>
<td>18 (8.70)</td>
<td>9 (4.05)</td>
<td>.08</td>
</tr>
<tr>
<td>Current Medication (by Class)^e, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Antipsychotic</td>
<td>8 (3.86)</td>
<td>0 (0.00)</td>
<td>9.30 × 10^−3^</td>
</tr>
<tr>
<td>Antidepressant</td>
<td>21 (10.14)</td>
<td>1 (0.45)</td>
<td>1.49 × 10^−5^</td>
</tr>
<tr>
<td>Stimulant</td>
<td>32 (15.46)</td>
<td>2 (0.90)</td>
<td>6.71 × 10^−8^</td>
</tr>
<tr>
<td>SHA</td>
<td>4 (1.93)</td>
<td>0 (0.00)</td>
<td>.11</td>
</tr>
<tr>
<td>Mood stabilizer</td>
<td>4 (1.93)</td>
<td>0 (0.00)</td>
<td>.11</td>
</tr>
<tr>
<td>Others</td>
<td>27 (13.04)</td>
<td>9 (4.05)</td>
<td>1.50 × 10^−3^</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Behavioral Patterns</th>
<th>ASD, n = 348</th>
<th>NT, n = 348</th>
<th>p Value, 2-Tailed^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>GU_1</td>
<td>nASD = 8, nNT = 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP_1</td>
<td>nASD = 14, nNT = 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KKI_1</td>
<td>nASD = 31, nNT = 90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYU_1</td>
<td>nASD = 42, nNT = 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OHSU_1</td>
<td>nASD = 23, nNT = 23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCD_1</td>
<td>nASD = 14, nNT = 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCLA_1</td>
<td>nASD = 17, nNT = 27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USM_1</td>
<td>nASD = 10, nNT = 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U_MIA_1</td>
<td>nASD = 8, nNT = 7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a For 14 data collections from ABIDE-II and GENDAAR were included in this study: BNI_1 (nASD = 21, nNT = 15), ETH_1 (nASD = 6, nNT = 21), GU_1 (nASD = 34, nNT = 26), IP_1 (nASD = 10, nNT = 5), IU_1 (nASD = 17, nNT = 15), KKI_1 (nASD = 31, nNT = 90), NYU_1 (nASD = 42, nNT = 23), OHSU_1 (nASD = 23, nNT = 23), TCD_1 (nASD = 14, nNT = 18), UCLA_1 (nASD = 15, nNT = 9), UCD_1 (nASD = 348, nNT = 348). P values that survived false discovery rate correction (q < .05).

^b Only 207 participants with ASD and 222 NT participants from the ABIDE-II + GENDAAR sample have medication information. As such, percentages were computed based on 207 participants with ASD and 222 NT participants (see Supplemental Table S13 for details on medication use in NT participants). We note that the number of participants taking medication by class does not match the number of participants taking medication by target, because 1 medication may have multiple targets. For example, mirtazapine belongs to the antidepressant class but it has multiple targets, such as serotonin and norepinephrine/noradrenaline.
correction, and alignment with anatomical T1. Framewise displacement and voxelwise differentiated signal variance were computed using fslnotion_outliers (36,37). Volumes with framewise displacement >0.2 mm or voxelwise differentiated signal variance >50 were marked as censored frames, together with 1 frame before and 2 frames after. Uncensored data segments lasting fewer than 5 contiguous volumes were also censored (38). Functional runs with >50% censored frames were removed (Supplemental Table S2).

We regressed out 18 nuisance regressors consisting of 6 motion parameters, averaged cerebrospinal ventricular signal, averaged white matter signal, global signal, and their temporal derivatives. Censored frames were ignored when regression coefficients were computed. Data were interpolated across censored frames using least-squares spectral estimation (39). We applied global signal regression (GSR) because of its effectiveness in removing motion-related and respiratory artifacts (40,41). Recent work has also suggested that GSR increases the associations between behavior and RSFC (34). Nevertheless, we performed control analyses using an alternative to GSR (see Control Analyses). Finally, the data were bandpass filtered (0.009 Hz ≤ f ≤ 0.08 Hz), projected onto FreeSurfer fsaverage6 surface space, smoothed with a 6-mm kernel, and downsampled onto FreeSurfer fsaverage5.

**Resting-State Functional Connectivity**

We utilized a cortical parcellation (42) comprising 400 cortical regions of interest (ROIs) (Figure 1C) and a subcortical segmentation (43) comprising 19 subcortical ROIs (Figure 1D). RSFC (Pearson’s correlation) was computed among the average time series of the 419 brain ROIs (ignoring censored frames), yielding a 419 × 419 RSFC matrix for each participant. Age, sex, head motion [mean framewise displacement (37)], and site differences were regressed from all participants’ RSFC data with a general linear model (GLM). Each RSFC entry (i.e., lower triangular entries since the RSFC matrices were symmetric) of the participants with ASD was z-normalized with respect to the 348 ABIDE-II+GENDAAR NT participants. A z score larger (or smaller) than zero indicates hyperconnectivity (or hypoconnectivity) relative to the connectivity observed in the NT participants.

**Latent Factors in the ABIDE-II+GENDAAR Combined Sample**

Latent ASD factors were identified using the ABIDE-II+GENDAAR dataset. We applied a Bayesian model (Figure 1A) to the z-normalized RSFC of the participants with ASD to estimate the latent factors. The model is a variant of Bayesian models that were previously used to discover latent atrophy factors in Alzheimer’s disease (44) and latent components subserving cognitive tasks (45). It assumes that each individual expresses one or more latent factors associated with distinct hypo- and hyper-RSFC patterns. Given the RSFC data and a user-defined number of factors, K, we can estimate the factor composition of each participant—that is, the probability that a participant expresses a latent factor, expressed as Pr(PrFactor|Participant); and the factor-specific hypo- and hyper-RSFC patterns, expressed as E(RSFC patterns|Factor). See Supplemental Methods.

We estimated 2 to 4 latent factors. The estimations were robust for 2 and 3 factors but unstable for 4 factors (Supplemental Figure S1). Therefore, we did not consider a larger number of factors. Furthermore, the 2-factor estimates were inconsistent in the control analyses (see Control Analyses), so we focused on the 3-factor estimates in subsequent analyses.

To estimate confidence intervals for the factor-specific hypo- and hyper-RSFC patterns, we applied a bootstrapping procedure (Supplemental Methods). To reduce multiple comparisons, we averaged the factor-specific hypo- and hyper-RSFC patterns across ROI pairs within and between the 17 networks and subcortical structures (Figure 1C, D), resulting in 18 × 18 matrices, before we computed bootstrapped z scores. The z scores were converted to p values and corrected using false discovery rate (q < .05) along with other tests (Supplemental Methods).

**Associations Between Participants’ Characteristics and Latent Factors in the ABIDE-II+GENDAAR Combined Sample**

We applied separate GLMs (or logistic regression for binary variables) to the factor compositions and each characteristic (age, sex, full-scale IQ, and head motion) of ABIDE-II+GENDAAR participants with ASD to investigate potential associations. For each GLM or logistic regression, the participants’ characteristic and factor compositions were treated as the dependent and independent variables, respectively (Supplemental Methods).

**Associations Between Behavioral Symptoms and Latent Factors in the ABIDE-II+GENDAAR Combined Sample**

Because the ABIDE-II+GENDAAR combined sample comprised datasets across independent sites, not all of the same behavioral measures were collected for all participants (Supplemental Table S3). If we considered all available behavioral measures jointly, we would have only 7 participants. Therefore, available behavioral subscores were divided into 5 groups to maximize the number of participants in each group. For example, the Social Responsiveness Scale (SRS) Autistic Mannerism subscale and Repetitive Behaviors Scale—Revised 6 Subscales (RBS-R6) were grouped together because they index aspects of restricted and repetitive behaviors (RRBs). The Autism Diagnostic Observation Schedule (ADOS) Stereotypy Behavior subscore was not included in the RRB domain because only 38 participants had ADOS Stereotyped Behavior, the SRS Autistic Mannerism and RBS-R6 subscales. The 5 groups of behavioral scores are shown in Supplemental Table S4.

We then applied canonical correlation analysis (CCA) (46) between each group of behavioral scores and each factor loading, i.e., Pr(PrFactor|Participant), which totaled 15 CCA s in the case of the 3-factor model (Supplemental Methods). Through CCA, we sought to find an optimal linear combination of behavioral scores that maximally correlated with the factor.
ADHD Factors With Distinct RSFC–Behavioral Patterns

loading. Age, sex, head motion, and sites were regressed out from behavioral scores and factor loadings before the CCA. Statistical significance was tested using 10,000 permutations accounting for different sites. Multiple comparisons were corrected using false discovery rate (q < .05) (Supplemental Methods).

**Control Analyses in the ABIDE-II + GENDAAR Combined Sample**

First, to ensure robustness to preprocessing strategies, we applied the Bayesian model to rs-fMRI processed using CompCor (47) instead of GSR. Second, we applied k-means clustering to the z-normalized RSFC data of ABIDE-II + GENDAAR participants with ASD (processed with GSR or CompCor) to ensure robustness to analysis strategies (k-means vs. Bayesian model). Third, we compared behavioral associations of the k-means clusters with those of the latent factors. Fourth, we randomly split the 306 participants with ASD in the ABIDE-II + GENDAAR combined sample into 2 groups (Supplemental Table S5) and estimated the latent factors in each group independently. Fifth, we removed 4 small samples (i.e., <10 individuals with ASD or NT individuals) and re-estimated the latent factors. Last, we removed 12 NT participants taking medication and re-estimated the latent factors. See Supplemental Methods for details.

**Drawbacks of Traditional Case-Control Analyses in ABIDE-I**

To explore the drawbacks of case-control analyses, we considered 166 participants with ASD and 150 NT participants from ABIDE-I (Supplemental Table S1) and inferred their factor compositions using latent factors estimated from the ABIDE-II + GENDAAR combined sample. These compositions were used to assign each individual to 1 of 3 subgroups. This subgrouping violated the spirit of our hybrid dimensional–categorical approach but was necessary for comparison with traditional case-control analysis. To ensure robustness, we experimented with 2 different criteria of assigning participants with ASD to subgroups (Supplemental Tables S6 and S7). RSFC differences between each ASD subgroup and demographically matched NT participants were computed and compared with traditional case-control analysis (i.e., RSFC differences between participants with ASD and NT participants without subgrouping). See Supplemental Methods for details.

**RESULTS**

**Latent ASD Factors With Dissociable Hypo- and Hyper–RSFC Patterns**

We applied the Bayesian model (Figure 1A) to 306 ABIDE-II + GENDAAR participants with ASD. An important model parameter is the number of latent factors, K. We experimented with K of 2, 3, and 4. The 4-factor model was unstable (Supplemental Figure S1), so we did not explore more factors. On the other hand, the 2-factor model was sensitive to the pre-processing strategy (Supplemental Results and Supplemental Table S8). Thus, we focused on the 3-factor solution. Each of the 3 factor-specific hypo- and hyper–RSFC patterns among the 400 cortical and 19 subcortical ROIs (Figure 1C, D) are shown in Figure 2A (unthresholded) and Figure 2B (statistically significant). Figure 2C illustrates the significant RSFC patterns averaged within and between the 17 networks and subcortical structures.

Factor 1 was associated with ASD-related hypoconnectivity (blue in Figure 2) within and between perceptual–motor networks (somatomotor A/B, visual A/B, salience/ventral attention A, dorsal attention A/B). By contrast, there was ASD-related hyperconnectivity (red in Figure 2) between perceptual–motor and association networks (default mode, control, salience/ventral attention B), as well as between somatomotor and subcortical regions (caudate and thalamus).

Factor 2 was associated with patterns of hypo- and hyper–RSFC almost opposite to those of factor 1 (r = −.57) but with subtle deviations. For example, regions within default mode networks A and B were strongly hyperconnected in factor 2 but only weakly hypoconnected in factor 1. Similarly, regions between somatomotor networks and the caudate were strongly hyperconnected in factor 1 but did not exhibit any atypical RSFC in factor 2.

Factor 3 was characterized by complex patterns of hypo- and hyper–RSFC. For example, there was hyperconnectivity between visual and somatomotor networks. There was also strong hypoconnectivity among regions within default mode networks A and B and among regions within the visual networks.

**Factor Compositions of Participants With ASD in the ABIDE-II + GENDAAR Combined Sample**

Figure 3 shows the factor compositions of participants with ASD in the ABIDE-II + GENDAAR sample. Most participants expressed multiple latent factors rather than a single factor. No single site showed predominantly one single factor, suggesting that latent factors were not driven by site differences.

**Default Mode Network Exhibits Abnormal Connectivity Across All 3 Factors**

Supplemental Figure S2 shows the statistically significant hypo- and hyper–RSFC unique to each latent factor, allowing for comparison of the RSFC patterns among the 3 factors (Figure 2B).

To examine hypo- and hyper–RSFC patterns that are shared across factors, we binarized (ignoring directionality of abnormality) within-network and between-network blocks with significant bootstrapped z scores (Figure 2C) and summed them across the 3 factors (Figure 4A). In addition, absolute values of hypo- and hyper–RSFC patterns that were significant across all 3 factors (Figure 2B) were summed to obtain the magnitude of hypo- and hyper–RSFC patterns common across factors (Figure 4B). Altered connectivity patterns within default mode A and B networks were notable, as were those between default mode and perceptual–motor networks (somatomotor A, salience/ventral attention A, dorsal attention B). In addition, hypo- and hyperconnectivity within salience/ventral attention A, within dorsal attention and between somatomotor and control B networks were also common across the factors.

Biological Psychiatry June 15, 2020; 87:1071–1082 www.sobp.org/journal 1075
Lastly, Figure 4C shows the strength of involvement of each ROI obtained by summing the rows of Figure 4B. The strength of the default mode network’s involvement was particularly striking. Although atypical default mode network connectivity was present across all factors, the directionality was inconsistent. For example, factor 2 exhibited hyperconnectivity within the default mode network, while factor 1 and 3 exhibited hypoconnectivity (Figure 2B).

**Participants’ Characteristics Across Latent Factors in the ABIDE-II+GENDAAR Combined Sample**

We used GLM (or logistic regression) to investigate whether characteristics (i.e., age, sex, full-scale IQ, head motion) of participants with ASD varied across factors in the ABIDE-II+GENDAAR combined sample. Factor 3 was preferentially expressed by male participants relative to factor 2 ($p < .001$) (Supplemental Figure S3A). Factor 3 was also associated with older participants compared with factors 1 and 2 ($p = .002$ and $p = .01$, respectively) (Supplemental Figure S3B). There was no difference in full-scale IQ or head motion across factors (Supplemental Figure S3C, D).

**Associations Between Latent Factors and Behavioral Symptoms**

To examine associations between latent factors and behavioral symptoms in participants with ASD in the ABIDE-II+GENDAAR combined sample, we performed CCA between each factor loading and each group of behavioral scores.
(see Methods and Materials). Five sets of CCAs remained significant after false discovery rate correction ($q < .05$) (Figure 5). Higher scores indicated worse symptoms, so positive values implied that a higher loading on the factor was associated with greater impairment. Factor 1 was associated with worse RRB ($r = .54, p = .002$) (Figure 5A), social deficits ($r = .27, p = .004$) (Figure 5B), and affective problems ($r = .35, p = .02$) (Figure 5C). Factor 2 was associated with both worse externalizing problems ($r = .35, p = .02$) (Figure 5D) and executive dysfunction ($r = .33, p = .02$) (Figure 5E). Factor 3 was not associated with any aggregated behavioral symptoms.

**Control Analyses**

Here, we summarize the results of the control analyses (see Supplemental Results for more details). First, results for factors 1 and 2 from the 3-factor model were similar regardless of processing pipeline (GSR or CompCor), but those for factor 3 were not (Supplemental Table S8). However, all 3 factors (estimated in primary analyses using GSR) (Figure 2) were similar to clusters obtained from $k$-means regardless of GSR or CompCor (Supplemental Table S9). Thus, overall, factors 1 and 2, and to some extent factor 3, were relatively robust to pre-processing and analysis strategies.

Compared with the latent factors, $k$-means clusters showed similar but weaker behavioral associations (Supplemental Figure S4), suggesting potential advantage of our hybrid dimensional–categorical model. Factors estimated from random splits of ABIDE-II–GENDAAR participants with ASD were similar to the original factors and with each other (Supplemental Table S10). Finally, factors remained highly similar after removing 4 small samples (Supplemental Table S11) or NT participants taking medication(s) (Supplemental Table S12).

**Traditional Case-Control Analysis Yields Smaller Effects and Misses Significant ASD-Related RSFC Associations**

To explore potential drawbacks of traditional case-control analyses, we computed RSFC differences between 166 participants with ASD and 150 NT participants from ABIDE-I. We also computed RSFC differences between participants with ASD and demographically matched NT participants in each ASD subgroup in ABIDE-I (see Methods and Materials). Despite the larger sample size, traditional case–control analysis yielded significantly weaker RSFC differences than the subgroup analyses and missed out on ASD-related RSFC differences (Supplemental Figures S5 and S6). See Supplemental Results for details.

**DISCUSSION**

In this study, we applied a Bayesian model to a large rs-fMRI cohort of individuals with ASD, revealing 3 latent factors with dissociable patterns of hypo- and hyper-RSFC. Each factor was expressed to different degrees across individuals, and each was associated with distinct behavioral and demographic features that are known sources of clinical heterogeneity in ASD, i.e., core ASD impairments, affective problems, externalizing symptoms, executive dysfunctions, age, and sex. Overall, these results suggest that each individual with ASD expresses a mosaic of latent factors. This interindividual variability has been missed in prior subtyping approaches that assigned each individual to a single subtype (7–12,48). By contrast, our approach allows each individual’s factor composition to be unique, thus retaining interindividual variability. This approach is consistent with models suggesting that ASD heterogeneity reflects the contribution of multiple mechanisms with different degrees of expression across subjects (6,23).

Factor 1 loading was the highest when averaged across participants with ASD, so unsurprisingly, the hypo- and hyper-RSFC pattern of factor 1 was the most similar to prior case-control whole-brain comparisons emerging from distinct analytical approaches that were applied to partially overlapping samples from ABIDE-I (28–30). ASD-related hypoconnectivity within sensory and salience networks, as well as hyperconnectivity between somatomotor and subcortical regions, were notable. Likely reflecting its greater prevalence among participants with ASD, this factor was more strongly associated with the core ASD symptoms as measured by SRS and RBS-R (20) but not ADOS scores. We note that the ADOS
is observer based, while RBS-R and SRS scores are not. However, because of the limited number of observer-based measures in the repository relative to questionnaire scores, further analyses determining the sensitivity of RSFC to direct observation versus reported ratings are not feasible.

Perhaps reflecting their associations with ASD symptoms, multiple functional networks that were previously associated with ASD severity were involved in factor 1. For example, social skills impairment and hypo-RSFC within default mode or salience networks have been previously reported (49,50). Although not often explored, hyperconnectivity between the thalamus and temporal cortex has also been associated with social reciprocity deficits in ASD (30), and an atypical RSFC balance in cortical striatal circuitry involving limbic somatomotor and frontoparietal networks has been associated with RRB indexed by RBS-R total scores (51).

Interestingly, the factors did not differentiate among core ASD symptoms (e.g., RRB vs. social reciprocity) but differentiated core ASD and comorbid affective symptoms (factor 1) from comorbid symptoms related to executive dysfunction and externalizing symptoms (factor 2). This finding underscores two important aspects of the ASD phenomenology. One is the strong intercorrelation of ASD symptom domains (20,52) and possibly their partially overlapping biological underpinnings. While neuroimaging studies with deeper phenotyping may be able to characterize the association between the RSFC patterns in a given factor and specific symptom subdomains, larger-scale phenomena reflected in the mosaic of atypical RSFC in a given factor are at play. This strong intercorrelation is suggested by recent reports of atypical hierarchical network organization in ASD relative to that of controls, which was also related to global metrics of ASD traits (53,54). The second aspect of the ASD phenomenology underscored by our brain–behavioral findings is that comorbidity critically contributes to ASD heterogeneity. Comorbidity is a clinical dimension that has been quantitatively and qualitatively overlooked in the neuroimaging literature, albeit with notable exceptions (55,56). Our findings suggest that comorbidity should be more consistently accounted for in ASD biomarker efforts.

In contrast with factor 1, factor 2 was characterized by hypo- and hyper-RSFC that included hyperconnectivity within the default mode and attentional networks, and hypoconnectivity between the default mode and attentional networks. As previously mentioned, factor 2 was also associated with executive dysfunction and externalizing symptoms. These findings are consistent with reports that poor executive control might result from abnormalities in

Figure 4. Patterns of hypo- and hyper-resting-state functional connectivity (RSFC) involved in all 3 factors. (A) Statistically significant within- and between-network hypo- and hyper-RSFC patterns (Figure 2C) were binarized and then summed across the 3 factors; hypo- and hyper-RSFC patterns that were significant in only 1 factor were set to zero. (B) Sum of absolute values of hypo- and hyper-RSFC (Figure 2B) that were significant across all 3 factors. (C) The summed absolute values of hypo- and hyper-RSFC from panel (B) were averaged across the rows for each region of interest (ROI) and projected onto a surface map for visualization. DorsAttn, dorsal attention network; Sal/VentAttn, salience/ventral attention network; SomMot, somatomotor network; TempPar, temporal parietal network; Subcor, subcortical structures.
attentional networks (57). Greater connectivity within the default mode network has also been linked to worse executive function in ASD (50). Furthermore, externalizing symptoms are frequently observed in large proportions of individuals with ASD (58–60). Finally, it remains unknown to what extent factor 2 and the other factors and their relationships with behavior are ASD specific or extend across diagnoses. An initial study (13) has suggested that shared factors across attention-deficit/hyperactivity disorder and ASD might exist, but the lack of a shared behavioral battery between the attention-deficit/hyperactivity disorder and ASD samples limited further exploration with specific symptom domains. Another study has shown that distinct clusters of RSFC patterns exist across individuals with and without ASD, and the subtypes revealed unique brain–behavior relations (14). The emergence and availability of deeply phenotyped transdiagnostic samples (61) will help address this gap.

In our study, factors 1 and 2 were associated with similar age but distinct behavioral deficits, suggesting that they might not simply reflect disease severity or neurodevelopmental stage. On the other hand, factor 3 was more frequently expressed in older participants and might thus reflect a neurodevelopmental stage. The prevalence of ASD is much higher in male participants than in female participants (62). Recent studies have suggested that ASD-related sex differences are associated with atypical brain connectivity and mentalization (63–65). Our study suggests that factor 3 was associated with male participants, but given the small percentage of female participants...
in the ABIDE-II+GENDAAR combined sample, replication with larger sex-balanced datasets is necessary.

Finally, along with factor-specific RSFC signatures, all 3 factors shared atypical RSFC involving the default mode network. This finding is consistent with reports in the larger RSFC ASD literature that have consistently described structural and functional ASD-related abnormalities involving the default mode network (66–68). The directionality of abnormal default mode network RSFC varied by factor, e.g., factor 2 exhibited hyperconnectivity within the default mode network, while factors 1 and 3 exhibited hypoconnectivity (Figure 2B). Similarly, factor 2 exhibited hypoconnectivity between the default mode and attentional networks, but factors 1 and 3 exhibited hyperconnectivity. The differences between factors may explain some of the inconsistent reports in prior studies of default mode network atypicalities in ASD (49,69–71). As shown by our results from additional analyses on the independent ABIDE-I sample, common case-control comparisons fail to appreciate heterogeneous ASD-related RSFC abnormalities, overall urging for future quantitative ASD subtyping efforts.

A major limitation of this study is the lack of harmonized deep phenotyping across all participants with ASD because the datasets were gathered post hoc. Therefore, despite the large sample size, the behavioral associations were separated into 5 sets of analyses performed on overlapping subsets of participants with ASD. A harmonized dataset would also allow the identification of latent factors from RSFC and behavioral deficits simultaneously using a multimodal variant of the current approach (72). We also note that motor deficits are common in ASD (73–75), but the lack of motor control measures precluded any association with the factor loadings. Indeed, whether motor control might be associated with factor 1 (core ASD symptoms) or factor 2 (other deficits) is an important question for future studies. Furthermore, since medication information was limited (e.g., only 8 participants with ASD in the ABIDE-II+GENDAAR sample were reported to be taking antipsychotic medications), we did not explore the association between factors and medications. Finally, longitudinal data are needed to clarify whether the latent factors reflect ASD heterogeneity or neurodevelopmental stages or complex interactions between heterogeneity and neurodevelopment.

Conclusions

Our study revealed 3 latent ASD factors with dissociable whole-brain hypo- and hyper-RSFC patterns. The factors were associated with distinct behavioral symptoms and demographics. Our approach allows each individual to express multiple latent factors to varying degrees rather than a single factor. Therefore, each individual’s factor composition is unique, which might be potentially useful for future biomarker development.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by Singapore Ministry of Education Tier 2 (Grant No. MOE2014-T2-2-016 [to BTTY]), NUS Strategic Research (Grant No. DPR7/944/09/14 [to BTTY]), NUS SOM Aspiration Fund (Grant No. R185000271720 [to BTTY]), Singapore National Medical Research Council (Grant No. CBGR/0088/2015 [to BTTY]), National University Of Singapore Young investigator Award [to BTTY], and the Singapore National Research Foundation Fellowship (Class of 2017 [to BTTY]) and partially supported by National Institute of Mental Health (NIMH) (Grant No. R01MH105506–01 [to ADM]). Our research also used resources provided by the Center for Functional Neuroimaging Technologies (Grant No. P41EB015896) and instruments supported by the Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital (Grant Nos. 1S10RR23401, 1S10RR19307, and 1S10RR020343). Data from GENDAAR Consortium were obtained from the National Institutes of Health (NICH)-supported National Database for Autism Research (NDAR) (NIMH Data Repositories Study No. 2021; NIMH Grant No. R01 MH100028; Principal Investigator: Kevin A. Pelphrey).NDAR (https://ndar.nih.gov/) is a collaborative informatics system created by the NICH to provide a national resource to support and accelerate research in autism. Our computational work was partially performed on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg). Details on the funding sources of each site can be found at http://fcon_1000.projects.nitrc.org/indi/abide/.

This article reflects the views of the authors and may not reflect the opinions or views of the NICH nor of the investigators who submitted original data to NDAR.

We thank the investigators and sites that contributed to the ABIDE repositories. We also thank the GENDAAR Consortium for their efforts in data collection and sharing. The GENDAAR Consortium comprises, in alphabetical order, Elizabeth H. Aylward, Raphael A. Bernier, Susan Y. Bookheimer, Mirella Dapretto, Nadine Gaab, Daniel H. Geschwind, Andrei Irimia, Allison Jack, Charles A. Nelson, Kevin A. Pelphrey, Matthew W. State, John D. Van Horn, Pamela Ventola, and Sara J. Webb. The authors report no biomedical financial interests or potential conflicts of interest.

This article was published as a preprint on bioRxiv: doi: https://doi.org/10.1101/692772.

ARTICLE INFORMATION

From the Department of Electrical and Computer Engineering (ST, NS, BTTY), Centre for Sleep and Cognition, Clinical Imaging Research Centre, N.I. Institute for Health, National University of Singapore; Centre for Cognitive Neuroscience (BTBY), Duke-National University of Singapore Medical School; National University of Singapore Graduate School for Integrative Sciences and Engineering (BTTY), Singapore, Republic of Singapore; Department of Electrical Engineering (ST), Stanford University, Stanford, California; Donders Center for Brain, Cognition and Behavior (DLF), Radboud University; Department for Cognitive Neuroscience (DLF), Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands; Computer Science and Artificial Intelligence Laboratory (XZ), Massachusetts Institute of Technology, Cambridge; Martinos Center for Biomedical Imaging (BTYY), Massachusetts General Hospital, Charlestown, Massachusetts; and Autism and Social Cognition Center (ADM), Child Mind Institute, New York, New York.

ST and NS contributed equally to this work.

Address correspondence to B.T. Thomas Yeo, Ph.D., ECE, CiRfC, N.I. & MNP, National University of Singapore, Singapore, Republic of Singapore. E-mail: thomas.yeo@nus.edu.sg.

Received Jul 4, 2019; revised Oct 15, 2019; accepted Nov 4, 2019.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsych.2019.11.009.

REFERENCES


ASD Factors With Distinct RSFC–Behavioral Patterns


