data

MDPI

Data Descriptor

ICA-Based Resting-State Networks Obtained on Large Autism fMRI Dataset ABIDE

Sjir J. C. Schielen [1,*], Jesper Pilmeyer [1], Albert P. Aldenkamp [1,2], Danny Ruijters [1,3] and Svitlana Zinger [1]

[1] Department of Electrical Engineering, Eindhoven University of Technology, 5612 AP Eindhoven, The Netherlands; j.pilmeyer@tue.nl (J.P.); aldenkampb@kempenhaeghe.nl (A.P.A.); d.ruijters@tue.nl (D.R.); s.zinger@tue.nl (S.Z.)
[2] Epilepsy Center Kempenhaeghe, Department of Behavioral Sciences, 5591 VE Heeze, The Netherlands
[3] Philips Healthcare, Image Guided Therapy, 5684 PC Best, The Netherlands
* Correspondence: s.j.c.schielen@tue.nl

Abstract

Functional magnetic resonance imaging (fMRI) has become instrumental in researching the functioning of the brain. One application of fMRI is investigating the brains of people with autism spectrum disorder (ASD). The Autism Brain Imaging Data Exchange (ABIDE) facilitates this research through its extensive data-sharing initiative. While ABIDE offers raw data and data preprocessed with various atlases, independent component analysis (ICA) for dimensionality reduction remains underutilized. ICA is a data-driven way to reduce dimensionality without prior assumptions on delineations. Additionally, ICA separates the noise from the signal, and the signal components correspond well to functional brain networks called resting-state networks (RSNs). Currently, no large, readily available dataset preprocessed with ICA exists. Here, we address this gap by presenting ABIDE's data preprocessed to extract ICA-based resting-state networks, which are publicly available. These RSNs unveil neural activation clusters without atlas constraints, offering a perspective on ASD analyses that complements the predominantly atlas-based literature. This contribution provides a resource for further research into ASD, benchmarking between methodologies, and the development of new analytical approaches.

Dataset: https://doi.org/10.5281/zenodo.15100689.

Keywords: resting-state functional MRI; independent component analysis; autism spectrum disorder; ABIDE; preprocessing

Academic Editor: Rüdiger Pryss

Received: 25 April 2025
Revised: 23 June 2025
Accepted: 30 June 2025
Published: 3 July 2025

Citation: Schielen, S.J.C.; Pilmeyer, J.; Aldenkamp, A.P.; Ruijters, D.; Zinger, S. ICA-Based Resting-State Networks Obtained on Large Autism fMRI Dataset ABIDE. *Data* **2025**, *10*, 109. https://doi.org/10.3390/data10070109

1. Summary

In the exploration of potential differences between people with autism spectrum disorder (ASD) and healthy controls, analyses progressing from functional magnetic resonance imaging (fMRI) have gained popularity [1,2]. These studies seek to find a neural signature of ASD by analyzing the blood-oxygen-level-dependent (BOLD) signal, a measure that reflects the concentration of deoxygenated hemoglobin resulting from neural activity. As this signal is measured per voxel, the brain is typically parcellated into regions of interest to reduce dimensionality.

Commonly, this parcellation follows a predefined brain atlas. The popularity of dividing the brain according to a predefined atlas can partly be explained by the easy

accessibility of data through the Autism Brain Imaging Data Exchange (ABIDE), which also offers already preprocessed data parcellated with various atlases [3–5].

A more flexible alternative for dimensionality reduction is independent component analysis (ICA), a data-driven form of blind source separation that allows the decomposition into spatially non-overlapping maps and their time series [6]. These maps either contain noise or patterns of activity scattered over the brain. As these patterns are observable without exerting external stimuli (i.e., task-free), they are called resting-state networks (RSNs) [7,8].

While less frequently explored in the literature [2], ICA offers some unique benefits. Where atlas-based parcellation segments the brain based on a standard template, ICA derives components directly from the study's data. This tailors the delineation to the studied sample based on similar voxel activity without presuming delineations [6]. Other dimensionality reduction techniques, such as principal component analysis (PCA) and non-negative matrix factorization (NMF), have also been applied in fMRI. PCA uncorrelates data into orthogonal components, but this does not provide independence [6]. NMF aims to decompose high-dimensional non-negative data into additive, non-negative components, but this can result in weakly identifiable components compared to ICA [9]. The benefit of using ICA is that it optimizes components that have shown good correspondence to RSNs to be statistically independent [10]. Furthermore, it allows the separation between the noise and signal.

The time series resulting from group ICA and dual regression can be interpreted as the amplitudes of the average BOLD signal over an RSN without the influence of the components identified as noise. Similarly to an atlas-based parcellation, each region of interest is described by one time series. Therefore, any type of analysis applicable to atlas-based time series (e.g., connectivity or statistical) can be applied to dual-regressed time series. However, since the dual regression follows group ICA, the results directly relate to functional brain networks, which improves interpretability.

Despite the advantages of ICA, no readily available dataset with extracted RSNs existed hitherto. To provide easy accessibility to a large dataset that allows analyzing RSNs, we have preprocessed the resting-state fMRI data of 900 individuals (417 with ASD and 483 controls) selected from ABIDE I and II. Group ICA was performed on the preprocessed functional scans to identify group RSNs. Subject-specific time series and spatial maps were obtained using dual regression [11,12]. These data are publicly available to complement ABIDE's repository of preprocessed data: https://doi.org/10.5281/zenodo.15100689.

With this alternative approach to atlas-based parcellation, we provide a basis for several other types of analyses. A readily available, preprocessed dataset provides two main benefits. Subsequent analyses can use the dataset directly without preprocessing, which saves time and resources. More importantly, it allows easier benchmarking between studies that could otherwise have heterogeneous preprocessing pipelines.

2. Data Description

This section provides an overview of and information on the available data. All data can be found on Zenodo: https://doi.org/10.5281/zenodo.15100689. Version 2 contains the subject-specific spatial maps. The repository saved under version 1 is described below.

2.1. Structure of the Dataset

The repository contains two main folders. The folder 'code' contains the scripts and resources used to preprocess the raw data from ABIDE. The folder 'data' contains the main dataset in the file 'ABIDE.zip' and the components that resulted from group ICA. The 'ABIDE.zip' follows a folder structure similar to the ABIDE website [13] in which

participants' data is grouped according to the site that acquired the data. The independent components resulting from group ICA are provided in the folder 'data/melodic_IC'. Detailed instructions are provided in the 'readme.md' files in the main folder and in the folders 'code' and 'data'. Furthermore, 'supportingFiles.zip' contains files with information on imaging protocols, framewise displacements, and the identified RSNs.

2.2. Loading the Dataset

The main dataset 'ABIDE.zip' contains the time series resulting from dual regression. One file named 'dr_stage1_subjectXXXXXXX.txt' exists per participant in which 'XXXXXXX' specifies the subject identifier. Each site's subfolder also contains a file named 'phenotypic.csv', which contains the phenotypic information of the participants in the subfolder. Information like the age, sex, and, most importantly, the diagnostic label of the participant is provided in the phenotypic files. The time series and phenotypic data can be loaded in any preferred programming language. As an example, there is a Python notebook 'code/phenotypicDataLoader.ipynb' that acts as sample code to load the phenotypic data.

The text files with the dual regressed time series contain 32 columns because 32 components were selected for group ICA. A list of column indices that correspond to the components identified as RSNs is provided in the file 'RSNs32.xlsx', which can be found in 'data/supportingFiles.zip'. For example, to use the time series corresponding to the default mode network, column 1 should be read. Note that the counting of these components starts at 1, so programming languages that start indexing at 0 (e.g., Python) should account for this.

2.3. Resting-State Networks

The group ICA components identified as resting-state networks are overlaid with the MNI 152 ICBM template [14], as shown in Figure 1. As described in the literature [15], increasing the number of components led to networks being split into subnetworks. While using 50 or 74 components split most networks into subnetworks, using 30 components resulted in networks that were not sufficiently separated from noise. The 32-component ICA showed components separated from noise, while only a few networks were split into subnetworks. Therefore, components were selected from the 32-component ICA. Given the variability in naming conventions for RSNs across studies [16], we have proposed names for the networks in Figure 1 that align with prevalent terminology in the field.

For research flexibility, we have made the output components available on Zenodo, allowing users to select components that best suit their research interests and naming conventions. These can be found in the folder 'data/melodic_IC'. The components are numbered. This number corresponds to the columns in the text files that contain the time series 'dr_stage1_subjectXXXXXXX.txt'. To use the time series from a chosen component, the column with the same index should be used.

2.4. Phenotypic and Demographic Information

A phenotypic overview of the dataset is listed in Table 1. As a result of combining different datasets and exclusions, the dataset is not matched in, e.g., performance intelligence quotient (PIQ) or age. This might warrant considerations for future studies, for example, in the following phenotypes. While the dataset reflects the commonly observed male predominance in ASD diagnoses [17], it also presents an opportunity to explore the less-represented female perspective in ASD research. Different sexes can be considered in studies as differences in certain brain functions (e.g., in working memory or planning) have been associated with sex [18,19]. The group difference in age can be considered as age-related changes occur in the brain, which can affect the fMRI measurement [20]. However, with small effect sizes, eye status was shown to significantly affect the connectivity

within or between RSNs [21]. To accommodate studies requiring cohorts matched on certain phenotypes, the dataset allows for the selection of subsets. Another approach is to consider phenotypic information as covariates in analyses.

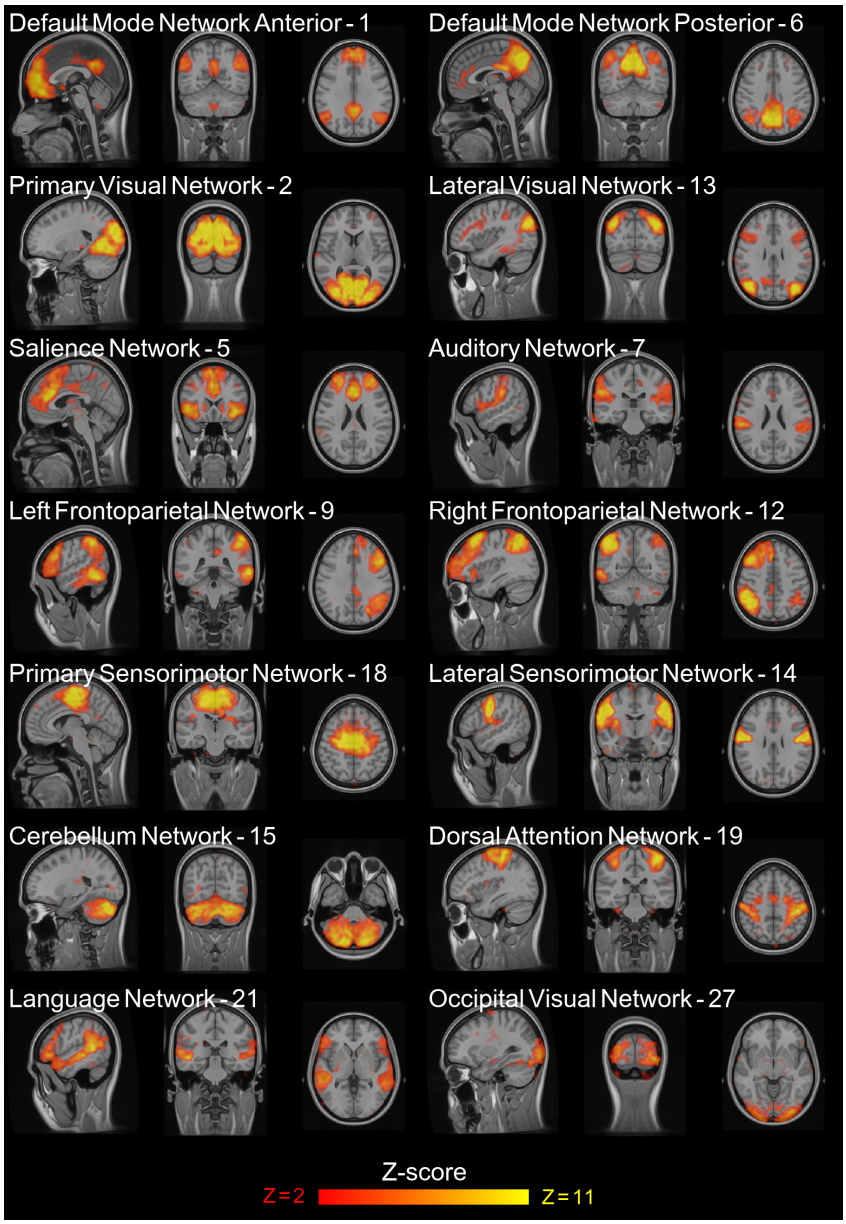

Figure 1. The group independent components that were identified to correspond to resting-state networks, where the Z-score represents voxels' contribution to the component expressed in standard deviations from the mean. Note that the number after the network's name corresponds to the component number as output of group ICA.

Site distributions of diagnosis, sex, age, PIQ, and mean FDs are shown in Figure 2a–e. The occurrences of comorbidities in the dataset are shown in Figure 2f. Note that the comorbidities in Figure 2f do not sum to the total number of people with comorbidities as 32 of the 104 people with comorbidities in the dataset have multiple.

Table 1. Demographic and phenotypic information summarized on the group level, where age and PIQ are reported following the convention mean ± standard deviation; the group differences were tested using the Chi-Square test for sex, eye status, and handedness; the Mann–Whitney U test was used for age; and the Welch test was used for PIQ. Significant differences are in bold assuming a significance level of 0.05. Abbreviations: ASD, autism spectrum disorder; HC, healthy control; m/f, male/female; PIQ, performance intelligence quotient; o, open; c, closed; u; unknown; r, right; l, left; ambi, ambidextrous.

	ASD	HC	*p*-Value	Statistic
Number	417	483	-	-
Sex (m/f)	361/56	377/106	**0.0117**	11.00
Age (years)	12.84 ± 5.04	13.84 ± 5.20	**0.0017**	88519
PIQ	105.78 ± 17.31	108.94 ± 14.54	**0.0040**	−2.89
Eye status (o/c/u)	301/81/35	352/97/34	0.7412	0.5591
Handedness (r/l/ambi/u)	270/32/19/96	330/26/11/116	0.1201	5.832

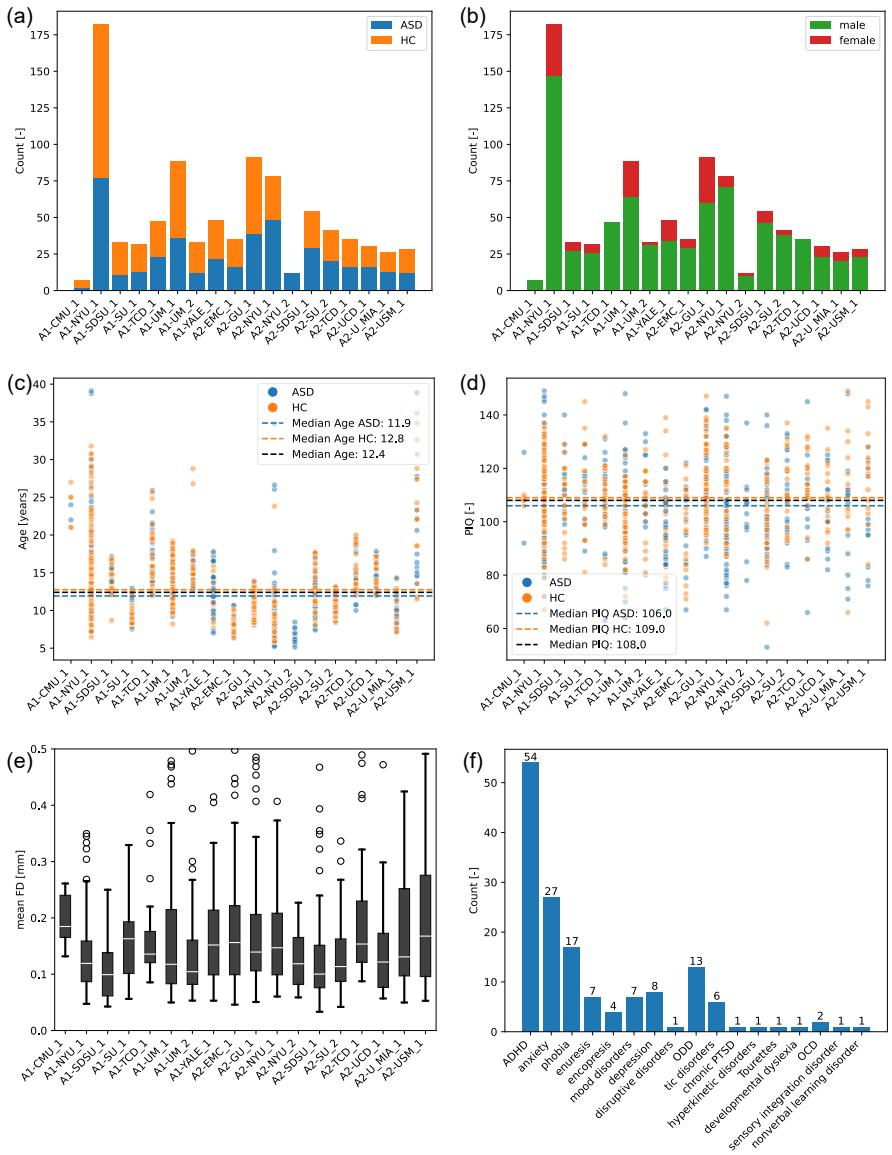

Figure 2. Information on phenotypes and motion. (**a**) Subjects with diagnostic labels per included site. (**b**) Subjects and sex distributions per included site. (**c**) The distribution of age and diagnosis per

included site. (**d**) The distribution of PIQ and diagnosis per included site. (**e**) Mean framewise displacements per site. (**f**) Comorbidities present in the dataset. Abbreviations: ASD, autism spectrum disorder; HC, healthy control; A1, ABIDE-I; A2, ABIDE-II; PIQ, performance intelligence quotient; FD, framewise displacement; CMU, Carnegie Mellon University; NYU, New York University; SDSU, San Diego State University; SU, Stanford University; TCD, Trinity College Dublin; UM, University of Michigan; EMC, Erasmus Medical Center; GU, Georgetown University; UCD, University of California Davis; UMIA, University of Miami; USM, University of Utah School of Medicine; ADHD, attention deficit/hyperactivity disorder; ODD, oppositional defiant disorder; PTSD, post-traumatic stress disorder; OCD, obsessive compulsive disorder.

3. Methods

The Methods section is divided into three parts. The first part outlines the steps involved in selecting data from ABIDE. The second part contains descriptions of the preprocessing steps. The validation of the preprocessed dataset is described in the third part.

3.1. Data Selection

To ensure the dataset was suitable for group ICA and comparisons between individuals, we made the selections summarized in Figure 3a. Initially, we considered all cross-sectional data from ABIDE. As the dataset is preprocessed for group ICA, it requires consistent repetition times (TRs) across subjects because signals are temporally concatenated [22]. Thus, subjects scanned with a TR other than 2000 ms (the most common) were excluded (n = 1026). While temporal interpolation could have allowed the inclusion of more subjects, we prioritized maintaining signal integrity where possible.

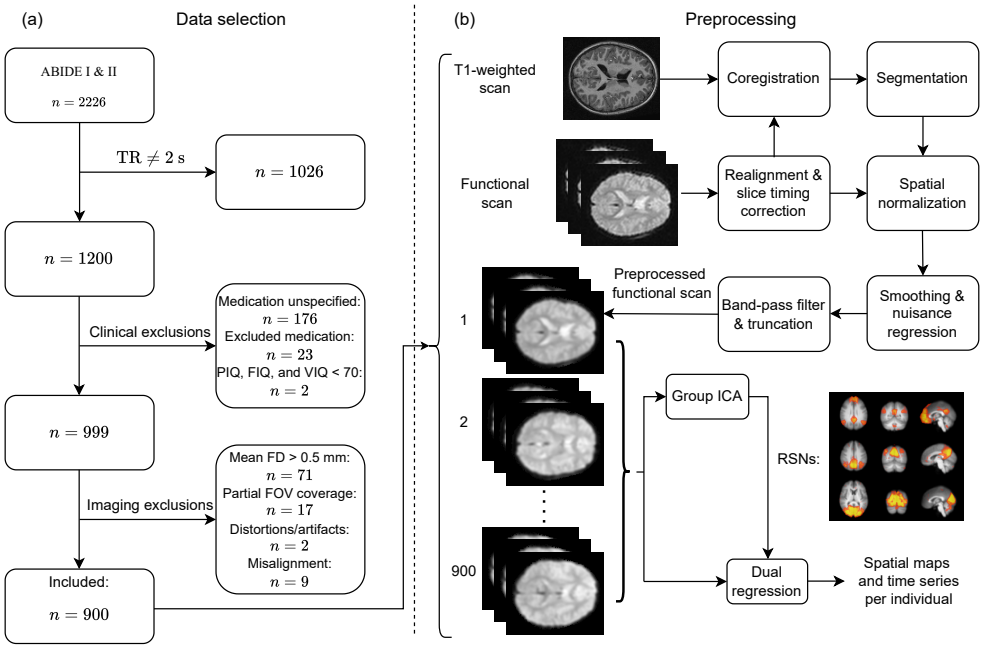

Figure 3. The schematic presentation of the selection process (**a**) and the preprocessing steps (**b**), where n is used to indicate the sample size; TR, the repetition time; P, F, and VIQ, respectively, the Performance, Full-scale, and Verbal Intelligence Quotient; FD, framewise displacement; FOV, the field of view; ICA, independent component analysis; and RSNs, resting-state networks.

Then, potential sources of clinical heterogeneity were considered. Although there is variation throughout the ABIDE sample in diagnostic assessment and instructions for the resting task (e.g., eyes open or closed), we deemed the potential effects of these variations

negligible given the sample size. We recognize that psychoactive medication can significantly affect brain function [23]. Consequently, all patients on centrally active medication potentially affecting fMRI (i.e., valproic acid, oxcarbazepine, topiramate, risperidone, citalopram, and lamotrigine) and the sites that did not report this information were excluded (n = 23 and n = 176, respectively).

Although comorbidities can obscure the interpretation of fMRI results, their high prevalence in ASD led us not to exclude comorbidities [24–26]. Specifically, 59% of the autistic population of ABIDE-II has at least one comorbidity [4]. Rather than excluding comorbidities, we only excluded subjects for which the performance, full-scale, and verbal intelligence quotient ({P, F, V} IQ) were all lower than 70 (n = 2). As these cases were rare in the sample, participants were not excluded if any of the IQ types were not reported.

After the clinical exclusions, 999 subjects were included for preprocessing. As motion can degrade signal quality and introduce spurious findings in fMRI [27], subjects with a mean framewise displacement over 0.5 mm were excluded (n = 71). Furthermore, exclusions based on artifacts were made if the issue caused inappropriate comparisons between individuals. In some scans (n = 17), brain parts of interest (often the cerebellum but occasionally parts of the occipital or temporal lobe) fell outside the field of view (FOV). While some artifacts are rather common (e.g., intensity non-uniformity or susceptibility), exclusions (n = 2) were only made if the artifact changed over time or caused signal loss in areas of interest. When the scans were brought into the standard reference space [28], exclusions (n = 9) were made if the brain significantly deviated from alignment with other brains.

Finally, 900 individuals were included from the ABIDE-I sites: Carnegie Mellon University (CMU), University of Michigan (UM; two samples), and Yale Child Study Center (Yale); the ABIDE-II sites: Erasmus University Medical Center Rotterdam (EMC), Georgetown University (GU), University of California Davis (UCD), University of Miami (UMia), and University of Utah School of Medicine (USM); and the sites both in ABIDE-I and ABIDE-II: New York University Langone Medical Center (NYU; both ABIDE-II samples), San Diego State University (SDSU), Stanford University (SU), and Trinity Centre for Health Services (TCD). As some sites are in both iterations of ABIDE, site abbreviations contain the prefix A1 or A2 to represent the sample from ABIDE-I and ABIDE-II when necessary.

3.2. Preprocessing

Figure 3b illustrates the step-by-step preprocessing approach aimed at preparing functional scans for group ICA. Then, dual regression was used to obtain spatial maps and time series per individual. It is worth noting that the preprocessing pipeline is uniformly applied to each participant, with the exception of group ICA, which is performed over each individual's preprocessed functional scan collectively.

The preprocessing approach and scripts were adapted from the work of Heunis (2021) [29] to handle multiple sites. The established software packages FSL (version 6.0.7.6) [30] and SPM (SPM12) [31] are used in the pipeline. The preprocessing steps up to and including smoothing follow the order of the SPM manual [32]. Then, nuisance regression was placed in between smoothing and temporal filtering as recommended by Pruim et al. [33]. Finally, truncation is performed just before group ICA as it provides equal weights between scans. The next paragraphs contain descriptions of each preprocessing step following the order of Figure 3b.

Realignment primarily minimizes misalignment between subsequent scanned volumes via image registration. After discarding the first four volumes of each functional scan to compensate for non-steady-state signals, each volume in the scan was registered to the first using the MCFLIRT function of FSL. Image registrations were performed with six degrees of freedom (motion parameters): three for rotation and three for translation.

The summation of the absolute difference between motion parameters per time step, i.e., the framewise displacement (FD), was used to exclude subjects if the mean FD exceeded 0.5 mm. While it is argued that even movements of small magnitude (less than 0.05 mm) can have artifactual effects on fMRI data [34], motion is common in the dataset, and there is no consensus on quality assessment standards [4]. Therefore, fMRI-based findings, e.g., biomarkers, should be robust to some extent of motion, which we set at a mean FD of 0.5 mm. Although ICA separates noise caused by motion from the signal, in cases of severe motion, errors in realignment increase. The threshold of 0.5 mm also limits the influence of these cases.

Slice timing correction temporally interpolates slices to account for acquisition times within a single volume. As this dataset is intended for methodologies that investigate differences between ASD and controls (e.g., group differences, diagnostic tools, or biomarker identification), the preprocessing is tailored to accommodate multiple potential analyses. Common methods are analyses progressing from functional or effective connectivity [35], where the latter estimates a sense of causality that requires correct timing. For example, dynamic causal modeling is a method of estimating effective connectivity, which requires slice timing correction [36]. Moreover, different acquisition protocols were used (sequential and interleaved slice acquisitions), which means comparisons between scans acquired with different slice orders are less appropriate than when corrected. Slice timing correction was performed referencing the middle slice using SPM.

Coregistration aligns the anatomical scan with the functional scan, leveraging the anatomical scan's higher spatial resolution for segmentation. Coregistration was performed using SPM with the default settings. For some subjects, the spatial alignment between the functional and anatomical scans differed significantly causing coregistration to fail. In these cases, a manual translation was performed to create a better starting position for the registration. If the coregistration failed after this effort and the misalignment caused incorrect spatial normalization, the subject was excluded (n = 9).

Segmentation was performed in SPM with default settings to obtain probability maps of air, soft tissue, bone, grey matter, white matter, and cerebrospinal fluid in the anatomical scan. These maps were later used for masking. In the segmentation step, the forward transformation of native space to Montreal Neurological Institute (MNI) space was obtained [28]. The forward transformation was applied to the functional image in the spatial normalization step, which also involved 4th-order B-spline interpolation to make the spatial resolution 2 × 2 × 2 mm. Normalization to the same space and a common voxel size among subjects scanned with different acquisition protocols was necessary to ensure correct spatial comparison between subjects.

Smoothing is commonly used in fMRI preprocessing pipelines to increase the signal-to-noise ratio and lower inter-subject variability [37]. SPM was used to perform smoothing with a three-dimensional Gaussian kernel with a full width at half maximum of 5 mm in each dimension. Despite smoothing, realignment, and exclusions of excessive FD, secondary effects of motion show up artifactually in the fMRI signal, which may result in spurious findings of, e.g., functional connectivity [27,38]. Therefore, ICA-based Automatic Removal of Motion Artifacts (ICA-AROMA) was used to decompose the fMRI signal and automatically classify and regress out motion-related noise components [33]. We chose ICA-AROMA for its effectiveness in removing motion-induced correlations between distant voxels while preserving functional network identifiability [39]. Extending ICA-AROMA with global signal regression (GSR) could eliminate more noise components from the signal at the risk of losing signal valuable to, e.g., functional network identifiability [40]. Considering the manual selection of components was performed after group ICA, we

decided not to extend ICA-AROMA with GSR as only the components corresponding to RSNs were selected.

A band-pass filter passing 0.01 to 0.1 Hz was applied to restrict the fMRI signal to the neural frequency range while minimizing cardiac and respiratory interference [41,42]. This was implemented with a second-order zero-phase digital Butterworth filter in Matlab R2023b. To ensure equal weighting between individuals in group ICA, all fMRI scans were truncated to 146 volumes, corresponding to the shortest length of the included scans minus the four discarded volumes at the start. The truncated volumes all contain the first 146 volumes (after discarding) as exclusion-worthy artifacts or distortions were mainly observed later in the scans.

Group ICA was performed using FSL's Multivariate Exploratory Linear Optimized Decomposition into Independent Components (MELODIC). In MELODIC, preprocessed functional scans were temporally concatenated, and an initial PCA was used to whiten the data and reduce dimensionality. Then, the implementation of probabilistic ICA was used to perform the decomposition [22]. Increasing the model order (the number of components) tends to decompose networks into subnetworks [15]. Therefore, multiple iterations of this process were run, each with a different number of components: 30, 32, 50, and 'default' in which the number of components was estimated automatically to be 74. It is worth noting that a set number of components for ICA still maintains the flexibilities of ICA: data-driven delineations directly on the sample and denoising by disregarding the noise components.

The components corresponding to RSNs were selected following the approach of Bernas et al. (2018) [43]. The 'goodness-of-fit' (GoF) was computed between the template spatial maps provided in the Smith atlas [10] and the spatial maps from group ICA on our sample for a threshold of $Z = 3$. The GoF is the difference between the average Z-scores of the voxels (of our spatial maps) falling within the template spatial map and the Z-scores of the voxels falling outside the template spatial map. Using the best fits between the obtained RSNs and the template spatial maps as an initial reference, the results were further validated with visual inspection and discussion until a consensus between the authors was reached. Further visual inspection is necessary as RSNs not included in the Smith atlas or subnetworks can occur.

The goodness-of-fit scores are shown in Figure 4. The components are represented by their index in the output components of group ICA, which are shown in Figure 1. Overall, there is good correspondence between the networks of the Smith atlas and the components from group ICA. However, some differences can be observed. For example, the default mode network (anterior and posterior) and the sensorimotor network (primary and lateral) are split into subnetworks. Furthermore, we found two networks that were not included in the Smith atlas: the dorsal attention network and the language network.

Dual regression was used to calculate each individual's version of the group-level components and their associated time series [44]. The benefit of dual regression is that the group ICA components serve as the design matrix in the general linear model, which is optimized to resemble the individual's preprocessed functional scan. This captures the individual's variability while preserving the spatial characteristics of the group ICA.

3.3. Validation

In the ABIDE-II initiative, data was shared regardless of imaging quality because of the absence of a consensus on quality criteria and to accommodate the development of artifact correction methods [4]. The dataset that we present here was selected to perform group ICA from which further analyses can progress. Therefore, care was taken to ensure proper comparisons between subjects in the dataset.

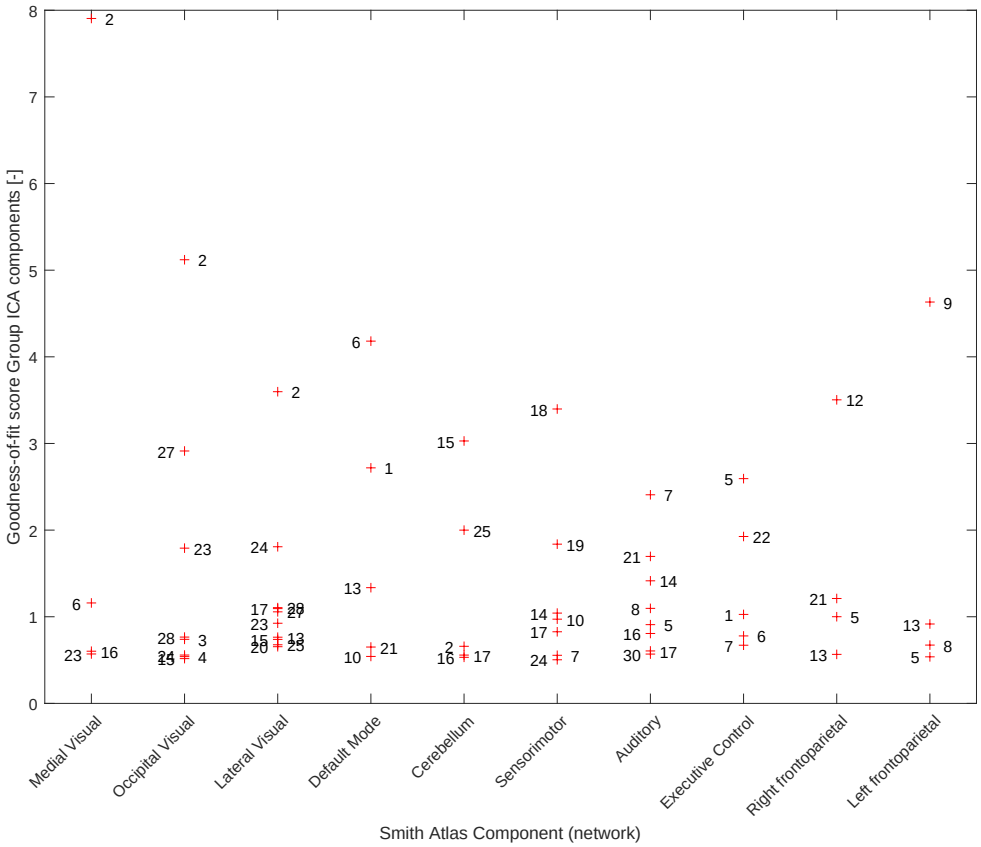

Figure 4. The goodness-of-fit between the components obtained with group ICA and the Smith atlas's template RSNs. To aid readability, only the scores above 0.5 are plotted. The numbers indicate the component number from group ICA: 1, Default Mode Network Anterior; 2, Primary Visual Network; 5, Salience Network; 6, Default Mode Network Posterior; 7, Auditory Network; 9, Left Frontoparietal Network; 12, Right Frontoparietal Network; 13, Lateral Visual Network; 14, Lateral Sensorimotor Network; 15, Cerebellum Network; 18, Primary Sensorimotor Network; 19, Dorsal Attention Network; 21, Language Network; 27, Occipital Visual Network. The numbers 8, 10, 16, 17, 23, 24, 25, 28, and 30 were classified as noise.

To validate that preprocessed functional scans can be compared properly, we visually inspected all 999 preprocessed functional scans in the dataset and discussed potential exclusions until a consensus was reached among the authors. We prioritized two main points: spatial alignment should be correct to ensure that the same regions are considered among participants in ICA and artifacts or distortions should not change over time, which might cause spurious activation patterns. This led to the exclusion of participants, of which the following subject identifiers (SIDs) are examples: time-varying artifacts (SID 28901, A2-SDSU), partial field-of-view coverage (SID 50653, A1-CMU), and improper coregistration (SID 50561 A1-Yale).

Note that not all artifact types were excluded. The types of artifacts that are still present in the dataset are both darker (e.g., SID 50958 A1-NYU) and brighter (e.g., SID 28755 A2-GU) susceptibility artifacts, intensity non-uniformity (e.g., SID 50603 A1-Yale), and slight wrap-around artifacts (e.g., SID 28755 A2-GU). Artifacts are also part of a realistic dataset, so subsequent analysis should have the opportunity to incorporate some robustness.

Furthermore, we tested for spatial differences between the subject-specific RSNs of the ASD and HC groups using randomized permutation testing. The second step in dual regression involves regressing the subject-specific time series per RSN (as temporal regressors in a multiple regression) into the subject's preprocessed functional scan, resulting

in subject-specific versions of each group-level spatial map. Each subject-specific RSN was then used in FSL's randomise permutation-testing tool to perform a two-sample unpaired t-test for differences between the ASD and control group while accounting for the nuisance variables age and sex [12].

No significant differences were found with permutation testing. The lowest p-value ($p = 0.078$) was found for the occipital visual network. The full results of permutation testing are listed in Table 2. The lack of significant p-values shows no significant differences in the subject-specific spatial maps between subjects with ASD and controls. As the subject-specific spatial maps result from regressing the subject-specific time series into a subject's 4D scan, there is no significant difference in which voxels contribute to this temporal behavior between ASD and control for each RSN. Therefore, no significant structural differences in the RSNs were found between ASD and control.

Table 2. The results of two-sample t-statistic permutation testing using FSL's randomise. The reported p-values are the lowest voxel values (or the highest (1-p)-value as this is how the results are returned) per resting-state network. This means there are no voxels closer to significance than the ones reported here. Note that p-values are corrected for multiple comparisons over voxels but not for multiple comparisons over RSNs, which would require further correction pushing them further from significance. Abbreviations: RSN, resting-state network.

RSN	p-Value
Default Mode Network Anterior	0.5158
Default Mode Network Posterior	0.1654
Primary Visual Network	0.3296
Lateral Visual Network	0.7056
Salience Network	0.3076
Auditory Network	0.5048
Left Frontoparietal Network	0.7996
Right Frontoparietal Network	0.4128
Primary Sensorimotor Network	0.4156
Lateral Sensorimotor Network	0.6470
Cerebellum	0.8012
Dorsal Attention Network	0.1766
Language Network	0.3550
Occipital Visual Network	0.0782

4. Conclusions

In this paper, we presented the considerations of preprocessing a dataset sampled from ABIDE to obtain ICA-based resting-state networks. The main contribution is a readily available, preprocessed dataset that provides a basis for analyzing resting-state networks in a large sample consisting of individuals with ASD and controls.

The presented dataset is based on group ICA, which is inherently limited to a single repetition time. Consequently, the presented dataset can be viewed as a case study, which requires adopting the same selection criteria while allowing for further personalized selections. Future work could investigate an approach to perform group ICA on scans taken with different repetition times while preserving the temporal information in the signals. This would allow a larger sample to be preprocessed. The proposed dataset can be

used to, e.g., analyze RSNs directly or the interactions between them to investigate group differences, an individual-level diagnosis, or features related to symptom severity in ASD.

Author Contributions: S.J.C.S.: conception, preprocessing, visual inspection, ICA component selection, and manuscript. J.P.: preprocessing, visual inspection, and ICA component selection. A.P.A.: patient selection, ICA component selection, and supervision. D.R.: preprocessing and supervision. S.Z.: conception, preprocessing, ICA component selection, and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This work is only funded by Eindhoven University of Technology, Department of Electrical Engineering.

Institutional Review Board Statement: While this dataset involves human subjects, we were not involved in the data acquisition. We only preprocessed the data available online at [13] and shared the preprocessed version under the same license. Information about data acquisition is provided in [3,4,13].

Informed Consent Statement: While this dataset involves human subjects, we were not involved in the data acquisition. We only preprocessed the data available online at [13] and shared the preprocessed version under the same license. Information about data acquisition is provided in [3,4,13].

Data Availability Statement: The preprocessed dataset and the used preprocessing code can be found at Zenodo: https://doi.org/10.5281/zenodo.15100689. Note that this preprocessing pipeline was tailored to Matlab R2023b on Windows, which requires SPM12 [31] and the image processing toolbox. Furthermore, FSL 6.0.7.6 [30] and the scripts and resources for ICA-AROMA [33] were used. FSL was run in the Windows Subsystem for Linux (WSL), which involved adapting commands and the ICA-AROMA scripts to work in WSL. These versions are available in the repository.

Acknowledgments: A preprint article is associated with this paper: https://arxiv.org/html/2412.13798v1 [45] (18 December 2024). We acknowledge the significant contributions of ABIDE-I and ABIDE-II and their primary sources of funding: NIMH K23MH087770 and NIMH 5R21MH107045 [3,4]. Furthermore, as requested on the ABIDE website [13], we acknowledge the funding sources of the sites included in this work: Carnegie Mellon University: NICHD/NIDCD PO1/U19 to M.B. (PI: Nancy Minshew), which is part of the NICHD/NIDCD Collaborative Programs for Excellence in Autism, and Simons Foundation to M. B. (PI: David Heeger); NYU Langone Medical Center: NIH (R21MH102660; K23MH087770; R21MH084126; R01MH081218; R01HD065282), Autism Speaks, The Stavros Niarchos Foundation, The Leon Levy Foundation, and an endowment provided by Phyllis Green and Randolph Cowen and Goldman Sachs on behalf of Ram Sundaram; San Diego State University: NIH R01-MH081023 (PI: Müller) and NIH K01 MH-097972 (PI: Fishman); Stanford University: National Institute of Mental Health Career Development Award [K01MH092288] to L.U.; Stanford Institute for Neuro-Innovation & Translational Neurosciences; National Institutes of Health [DC011095, MH084164] to V.M., National Institutes of Health [K01MH102428] to D.A.A., The Singer Foundation, The Simons Foundation (VM), and The Autism Science Foundation (AP); Trinity Centre for Health Sciences: The Meath Foundation, Adelaide and Meath Hospital, incorporating the National Children's Hospital (AMNCH), Tallaght, and travel fellowship by the Kyulan Family Foundation, National Children's Research Centre (NCRC), Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland; University of Michigan: Autism Speaks (CSM), NIH U19 HD035482 and NIH MH066496 (CL), Autism Speaks Pre-doctoral Fellowship 4773 (JLW), Michigan Institute for Clinical and Health Research (MICHR), Pre-doctoral Fellowship UL1RR024986 (JLW) and NIH R21 MH079871 (SP); Yale Child Study Center: Simons Foundation (KP), Autism Speaks (KP), John Merck Scholars Fund (KP), Autism Science Foundation, NICHD (KP) and NIMH; Erasmus University Medical Center Rotterdam: "This study received support from the Simons Foundation Autism Research Initiative (SFARI - 307280) and a Dutch ZonMw TOP grant number 91211021 to Tonya White. MRI data acquisition was sponsored in part by the European Community's 7th Framework Programme (FP7/2008-2013, 212652). Supercomputing computations were supported by the NWO Physical Sciences Division (Exacte Wetenschappen) and SURFsara (Lisa compute cluster, www.surfsara.nl). The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the School of Law and Faculty

of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, the Rotterdam Homecare Foundation, Rotterdam and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam."; Georgetown University: National Institute of Mental Health (NIMH), Grant. No. MH084961, to C. J. V., and Intellectual and Developmental Disabilities Research Center, Children's National Medical Center, Grant No. HD040677-07, to W. D. G. and L. E. K.; University of California Davis: National Institute of Mental Health (1R21MH099250-01; The Neural Substrates of Higher Level Learning in Autism); University of Miami: "This work was supported by awards K01MH092288 and R01MH107549 from the National Institute of Mental Health, a Slifka/Ritvo Innovation in Autism Research Award from the International Society for Autism Research, and a NARSAD Young Investigator Award to LQU." University of Utah School of Medicine: National Institutes of Health [grant numbers K08 MH092697, RO1MH080826, P50MH60450, T32DC008553, and R01NS34783], Autism Speaks Mentor-based Predoctoral Fellowship [grant number 1677], University of Utah Multidisciplinary Research Seed Grant, NRSA Predoctoral Fellowship [grant number F31 DC010143], and Ben B. and Iris M. Margolis Foundation.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

A1	ABIDE-I
A2	ABIDE-II
ABIDE	autism brain imaging data exchange
ambi	ambidextrous
ASD	autism spectrum disorder
BOLD	blood oxygen level dependent
c	closed
CMU	Carnegie Melon University
DOI	digital object identifier
EMC	Erasmus University Medical Center
f	female
FD	framewise displacement
FIQ	full-scale intelligence quotient
fMRI	functional magnetic resonance imaging
FOV	field of view
FSL	fMRI Brain Software Library
GSR	global signal regression
GU	Georgetown University
GoF	goodness-of-fit
HC	healthy control
ICA	independent component analysis
ICA-AROMA	ICA-based Automatic Removal of Motion Artifacts
l	left
m	male
MCFLIRT	Motion Correction Using FMRIB's Linear Image Registration Tool
MELODIC	Multivariate Exploratory Linear Optimized Decomposition into Independent Components
MNI	Montreal Neurological Institute
n	number
NIMH	National Institute of Mental Health
NMF	Non-negative matrix factorization
NYU	New York University Langone Medical Center
o	open
PCA	principal component analysis

NYU	New York University Langone Medical Center
o	open
PCA	principal component analysis
PI	Principal Investigator
PIQ	performance intelligence quotient
r	right
RSN	resting-state network
SDSU	San Diego State University
SID	subject identifier
SPM	Statistical Parametric Mapping
SU	Stanford University
TCD	Trinity Centre for Health Services
TR	repetition time
u	unknown
UCD	University of California Davis
UM	University of Michigan
UMia	University of Miami
USM	University of Utah School of Medicine
VIQ	verbal intelligence quotient
WSL	Windows subsystem for Linux
Yale	Yale Child Study Center

References

1. Santana, C.P.; de Carvalho, E.A.; Rodrigues, I.D.; Bastos, G.S.; de Souza, A.D.; de Brito, L.L. rs-fMRI and machine learning for ASD diagnosis: A systematic review and meta-analysis. *Sci. Rep.* **2022**, *12*, 6030. [CrossRef]
2. Schielen, S.J.; Pilmeyer, J.; Aldenkamp, A.P.; Zinger, S. The diagnosis of ASD with MRI: A systematic review and meta-analysis. *Transl. Psychiatry* **2024**, *14*, 318. [CrossRef] [PubMed]
3. 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.; et al. The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **2014**, *19*, 659–667. [CrossRef] [PubMed]
4. Di Martino, A.; O'connor, D.; Chen, B.; Alaerts, K.; Anderson, J.S.; Assaf, M.; Balsters, J.H.; Baxter, L.; Beggiato, A.; Bernaerts, S.; et al. Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Sci. Data* **2017**, *4*, 170010. [CrossRef] [PubMed]
5. Craddock, C.; Benhajali, Y.; Chu, C.; Chouinard, F.; Evans, A.; Jakab, A.; Khundrakpam, B.S.; Lewis, J.D.; Li, Q.; Milham, M.; et al. The neuro bureau preprocessing initiative: Open sharing of preprocessed neuroimaging data and derivatives. *Front. Neuroinformatics* **2013**, *7*, 5.
6. Calhoun, V.D.; Liu, J.; Adalı, T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* **2009**, *45*, S163–S172. [CrossRef]
7. Niazy, R.K.; Cole, D.M.; Beckmann, C.F.; Smith, S.M. Resting-state networks. In *fMRI: From Nuclear Spins to Brain Functions*; Springer: New York City, NY, USA, 2015; pp. 387–425.
8. Damoiseaux, J.S.; Rombouts, S.A.; Barkhof, F.; Scheltens, P.; Stam, C.J.; Smith, S.M.; Beckmann, C.F. Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13848–13853. [CrossRef]
9. Xie, J.; Douglas, P.K.; Wu, Y.N.; Brody, A.L.; Anderson, A.E. Decoding the encoding of functional brain networks: An fMRI classification comparison of non-negative matrix factorization (NMF), independent component analysis (ICA), and sparse coding algorithms. *J. Neurosci. Methods* **2017**, *282*, 81–94. [CrossRef]
10. Smith, S.M.; Fox, P.T.; Miller, K.L.; Glahn, D.C.; Fox, P.M.; Mackay, C.E.; Filippini, N.; Watkins, K.E.; Toro, R.; Laird, A.R.; et al. Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13040–13045. [CrossRef]
11. Beckmann, C.F.; Mackay, C.E.; Filippini, N.; Smith, S.M. Group comparison of resting-state FMRI data using multi-subject ICA and dual regression. *Neuroimage* **2009**, *47*, S148. [CrossRef]
12. Nickerson, L.D.; Smith, S.M.; Öngür, D.; Beckmann, C.F. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Front. Neurosci.* **2017**, *11*, 115. [CrossRef] [PubMed]
13. Di Martino, A.; Mostofsky, S.; Milham, M.P. Autism Brain Imaging Data Exchange (ABIDE). 2017. Available online: http://fcon_1000.projects.nitrc.org/indi/abide/ (accessed on 4 October 2023).

14. Fonov, V.S.; Evans, A.C.; McKinstry, R.C.; Almli, C.R.; Collins, D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* **2009**, *47*, S102. [CrossRef]
15. Abou-Elseoud, A.; Starck, T.; Remes, J.; Nikkinen, J.; Tervonen, O.; Kiviniemi, V. The effect of model order selection in group PICA. *Hum. Brain Mapp.* **2010**, *31*, 1207–1216. [CrossRef]
16. Uddin, L.Q.; Yeo, B.; Spreng, R.N. Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topogr.* **2019**, *32*, 926–942. [CrossRef]
17. Zeidan, J.; Fombonne, E.; Scorah, J.; Ibrahim, A.; Durkin, M.S.; Saxena, S.; Yusuf, A.; Shih, A.; Elsabbagh, M. Global prevalence of autism: A systematic review update. *Autism Res.* **2022**, *15*, 778–790. [CrossRef]
18. Goldstein, J.M.; Jerram, M.; Poldrack, R.; Anagnoson, R.; Breiter, H.C.; Makris, N.; Goodman, J.M.; Tsuang, M.T.; Seidman, L.J. Sex differences in prefrontal cortical brain activity during fMRI of auditory verbal working memory. *Neuropsychology* **2005**, *19*, 509. [CrossRef]
19. Boghi, A.; Rasetti, R.; Avidano, F.; Manzone, C.; Orsi, L.; D'Agata, F.; Caroppo, P.; Bergui, M.; Rocca, P.; Pulvirenti, L.; et al. The effect of gender on planning: An fMRI study using the Tower of London task. *Neuroimage* **2006**, *33*, 999–1010. [CrossRef]
20. Richter, W.; Richter, M. The shape of the fMRI BOLD response in children and adults changes systematically with age. *NeuroImage* **2003**, *20*, 1122–1131. [CrossRef]
21. Patriat, R.; Molloy, E.K.; Meier, T.B.; Kirk, G.R.; Nair, V.A.; Meyerand, M.E.; Prabhakaran, V.; Birn, R.M. The effect of resting condition on resting-state fMRI reliability and consistency: A comparison between resting with eyes open, closed, and fixated. *Neuroimage* **2013**, *78*, 463–473. [CrossRef]
22. Beckmann, C.F.; Smith, S.M. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* **2004**, *23*, 137–152. [CrossRef]
23. Linke, A.C.; Olson, L.; Gao, Y.; Fishman, I.; Müller, R.A. Psychotropic medication use in autism spectrum disorders may affect functional brain connectivity. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **2017**, *2*, 518–527. [CrossRef] [PubMed]
24. Bartolotti, J.; Sweeney, J.A.; Mosconi, M.W. Functional brain abnormalities associated with comorbid anxiety in autism spectrum disorder. *Dev. Psychopathol.* **2020**, *32*, 1273–1286. [CrossRef] [PubMed]
25. Antshel, K.M.; Zhang-James, Y.; Wagner, K.E.; Ledesma, A.; Faraone, S.V. An update on the comorbidity of ADHD and ASD: A focus on clinical management. *Expert Rev. Neurother.* **2016**, *16*, 279–293. [CrossRef] [PubMed]
26. Lai, M.C.; Kassee, C.; Besney, R.; Bonato, S.; Hull, L.; Mandy, W.; Szatmari, P.; Ameis, S.H. Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *Lancet Psychiatry* **2019**, *6*, 819–829. [CrossRef]
27. Power, J.D.; Barnes, K.A.; Snyder, A.Z.; Schlaggar, B.L.; Petersen, S.E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **2012**, *59*, 2142–2154. [CrossRef]
28. Evans, A.C.; Collins, D.L.; Mills, S.; Brown, E.D.; Kelly, R.L.; Peters, T.M. 3D statistical neuroanatomical models from 305 MRI volumes. In *Proceedings of the 1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference, San Francisco, CA, USA, 31 October 1993–6 November 1993* ; IEEE: Piscataway, NJ, USA, 1993; pp. 1813–1817.
29. Heunis, S. jsheunis/fMRwhy: Release 0.0.2—for Referencing, 2021. Available online: https://zenodo.org/records/4527533 (accessed on 4 November 2023). [CrossRef]
30. Jenkinson, M.; Beckmann, C.F.; Behrens, T.E.; Woolrich, M.W.; Smith, S.M. FSL. *NeuroImage* **2012**, *62*, 782–790. [CrossRef]
31. Penny, W.D.; Friston, K.J.; Ashburner, J.T.; Kiebel, S.J.; Nichols, T.E. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*; Elsevier: Amsterdam, The Netherlands, 2011.
32. Ashburner, J.; Barnes, G.; Chen, C.C.; Daunizeau, J.; Flandin, G.; Friston, K.; Gitelman, D.; Glauche, V.; Henson, R.; Hutton, C.; et al. *SPM12 Manual*; Wellcome Centre for Human Neuroimaging, UCL: London, UK, 2021. Available online: https://www.fil.ion.ucl.ac.uk/spm/doc/spm12_manual.pdf (accessed on 5 November 2023).
33. Pruim, R.H.; Mennes, M.; van Rooij, D.; Llera, A.; Buitelaar, J.K.; Beckmann, C.F. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **2015**, *112*, 267–277. [CrossRef]
34. Byrge, L.; Kennedy, D.P. Identifying and characterizing systematic temporally-lagged BOLD artifacts. *NeuroImage* **2018**, *171*, 376–392. [CrossRef]
35. Friston, K.J. Functional and effective connectivity: A review. *Brain Connect.* **2011**, *1*, 13–36. [CrossRef]
36. Friston, K.J.; Harrison, L.; Penny, W. Dynamic causal modelling. *Neuroimage* **2003**, *19*, 1273–1302. [CrossRef]
37. Blazejewska, A.I.; Fischl, B.; Wald, L.L.; Polimeni, J.R. Intracortical smoothing of small-voxel fMRI data can provide increased detection power without spatial resolution losses compared to conventional large-voxel fMRI data. *NeuroImage* **2019**, *189*, 601–614. [CrossRef] [PubMed]
38. Balachandrasekaran, A.; Cohen, A.L.; Afacan, O.; Warfield, S.K.; Gholipour, A. Reducing the effects of motion artifacts in fMRI: A structured matrix completion approach. *IEEE Trans. Med. Imaging* **2021**, *41*, 172–185. [CrossRef] [PubMed]
39. Ciric, R.; Wolf, D.H.; Power, J.D.; Roalf, D.R.; Baum, G.L.; Ruparel, K.; Shinohara, R.T.; Elliott, M.A.; Eickhoff, S.B.; Davatzikos, C.; et al. Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *Neuroimage* **2017**, *154*, 174–187. [CrossRef] [PubMed]

40. Murphy, K.; Fox, M.D. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *Neuroimage* **2017**, *154*, 169–173. [CrossRef]
41. Margulies, D.S.; Böttger, J.; Long, X.; Lv, Y.; Kelly, C.; Schäfer, A.; Goldhahn, D.; Abbushi, A.; Milham, M.P.; Lohmann, G.; et al. Resting developments: A review of fMRI post-processing methodologies for spontaneous brain activity. *Magn. Reson. Mater. Physics Biol. Med.* **2010**, *23*, 289–307. [CrossRef]
42. Boubela, R.N.; Kalcher, K.; Huf, W.; Kronnerwetter, C.; Filzmoser, P.; Moser, E. Beyond noise: Using temporal ICA to extract meaningful information from high-frequency fMRI signal fluctuations during rest. *Front. Hum. Neurosci.* **2013**, *7*, 168. [CrossRef]
43. Bernas, A.; Aldenkamp, A.P.; Zinger, S. Wavelet coherence-based classifier: A resting-state functional MRI study on neurodynamics in adolescents with high-functioning autism. *Comput. Methods Programs Biomed.* **2018**, *154*, 143–151. [CrossRef]
44. Filippini, N.; MacIntosh, B.J.; Hough, M.G.; Goodwin, G.M.; Frisoni, G.B.; Smith, S.M.; Matthews, P.M.; Beckmann, C.F.; Mackay, C.E. Distinct patterns of brain activity in young carriers of the APOE-ε4 allele. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 7209–7214. [CrossRef]
45. Schielen, S.J.; Pilmeyer, J.; Aldenkamp, A.P.; Ruijters, D.; Zinger, S. ICA-based Resting-State Networks Obtained on Large Autism fMRI Dataset ABIDE. *arXiv* **2024**, arXiv:2412.13798.