



Federated Analysis in COINSTAC Reveals Functional Network Connectivity and Spectral Links to Smoking and Alcohol Consumption in Nearly 2,000 Adolescent Brains

Harshvardhan Gazula¹ · Kelly Rootes-Murdy² · Bharath Holla^{3,44} · Sunitha Basodi² · Zuo Zhang⁴ · Eric Verner² · Ross Kelly² · Pratima Murthy³ · Amit Chakrabarti⁵ · Debasish Basu⁶ · Subodh Bhagyalakshmi Nanjaya⁶ · Rajkumar Lenin Singh⁷ · Roshan Lourembam Singh⁸ · Kartik Kalyanram⁹ · Kamakshi Kartik⁹ · Kumaran Kalyanaraman¹⁰ · Krishnaveni Ghattu¹⁰ · Rebecca Kuriyan¹¹ · Sunita Simon Kurpad¹² · Gareth J Barker¹³ · Rose Dawn Bharath¹⁴ · Sylvane Desrivieres⁴ · Meera Purushottam¹⁵ · Dimitri Papadopoulos Orfanos¹⁶ · Eesha Sharma¹⁷ · Matthew Hickman¹⁸ · Mireille Toledano¹⁹ · Nilakshi Vaidya^{20,21} · Tobias Banaschewski²² · Arun L. W. Bokde²³ · Herta Flor^{24,25} · Antoine Grigis¹⁶ · Hugh Garavan²⁶ · Penny Gowland²⁷ · Andreas Heinz²⁸ · Rüdiger Brühl²⁹ · Jean-Luc Martinot³⁰ · Marie-Laure Paillère Martinot^{30,31} · Eric Artiges^{30,32} · Frauke Nees^{21,22,24} · Tomás Paus^{33,34} · Luise Poustka³⁵ · Juliane H. Fröhner³⁶ · Lauren Robinson³⁷ · Michael N. Smolka³⁶ · Henrik Walter²⁸ · Jeanne Winterer^{28,38} · Robert Whelan³⁹ · IMAGEN Consortium · Jessica A. Turner^{2,40} · Anand D. Sarwate⁴¹ · Sergey M. Plis^{2,42} · Vivek Benegal³ · Gunter Schumann^{4,21,43} · Vince D. Calhoun^{2,40}

Accepted: 29 August 2022 / Published online: 25 November 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

With the growth of decentralized/federated analysis approaches in neuroimaging, the opportunities to study brain disorders using data from multiple sites has grown multi-fold. One such initiative is the Neuromark, a fully automated spatially constrained independent component analysis (ICA) that is used to link brain network abnormalities among different datasets, studies, and disorders while leveraging subject-specific networks. In this study, we implement the neuromark pipeline in COINSTAC, an open-source neuroimaging framework for collaborative/decentralized analysis. Decentralized exploratory analysis of nearly 2000 resting-state functional magnetic resonance imaging datasets collected at different sites across two cohorts and co-located in different countries was performed to study the resting brain functional network connectivity changes in adolescents who smoke and consume alcohol. Results showed hypoconnectivity across the majority of networks including sensory, default mode, and subcortical domains, more for alcohol than smoking, and decreased low frequency power. These findings suggest that global reduced synchronization is associated with both tobacco and alcohol use. This proof-of-concept work demonstrates the utility and incentives associated with large-scale decentralized collaborations spanning multiple sites.

Keywords COINSTAC · Decentralized analysis · Neuromark · IMAGEN · CVEDA · Adolescent health

Introduction

In the past decade, the field of neuroimaging has seen a rapid growth in data pooling/sharing initiatives (Eickhoff et al., 2016). This, in turn, has resulted in a major transformation in

our understanding of brain disorders as well as accelerating studies to identify brain-based markers for potential use in clinical settings (Wilcox et al., 2020). However, open data sharing can be limited by policy or proprietary restrictions or data re-identification concerns (Shringarpure & Bustamante, 2015; Sweeney, 2002; White et al., 2020). Such difficulties are further compounded in the case of sensitive patient data that cannot be easily shared or accessed. Yet another challenge is pooling the data centrally can require significant computational resources which may not be available at a given site.

Harshvardhan Gazula and Kelly Rootes-Murdy contributed equally to this work.

✉ Harshvardhan Gazula
hvgazula@umich.edu

✉ Bharath Holla
hollabharath@gmail.com

Extended author information available on the last page of the article

The aforementioned barriers can be offset by embracing federated analysis of neuroimaging data. Federated analysis does not require moving the data from the original sites but can generate results similar to pooled data. It can also be implemented with varying degrees of privacy, ranging from sharing of derived results only to full encryption. Towards this end, Plis et al. (2016) introduced a federated analysis framework and platform called the COllaborative Informatics and Neuroimaging Suite Toolkit for Anonymous Computation (COINSTAC; <http://coinstac.trendscenter.org>). There is no sharing of individual-level data in COINSTAC and thus the privacy of individual datasets is preserved while still allowing analyses to be performed. Multiple neuroimaging algorithms have been federated and can be run in COINSTAC. Examples of implemented algorithms include decentralized voxel-based morphometry (Gazula et al., 2018), decentralized t-distributed stochastic neighbor embedding (Saha et al., 2017, 2021), decentralized dynamic functional network connectivity (Baker et al., 2020), and decentralized support vector machine with differential privacy (Sarwate et al., 2014).

For the purpose of this study, we focus on decentralized variants of independent component analysis. Independent component analysis (ICA) is a versatile and powerful data-driven approach that can be used to analyze group fMRI data (Calhoun et al., 2009) to study the spatio-temporal structure of the fMRI signal. One of the earliest decentralized versions of ICA developed, for use with functional magnetic resonance imaging (fMRI) data, was the decentralized temporal ICA algorithm (Baker et al., 2015). More recently, (Baker et al., 2020) presented a decentralized group ICA algorithm to enable decentralized dynamic functional network connectivity analysis. Group ICA (Calhoun et al., 2009) enables group-level inferences while allowing for cross-subject variability. However, implementing group ICA within a decentralized context presents some challenges due to its fully data-driven output. For example, results typically require labelling of the output components, and while there are options for automated labelling (Salman et al., 2021) within the GIFT software (<http://trendscenter.org/software/gift>) the use of a data-driven approach involves considerable user interactions which can be inefficient in a federated environment. In addition, separate group ICA results may be challenging to compare with one another, hindering replication and cross-study comparison. To address this issue, Du et al. (2020) proposed the NeuroMark pipeline, an *a priori* driven and fully automated ICA approach informed by reliable network templates to achieve linked analyses among different datasets, studies, and disorders.

The decentralized NeuroMark approach will be the subject of discussion in this work. Our focus is on the impact of substance use on adolescent brain functional network connectivity. Recently, Gazula et al. (2021) performed a large-N decentralized voxel-based morphometry analysis of structural

magnetic resonance imaging data across two cohorts from 14 different sites to understand the structural changes in the brain as linked to age, body mass index and smoking. However, there has not yet been a large-N analysis of functional MRI data with ICA within a decentralized framework. In this work, we extend the aforementioned structural MRI results by presenting an exploratory analysis of functional network connectivity (FNC) and spectral analysis derived from independent component analysis on resting fMRI data. Our contributions in this paper can be summarized as follows.

1. Decentralized ICA on large datasets across multiple sites, IMAGEN from the Europe and cVEDA from India, in the COINSTAC framework and sharing observations.
2. Evaluating the association of smoking and alcohol use on functional network connectivity in the adolescent brain.
3. Decentralized MANCOVAN and univariate testing of ICA output.

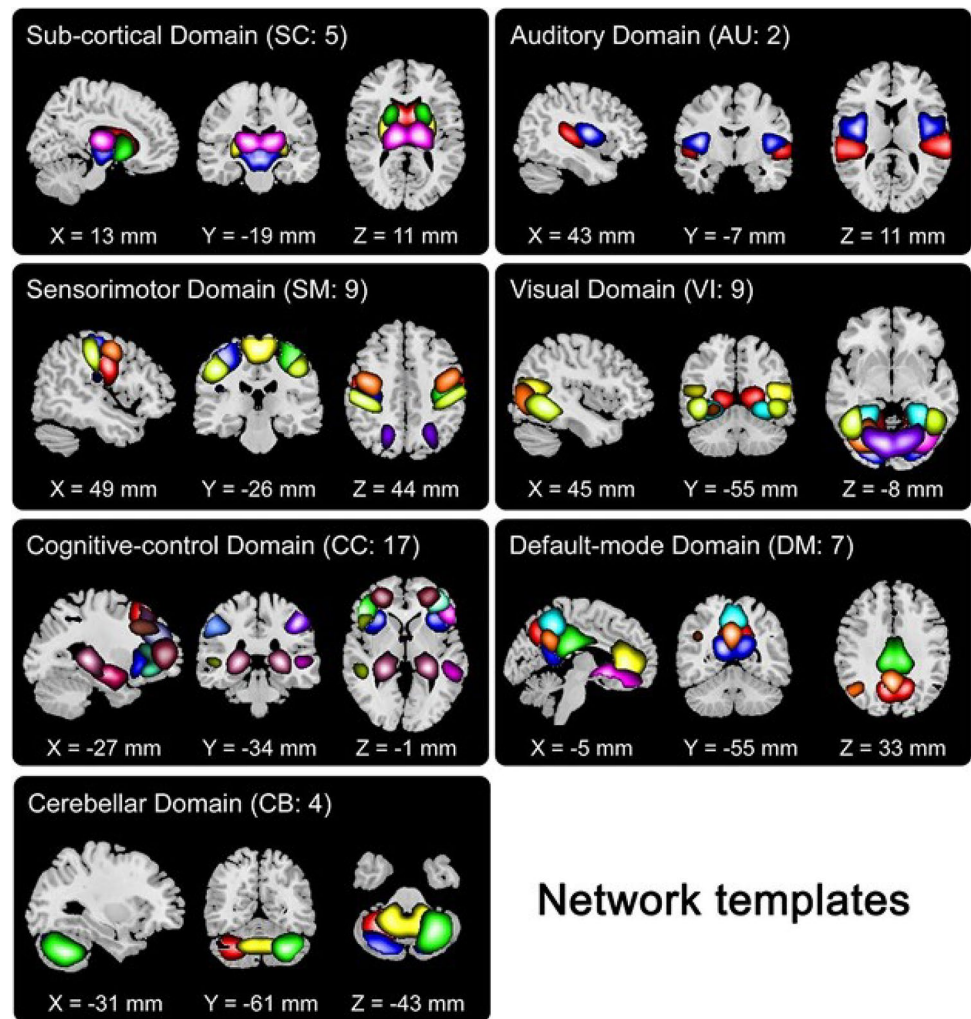
The outline of the current paper is as follows: In “**Methods**”, we discuss the Decentralized NeuroMark pipeline. In “**Data**”, we describe the IMAGEN and cVEDA data used in this study. In “**Results**” and “**Discussion**”, we present the results and discuss our experience implementing the NeuroMark pipeline in COINSTAC. We conclude the paper in “**Conclusion**”.

Methods

As mentioned earlier, the goal of group ICA is to enable group level inferences by allowing for subject-level variability. However, due to the data-driven nature of ICA, these inferences can turn out to be inconsistent when compared across different studies containing data from similar disorders. It is possible to systematically integrate these findings across studies and disorders by interlinking the estimation of subject-level spatial maps using a common set of spatial priors and NeuroMark (Du et al., 2020) precisely achieves this objective. The goal of NeuroMark is to fully automate the estimation and labeling of individual subject connectivity features using spatial network priors derived from independent large samples. Another advantage of NeuroMark is that the ICA estimates are computed on individual subjects, which completely avoids data leakage, facilitating cross-validation and independent analysis. The estimation of subject specific networks is performed by leveraging spatially-constrained ICA (Lin et al., 2010; Du & Fan, 2013).

The NeuroMark pipeline leverages component priors derived from replicable spatial maps estimated from large independent datasets. For this purpose, resting-state fMRI datasets from the Human Connectome Project (HCP,

Fig. 1 Domains identified in Du *et al.* Briefly, these seven identified network templates were divided based on anatomical and functional properties. In each subfigure, one color in the composite maps corresponds to an intrinsic connectivity network (ICN)



<http://www.humanconnectomeproject.org/data/>) and the Genomics Superstruct Project (Holmes et al., 2015) were utilized. Group ICA was applied separately to the two datasets, followed by a greedy spatial correlation analysis (coupled with expert inspection) to identify and label intrinsic connectivity networks (ICNs) that showed high replicability across the two separate group ICA analyses (Du et al., 2020) see Figs. 1 and 2.

For the Neuromark pipeline, these ICNs are used as network templates (aka priors) to estimate subject-specific functional networks and their associated time-courses (TCs) via an adaptive spatially constrained ICA method. Here we used multi-objective optimization ICA where one objective is to optimize the independence of networks (for the subject), while the other is to optimize the similarity between one subject-specific network and its related network template.

Once this is accomplished, various network features can be extracted from both static and dynamic perspectives. One such example of a measure is the static FNC (sFNC)

which is computed as the Pearson correlations between TCs of ICNs to yield an sFNC matrix reflecting the interaction between any two networks. While the spatial map of each ICN reflects intra-connectivity within brain functional network, sFNC matrix represents inter-connectivity strengths between different ICNs. The obtained network features and time courses can be further analysed statistically via decentralized MANCOVA for making inferences.

Decentralized ICA stats

Once the network features are derived from Neuromark, the researcher/user can perform any statistical analysis in a decentralized fashion. This helps in speeding up the analysis enabling multiple research groups to participate in larger studies, preserving their data privacy without sharing the actual data. Fig. 3 shows how the Neuromark pipeline is performed in a centralized way. There are two ways in which the decentralized statistics can be performed. Here we implement the MANCOVAN toolbox in GIFT (<https://>

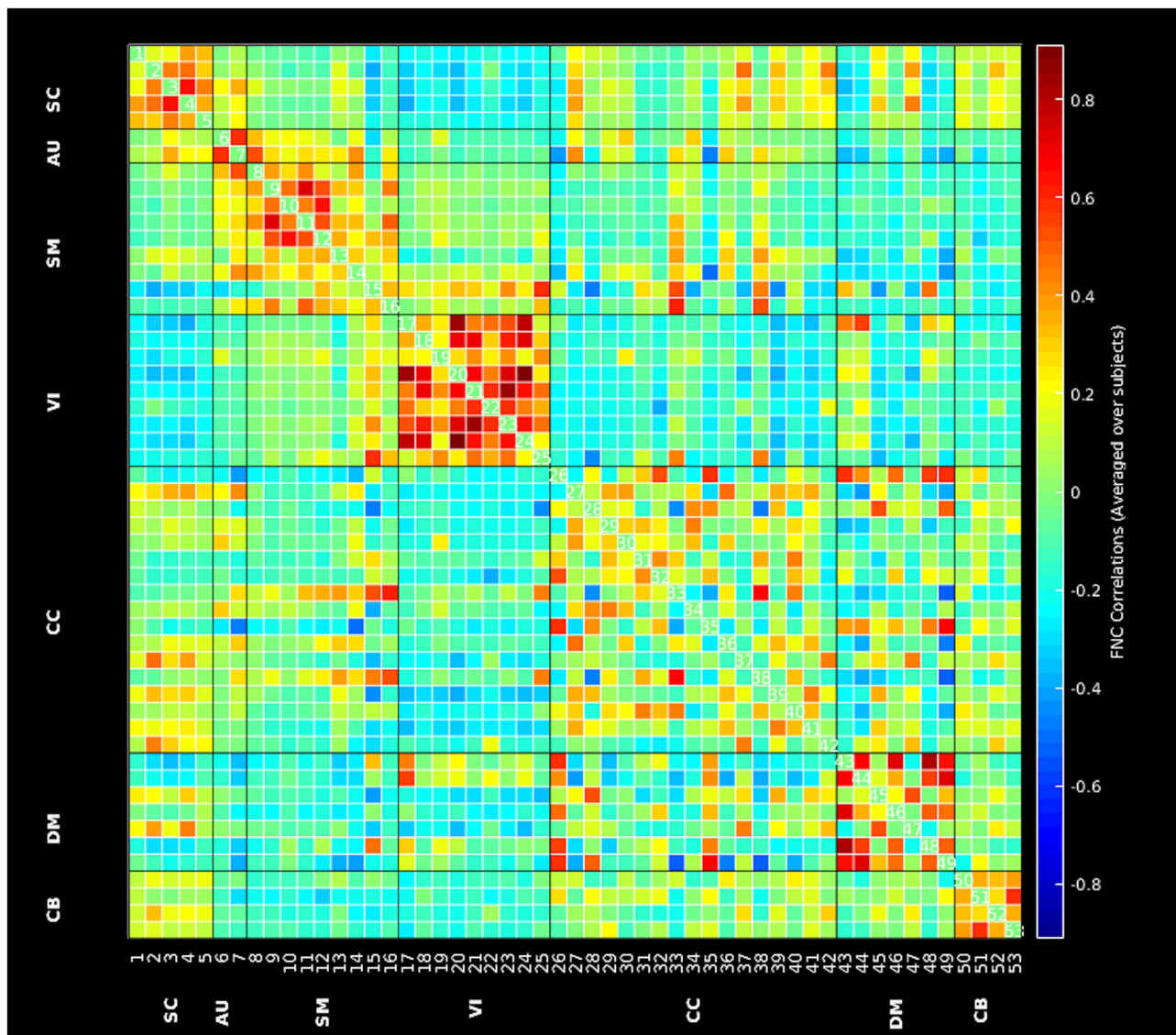


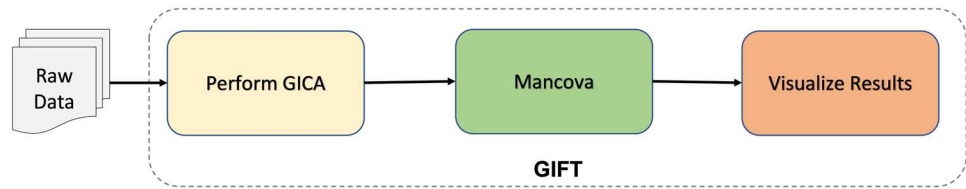
Fig. 2 Correlation map showing average correlation, across participants, for each domain. Cool colors represent anticorrelation and warm colors represent positive correlation. In general, there was high positive ($r > .5$) correlation within the visual, sensorimotor,

and default mode network domains across all participants. Of note. SC: sub-cortical domain, AUD: auditory domain, SM: sensorimotor domain, VIS: visual domain, CC: cognitive-control domain, DMN: default-mode domain, CR: cerebellar domain

trendscenter.org/software/gift/), which provides both MANCOV omnibus tested as well as univariate regression tests. In this work, we utilize the univariate regression functionality, though the MANCOV toolbox (MTB) is fully implemented. The first approach, *dMANCOV* (decentralized MANCOV), involves pooling all the time courses within a private aggregator (the application) and running the MTB on these pooled data. The results are then shared with every local site (Algorithm 1 and Fig. 4). The second and more secure approach *dMANCOV_{pa}* (decentralized MANCOV with private aggregator) involves performing an MTB analysis at each local site separately on

its corresponding timecourses and covariate data. From the local MTB results generated, relevant stats such as residual errors, contrast mean images, and other metadata (such as number of subjects) are extracted and sent to the private aggregator to compute global sum of squares residual errors and average beta weights. From these, the corresponding t -values and p -values are computed and sent back to all the local sites. Alg. 2 and Fig. 5 show the algorithm and pipeline corresponding to the improved decentralized MTB approach respectively. Alg. 2 is a decentralized multivariate analysis of covariance (a type of generalized linear model).

Fig. 3 Pipeline of GIFT framework computed in a centralized way



Algorithm 1: Decentralized GICA and MANCOVA

Data: $\mathbf{X}_i^m \in R$ be raw sMRI/fMRI data of m subjects available at site $i \in \{0, 1, 2, \dots, N\}$, and $\mathbf{Y}_i^k \in R$ representing k covariate information.

```

/* Local sites */
1 for all sites  $i$  in  $1 \dots N$  do
2   Perform group independent component analysis on the group of subjects ( $\mathbf{X}_i^m$ )
3   Send time courses information  $\mathbf{TC}_i^m$ , covariate files  $\mathbf{Y}_i^k$  and other metadata to the aggregator site
4 end
/* aggregator site */
5 if site  $i == 0$  then
6   Gather time courses and covariate files from all local sites  $i=1..N$ 
7   Combine time courses  $\mathbf{TC}_{agg} = \text{Aggregate}(\mathbf{TC}_i^m), i \in \{1, 2, 3, \dots, N\}$ 
8   Combine covariates  $\mathbf{Y}_{agg} = \text{Aggregate}(\mathbf{Y}_i^k), i \in \{1, 2, 3, \dots, N\}$ 
9   Perform mancova analysis  $\mathbf{M}_{agg}$  with  $\mathbf{TC}_{agg}$ ,  $\mathbf{Y}_{agg}$  and other metadata
10  Send the analysis results to all the local sites  $i \in \{1, 2, \dots, N\}$ 
11 end

```

Algorithm 2: Decentralized GICA and MANCOVA with Private Aggregator

Data: $\mathbf{X}_i^m \in R$ be raw sMRI/fMRI data of m subjects available at site $i \in \{0, 1, 2, \dots, N\}$, and $\mathbf{Y}_i^k \in R$ representing k covariate information.

```

/* Local sites */
1 for all sites  $i$  in  $1 \dots N$  do
2   Perform group independent component analysis on the group of subjects ( $\mathbf{X}_i^m$ )
3   Perform MANCOVAN analysis  $\mathbf{M}_i$  with  $\mathbf{TC}_i$ ,  $\mathbf{Y}_i$  and other metadata
4   Extract stats information (i.e., residual errors, contrast, and mean images) and other metadata (such as the number of subjects), save it as a file  $\mathbf{S}_i$ , and send it to the aggregator site.
5 end
/* aggregator site */
6 if site  $i == 0$  then
7   Gather stats  $\mathbf{S}_i$  files from all local sites,  $\mathbf{S}_{agg} = \text{Aggregate}(\mathbf{S}_i), i \in \{1, 2, 3, \dots, N\}$ 
8   Compute the global sum of squares residual errors and average beta weights.
9   Estimate the p – values from the aggregated stats  $\mathbf{S}_{agg}$ .
10  Compute and send the analysis results to all the local sites  $i \in \{1, 2, \dots, N\}$ .
11 end

```

COINSTAC

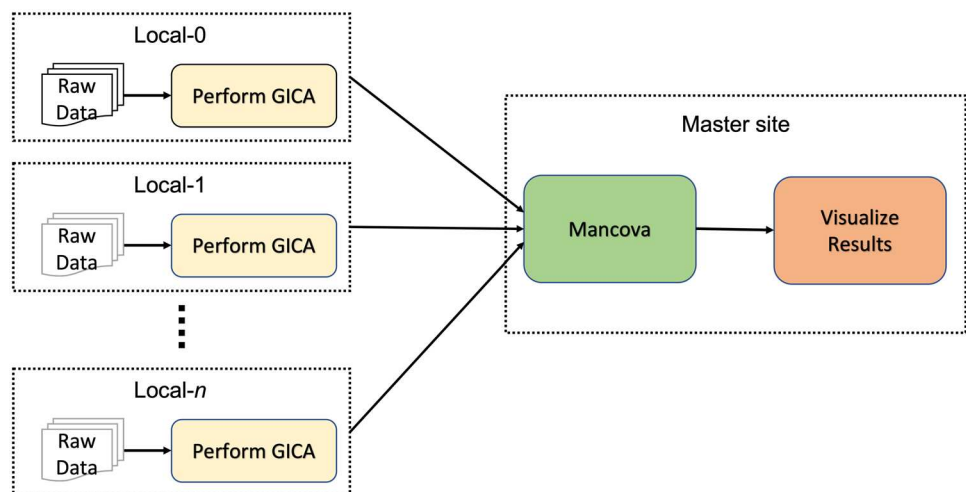
For this project, the entire analysis, neuromark and the decentralized statistical analysis, was performed in COINSTAC. In brief, COINSTAC (Gazula et al., 2020; Ming et al., 2017; Plis et al., 2016) is a fully open-source, federated learning platform geared towards neuroimaging research enabling researchers to perform decentralized analysis of neuroimaging data in a secure and privacy-preserving environment via a message-passing infrastructure. More specific instructions on how to install the application and run an example analysis can be found online at <https://github.com/trendscenter/coinstac-instructions>. In COINSTAC, sites perform computations on their local data and intermediate results are collected by a cloud-based “private aggregator” which synthesizes the results and returns them to the sites.

Data

Cohort

The data from two large multi-centric adolescent cohorts in India and Europe were used for the present analysis. The Consortium on Vulnerability to Externalising Disorders and

Fig. 4 Decentralized GICA and mancova implemented in COINSTAC using GIFT toolbox. This approach involves pooling all the time courses within a private aggregator and running the MTB on this pooled data following which the results are shared with every local site



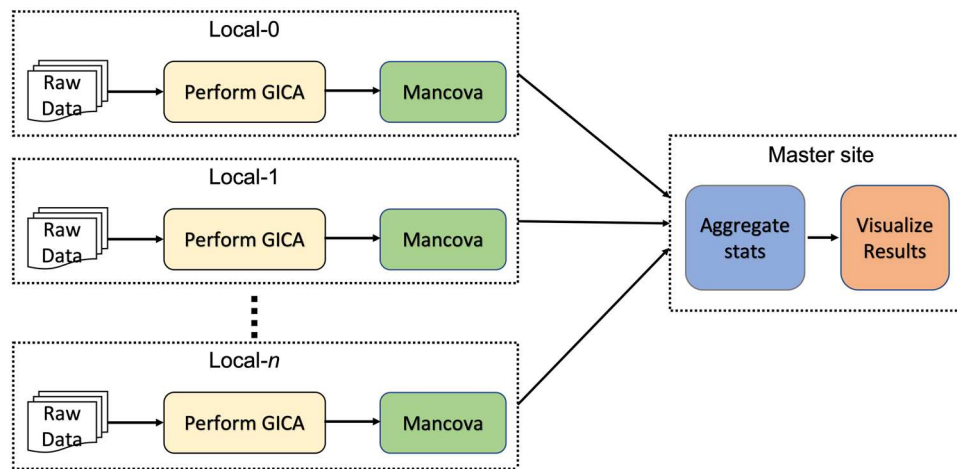


Fig. 5 This version of decentralized Mancova involves performing an MTB analysis at each local site separately on its corresponding timecourses and covariate data. From the local MTB results generated, relevant stats information such as residual errors, contrast mean images, and other metadata (such as number of subjects) are extracted

and sent to the master site. At the master site, these stats and meta-data from all the local sites are combined to compute global sum of squares residual errors and average beta weights. From these, t-values and p-values are computed and these results are sent to all the local sites

Addiction (cVEDA) is an accelerated longitudinal cohort in India (Sharma et al., 2020) that covers an age span of 5-24 years. Participants who provided brain imaging data were recruited at six study sites (see Zhang et al. (2020) for cohort profile and Sharma et al. (2020) for study protocol). The IMAGEN project is a longitudinal study of adolescent brain development and mental health in Europe (Schumann et al., 2010). Participants were recruited from eight study sites in England, Ireland, France, and Germany at the age of 14, with follow-up assessments at ages 16, 19, and 22. Each study site obtained ethical approval from the local research ethics committee. For both studies written consent (or verbal) assent were acquired from all the participants and parents prior to participation.

In this research, we used resting functional neuroimaging, and smoking and drinking assessments from the baseline cVEDA cohort ($n=1140$) and the second follow-up wave in IMAGEN ($n=839$), acquired at age 19. See Table 1 for the cohort description.

MRI Dataset Acquisition and Preprocessing

Although COINSTAC also implements fMRI preprocessing, for this study data was previously preprocessed by each site separately outside of COINSTAC, and we focused on using COINSTAC to implement a group ICA + stats analysis. The preprocessing steps are briefly described below. For a complete description of the preprocessing of the IMAGEN dataset see Schumann et al. (2010) and for the cVEDA dataset see (Sharma et al., 2020; Zhang et al., 2020).

For IMAGEN, data preprocessing was performed using the statistical parametric mapping (SPM) tool (<http://www.fil.ion.ucl.ac.uk/spm/>) toolbox. fMRI BOLD images were first realigned, corrected for head motion and slice timing (to account for timing difference in slice acquisition) and co-registered to T1w (MPRAGE) images. The fMRI data were normalized to the EPI template and resampled to 3 mm^3 isotropic voxels. Images were spatially smoothed using a Gaussian kernel with a 6 mm full width at half maximum. Subjects with head

Table 1 Sample Characteristics

Variable	cVEDA (N=1140)	IMAGEN (N=839)	Total (N=1979)	<i>t</i>	χ^2	<i>p</i>
Age [y, Mean (SD)]	15.677 (4.346)	18.998 (0.718)	17.085 (3.714)	-3.321		< 0.001
Sex [Male, n (%)]	638 (56.0%)	389 (46.4%)	1027 (51.9%)		17.8429	2.4e-5
BMI [kg/m^2 , Mean (SD)]	19.237 (4.509)	22.836 (3.946)	20.763 (4.634)	-3.599		< 0.001
Smoking Risk, n (%): 0	1073 (94.1%)	242 (28.8%)	1315 (66.4%)		923.7725	< 0.00001
1	67 (5.9%)	597 (71.2%)	664 (33.6%)			
Alcohol Risk, n (%): 0	1088 (95.4%)	224 (26.7%)	1312 (66.3%)		1022.0692	< 0.00001
1	43 (3.8%)	516 (61.5%)	559 (28.2%)			
2	9 (0.8%)	99 (11.8%)	108 (5.5%)			

motion < 0.25 mm (frame-wise displacement) and with functional data providing near full brain successful normalization (by comparing the individual mask with the group mask) were selected for further analysis.

For cVEDA, the parameters and acquisition protocols were similar to those described above in IMAGEN. More detailed information can be found at <https://cveda-project.org/standard-operating-procedures/>. The functional data processing for cVEDA in brief included the following steps. Motion correction by applying a rigid body registration of each volume to the middle volume (FSL MCF-LIRT). Slice-timing correction to account for timing difference in slice acquisition. Non-brain tissue removal (FSL BET) and co-registration to high-resolution T1 image (FSL FLIRT using the BBR algorithm). Motion correcting transformations, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using Advanced Normalization Toolbox (ANTs v2.1.0) and Lanczos interpolation. An ICA-based Automatic Removal Of Motion Artifacts (AROMA) was used to generate non-aggressively denoised data. And finally these data were resampled to 3mm isotropic and smoothed using a 6mm non-linear filter using FSL SUSAN. Only those subjects who had head motion FD of < 0.25 mm were retained.

Measure of Tobacco and Alcohol use

For IMAGEN, measures of tobacco smoking were obtained from self-reports in the Fagerstrom test for nicotine dependence (FTND) questionnaire (Heatherton et al., 1991). The answer to the question regarding lifetime smoking experience was used to create a binary variable, indicating whether the participant had ever smoked (labeled as 1) or not (labeled as 0). For alcohol risk, the Alcohol Use Disorders Identification Test–Consumption (AUDIT-C) score < 3 for females or < 4 for males were labeled as 0 (low risk), score of 3–7 for females or 4–7 for males were labeled 1 (moderate risk) and score ≥ 8 was labelled as 2 (high risk).

For cVEDA, measures of tobacco smoking and alcohol use were obtained from the World Health Organization's Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) questionnaire, which provides a specific substance involvement score indicating the risk levels. For smoking risk, a tobacco involvement score of > 3 was labeled as 1 (present) and a score of ≤ 3 was labeled as 0 (absent). For alcohol risk, an involvement score of ≤ 3 was labeled 0 (low risk), a score 4–26 was labeled 1 (moderate risk) and ≥ 27 was labeled as 2 (high risk).

Other covariates from both the cohorts were mean-centered indices of head motion during resting state

functional MRI (measured as framewise displacement in mm), body mass (BMI) (weight in kilograms divided by the squared height in meters), age in years, and dummy-codes for site and sex. (refer to Table 1 for sample characteristics).

Results

We provide a summary of the results here and then discuss them further in the following section. Please note that all the maps have been thresholded with a false discovery rate (FDR) correction of 0.05 (Benjamini & Hochberg, 1995).

FNC for Alcohol

Lifetime risk for alcohol showed more mean functional connectivity strengths than lifetime risk for tobacco. Across the seven domains (default mode network (DMN), subcortical (SC), auditory (AUC), sensorimotor (SM), visual (VIS), cognitive-control (CC), & cerebellar (CR)), there were roughly 27 positive pairs and 48 negative pairs, the SC domain was most positively associated with the other domains (8 positive pairs) and the SM domain and the CC domains were the most negatively associated with the other domains (20 and 21 pairs, respectively). Increased functional connectivity was observed between the DMN and cognitive control. See Fig. 6 for more details.

FNC for Tobacco

In addition to less overall connectivity strength, lifetime risk for tobacco showed almost exclusive negative connectivity pairs across the seven domains. To a lesser extent than alcohol, the DMN domain had increased functional connectivity with the cognitive control domain, a total of four positive connectivity pairs. The DMN also had one positive connectivity pair with the cerebellar domain and the visual domain had one positive connectivity pair with the sensorimotor domain. All of the other 27 connectivity pairs were negatively associated. See Fig. 7 for more details.

For both lifetime risk of alcohol and tobacco, there was a predominant decrease in a number of the connections, suggesting a common reduction of connection effect for both substances. Taken together, there was a decreased functional network connectivity observed in the sensorimotor, auditory, and visual networks in both alcohol and tobacco use.

The univariate results showed reduced power in the high frequency region for lifetime alcohol and tobacco use (both above 0.15 Hz) with larger changes for alcohol than tobacco. Alcohol use also showed increased power in the mid/BOLD

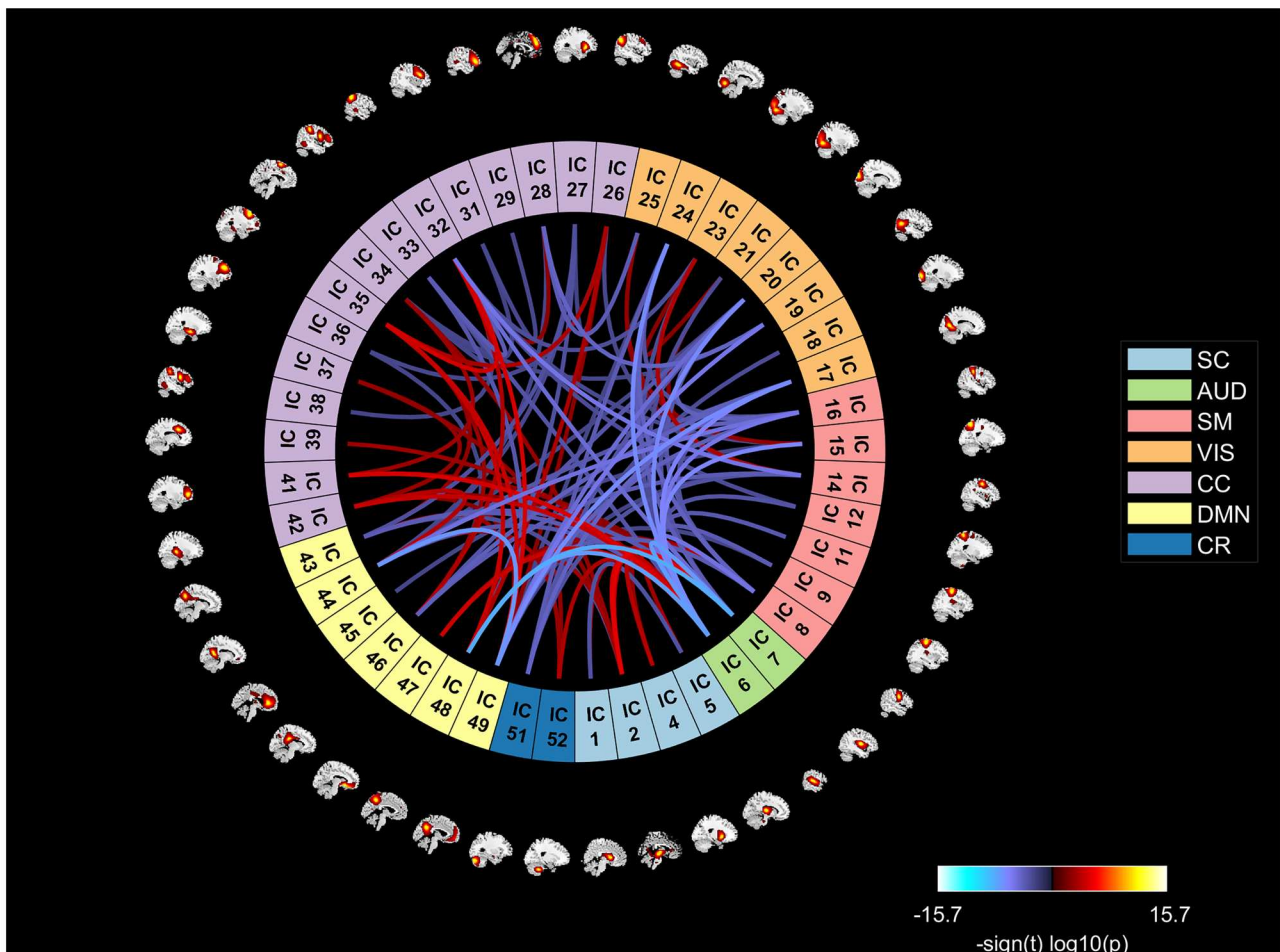


Fig. 6 Connectogram showing group mean connectivity across domains for alcohol users. Cool colors represent reduced connectivity and warm colors represent increased connectivity. In alcohol users, there was increased connectivity mostly linked to the DMN

domains and to a lesser extent, the cognitive control and sub-cortical domains. Of note. SC: sub-cortical domain, AUD: auditory domain, SM: sensorimotor domain, VIS: visual domain, CC: cognitive-control domain, DMN: default-mode domain, CR: cerebellar domain

frequency range, see Figs. 8 and 9 for more details. As a whole, alcohol use indicated two sets of signal frequencies, reduced power (roughly 0.2 Hz) at high frequency and high power (between 0.05 - 0.1 Hz) at low frequency.

Decentralized MANCOVA

We test our approach using the subset of the fMRI dataset used in Allen et al. (2011). Since, it is a large dataset, we select a subset of 50 subjects and run mancova algorithm as shown in Fig. 3. The same data is partitioned into two subsets to test our decentralized mancova analysis using algorithms Alg. 1 and Alg. 2. In our results, we compare FNC correlations and significant effects of gender outputs across different algorithms.

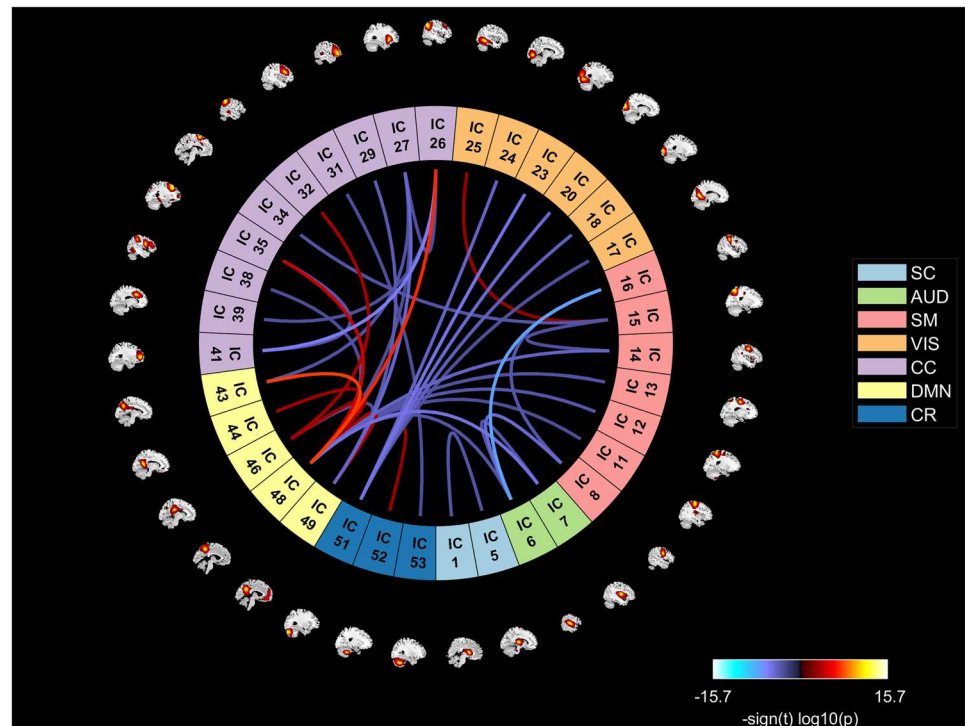
Figures 10, 11 and 12 show the FNC correlations, significant effects of gender and effect sizes per gender. These are the general studies performed using mancova analysis.

From the figures, it can be seen that Alg. 1 has same output as the GIFT interface in all the figures. Alg. 2 has the same output for FNC correlations (Fig. 10b, c) but varies slightly in terms of significant effects of gender (Figs. 11b, c and 12b, c). This is because the mancova output from all the local sites were used in computing mancova output in Alg. 2 rather than computing mancova at the remote site as in Alg. 1. This improvisation in Alg. 2 reduces running time by 50% as compared to Alg. 1.

Discussion

The goal of this paper was two-fold: to demonstrate the feasibility of COINSTAC (Gazula et al., 2018) for performing decentralized analysis on large datasets present across multiple sites, compare the results of the decentralized ICA and Neuromark, and to use this experiment to evaluate the

Fig. 7 Connectogram showing group mean connectivity across domains for tobacco users. Cool colors represent reduced connectivity and warm colors represent increased connectivity. In tobacco users, there was decreased connectivity linked to sensorimotor, auditory, and visual domains. Of note. SC: sub-cortical domain, AUD: auditory domain, SM: sensorimotor domain, VIS: visual domain, CC: cognitive-control domain, DMN: default-mode domain, CR: cerebellar domain



relationships of both smoking and tobacco use with functional connectivity changes. Using a large sample of tobacco and alcohol users, this study investigated resting state functional connectivity in seven brain networks (SC, AUD, SM, VIS, CC, DMN, & CR). Using previously-defined networks (Du

et al., 2020) we identified a trend of general decreased connectivity across all networks in both tobacco and alcohol users.

Starting with tobacco users, there was decreased connectivity in most of the pairings but the most prominent result was hypoconnectivity between the auditory and

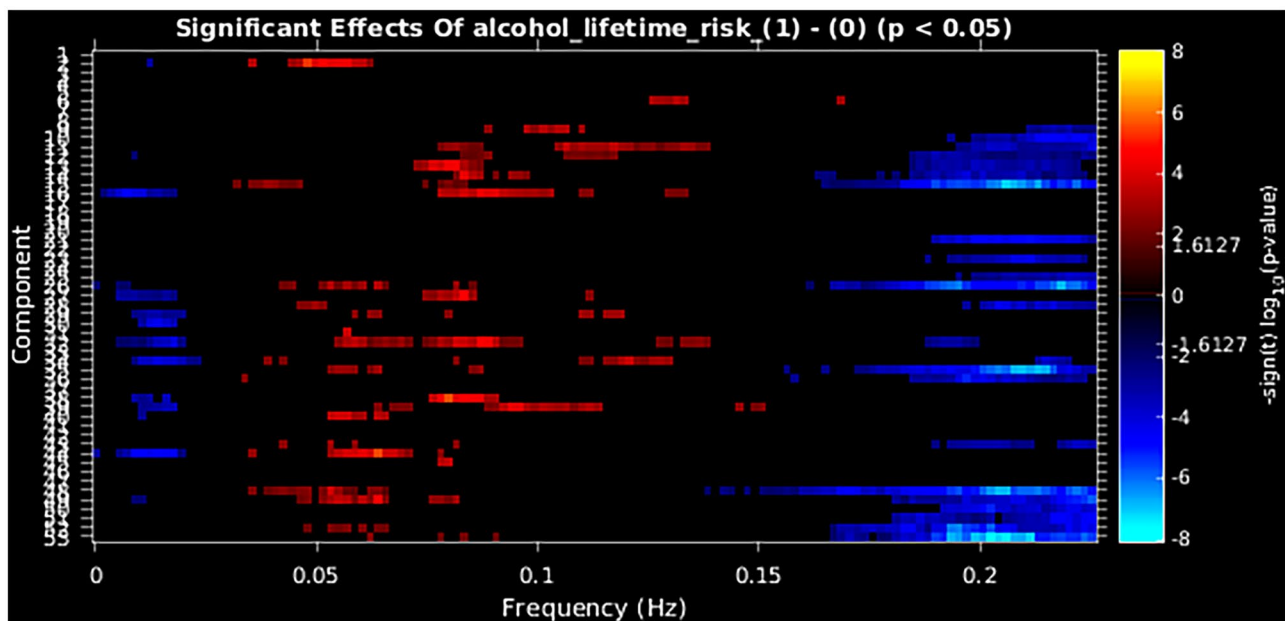


Fig. 8 Power spectra of each component for alcohol risk (condition 1) vs no risk (condition 0) flagged red or blue based on the direction of effect for every value where $p < 0.05$. There is a general increase in

the mid/BOLD frequency range (between 0.05 - 0.1 Hz) and reduced high frequency power above 0.15 Hz

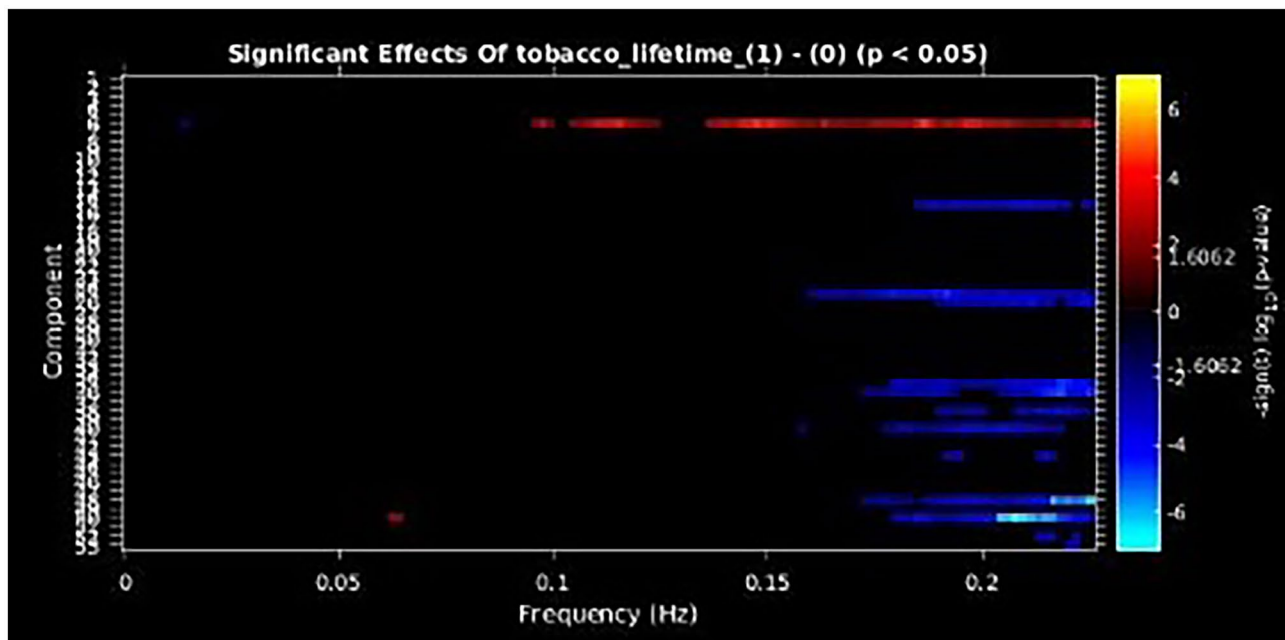


Fig. 9 Power spectra of each component for effect of tobacco risk (condition 1) vs no risk (condition 0) flagged red or blue based on the direction of effect for every value where $p < 0.05$. There was reduced BOLD high frequency power (above 0.15 Hz) across all components

sensorimotor networks. These findings are in support of the reduction of global cortical efficacy in smokers (Fedota & Stein, 2015). There was an exception with the DMN, that had increased within-network connectivity but also with the cognitive control and cerebellar domains. The increased connectivity between the DMN, which is associated with self-referential processes (Andrews-Hanna et al., 2010) and the cerebellum may be related to associative reward learning and potentially cravings (as it relates to these data) as the cerebellar domain has been previously associated with encoding of predicted internal states (Ebner & Pasalar, 2008; Popa & Ebner, 2019). Additionally, hyperconnectivity specifically within the DMN has been noted before and was hypothesized as an indication of the body being in a state of withdrawal (Vergara et al., 2017). A similar effect was identified in cannabis

users (Pujol et al., 2014). Therefore, we speculate that our results of increased connectivity may indicate some states of cravings or withdrawal (e.g., time in the scanner prevented participants from smoking) and not a direct connectivity alteration relating to tobacco use. There is a possible limitation of the validity of self-reports utilized in this study however, these data and protocols have been previously reported in multiple publications (Schumann et al., 2010; Sharma et al., 2020).

In alcohol users, there was significantly more connectivity across all networks when compared to tobacco users. Generally, there was hyperconnectivity between the DMN, CC, and SC domains. One explanation for this increased connectivity may be compensation related to impaired functionality that results from prolonged or excessive alcohol use. In other words, the hyperconnectivity may be a sign of overactivation,

Fig. 10 Functional Network Connectivity matrices generated using GIFT interface, *d*MANCOVA (Alg. 1) and *d*MANCOVA_{pa} (Alg. 2)

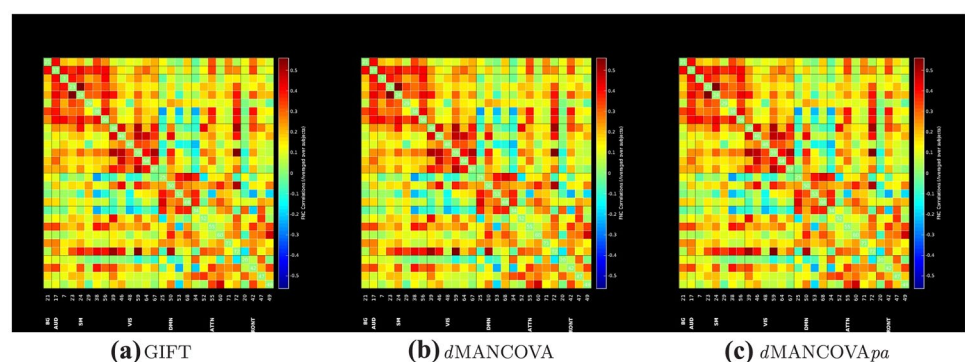
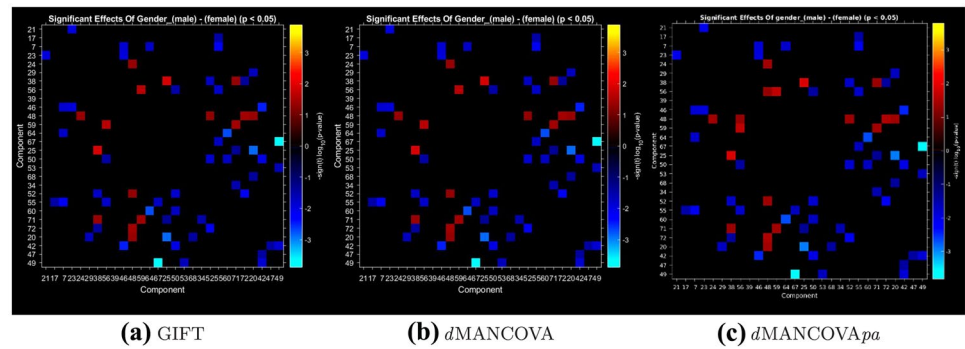


Fig. 11 Significant effects of gender generated using GIFT interface, *d*MANCOVA (Alg. 1) and *d*MANCOVA_{pa} (Alg. 2)



or enhanced neural effort, that is needed to achieve the original (or baseline) cortical performance in areas that have been negatively affected by alcohol (Jansen et al., 2015; Zhu et al., 2017). As some of these hyper-connections (within DMN and DMN and CC) were seen with the tobacco users, another potential explanation is the previously stated hypothesis that the hyperconnectivity is representing a state of withdrawal and not a direct change from alcohol use. However, given the limitations of this cross-sectional study, these results would benefit from future studies examining how much of this connectivity increase is related to the stages (and degree) of withdrawal or if this increase is still identified after substance consumption.

The remainder of the 75 identified pairs in alcohol use indicated hypoconnectivity. The negative pairs were across all seven domains, showing support for global cortical deficits in alcohol users. The strongest of those were within the DMN and between the DMN and AUD domains and between AUD and VIS domains. The visual region, specifically the middle occipital gyrus (Camchong et al., 2012), has previously been identified as a primary area of impact in alcohol users (Vergara et al., 2017). The hypoconnectivity between the DMN and VIS indicate a reduction in domains known for self-reference and information processing related to one's own body, aspects that when not functioning properly could explain excessive consumption, among other decisions.

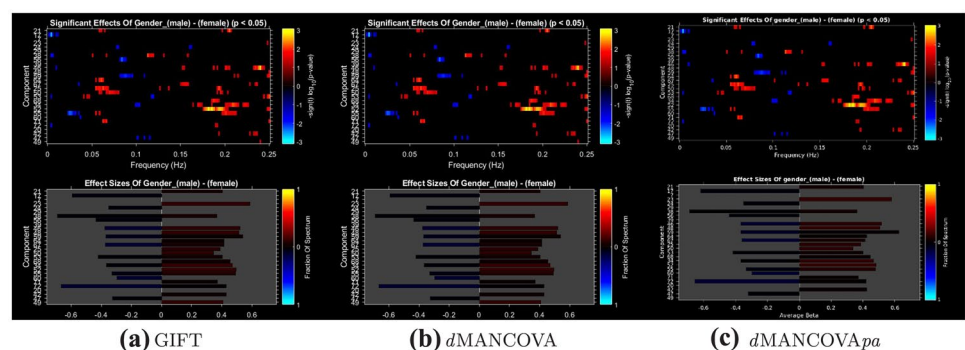
Finally, both tobacco and alcohol users had general hypoconnectivity within the DMN, SM, VIS, AUD, CC, and CR

domains. Alcohol users had more connections than tobacco users, with the majority of those pairs also being negative. These findings suggest that global reduced synchronization is associated with both tobacco and alcohol use. Specific domains, namely the DMN, CC, and VIS domains may explain some of the mechanisms behind substance use and how addictions are formed.

The results of the decentralized MANCOVA with a private aggregator show that it is possible to obtain results similar to results where the analysis is performed locally. In fact, it can also be seen that the results also qualitatively match the ones from a centralized analysis. This is a good indicator of how end-to-end large scale pipelines can be successfully implemented in COINSTAC without concerns about loss of significance.

There are some limitations to this project. First, the two datasets were preprocessed using different softwares and approaches. For this analysis we used the Neuromark spatially constrained ICA pipeline, which included templates intentionally derived from components that replicated across multiple large datasets collected with different parameters. The spatial constraint allows for automated single subject ICA results, capturing individual differences, and also helping to regularize the results, providing additional robustness to noise and dataset variation (Du et al., 2016, 2020). Second, the two datasets did significantly differ on demographics, see Table 1. However, the purposes of this study were to examine the similarities that were present between

Fig. 12 Significant effects and effect sizes of gender generated using GIFT interface, *d*MANCOVA (Alg. 1) and *d*MANCOVA_{pa} (Alg. 2)



the datasets. To evaluate site specific effects, we performed an additional analysis of the effects for each site separately and results were largely consistent with the pooled analysis (significant pairs were correlated above 0.8 for alcohol and tobacco), providing additional support for the reported results.

In summary, we proposed an ICA-based framework to generalize and standardize the calculation of possible functional connectivity features that leverages the benefits of a data-driven approach and also provides comparability across multiple analyses. In the current study, the experimental conditions for data collection and preprocessing is specific to each site thus enabling a straightforward implementation of adaptive-ICA for estimation of subject-specific components. Future work includes the development of iterative estimation schemes for data shared between sites. One of the advantages of decentralized analysis pipelines is that only intermediary statistics are passed between sites, and full patient records are never released across the network. These kinds of decentralized algorithms are “plausibly private” (Sarwate et al., 2014), due to the lack of directly identifiable records in the global data network. We believe work presented here provides a clear direction for future work in improving our understanding of brain disorders by leveraging data from different sites.

Conclusion

In this work, we demonstrated the feasibility of performing a multi-site large-sample ICA-based framework called Neuromark. The goal of Neuromark is to generalize and standardize the calculation of possible functional connectivity features that leverages the benefits of a data-driven approach and also provides comparability across multiple analyses. An exploratory analysis was performed on COINSTAC, a decentralized framework, using functional MRI data collected in the UK and India. Additionally, we also demonstrated the decentralized statistical analysis of subject-specific network features for group level-inferences. We hope this proof-of-concept will be a useful stepping stone towards eventual application of such decentralized approaches for analyzing data from different sites and for further adoption in the clinical setting.

Information Sharing Statement

More specific details about accessing the datasets used in the study can be found at (Zhang et al., 2020) for cVEDA and (Schumann et al., 2010) for IMAGEN. The COINSTAC software can be accessed at <https://github.com/trendscenter/coinstac>.

Acknowledgements This work was funded by the National Institutes of Health (grants: R01DA040487, R01MH121246, R01DA049238, and 1R01DA040487) and the National Science Foundation (grant: 2112455). This work received support from the following sources: Consortium on Vulnerability to Externalizing Disorders and Addictions (cVEDA) is jointly funded by the Indian Council for Medical Research (ICMR/MRC/3/M/2015-NCD-I) and the Newton Grant from the Medical Research Council(MR/N000390/1), United Kingdom, the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant ‘STRATIFY’ (Brain network based stratification of reinforcement-related disorders) (695313), Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the National Institute of Health (NIH) (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers), the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King’s College London, the Bundesministerium für Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, NE 1383/14-1), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 - MRF-058-0004-RG-DESR1 and MR/S020306/1 - MRF-058-0009-RG-DESR-C0759), the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: the ANR (ANR-12-SAMA-0004, AAPG2019 - GeBra), the Eranet Neuron (AF12-NEUR0008-01 - WM2NA; and ANR-18-NEUR00002-01 - ADORé), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l’Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau; the National Institutes of Health, Science Foundation Ireland (16/ERC/D/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence. ImagenPathways “Understanding the Interplay between Cultural, Biological and Subjective Factors in Drug Use Pathways” is a collaborative project supported by the European Research Area Network on Illicit Drugs (ERANID). This paper is based on independent research commissioned and funded in England by the National Institute for Health Research (NIHR) Policy Research Programme (project ref. PR-ST-0416-10001). The views expressed in this article are those of the authors and not necessarily those of the national funding agencies or ERANID. Author ZZ received a fellowship from the Medical Research Foundation (MRF-058-0014-F-ZHAN-C0866).

Author Contributions HG led the manuscript writing with support from KR, BH and SB. BH contributed data and was instrumental in performing the analysis in COINSTAC. ZZ contributed data to the study as well as contributed to describing the data. EV and RK manage the COINSTAC project. SP, AD and JT are part of the COINSTAC team. VB and GS are the principal investigators of the cVEDA-IMAGEN consortium and have been instrumental in facilitating this multi-site study. VC is also a co-investigator and leads the COINSTAC team, formed the vision for this work, and helped interpret the results. All others who were not mentioned here are part of either the cVEDA or IMAGEN consortia.

Data Availability Statement Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Declarations

Conflict of Interest Dr. Banaschewski served in an advisory or consultancy role for Lundbeck, Medice, Neurim Pharmaceuticals, Oberberg GmbH, Shire. He received conference support or speaker's fee by Lilly, Medice, Novartis and Shire. He has been involved in clinical trials conducted by Shire & Viforpharma. He received royalties from Hogrefe, Kohlhammer, CIP Medien, Oxford University Press. The present work is unrelated to the above grants and relationships. Dr. Barker has received honoraria from General Electric Healthcare for teaching on scanner programming courses. Dr. Poustka served in an advisory or consultancy role for Roche and Viforpharm and received speaker's fee by Shire. She received royalties from Hogrefe, Kohlhammer and Schattauer. The present work is unrelated to the above grants and relationships. The other authors report no biomedical financial interests or potential conflicts of interest.

References

- Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Frontiers in Systems Neuroscience*, 5, 2.
- Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., & Buckner, R. L. (2010). Functional-anatomic fractionation of the brain's default network. *Neuron*, 65, 550–562.
- Baker, B. T., Damaraju, E., Silva, R. F., Plis, S. M., & Calhoun, V. D. (2020). Decentralized dynamic functional network connectivity: State analysis in collaborative settings. *Human Brain Mapping*.
- Baker, B. T., Silva, R. F., Calhoun, V. D., Sarwate, A. D., & Plis, S. M. (2015). Large scale collaboration with autonomy: Decentralized data ica. In *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). IEEE.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57, 289–300.
- Calhoun, V. D., Liu, J., & Adalı, T. (2009). A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage*, 45, S163–S172.
- Camchong, J., Stenger, A., & Fein, G. (2012). Resting-State Synchrony During Early Alcohol Abstinence Can Predict Subsequent Relapse. *Cerebral Cortex*, 23, 2086–2099.
- Du, Y., Allen, E., He, H., Sui, J., Wu, L., & Calhoun, V. (2016). Artifact removal in the context of group ica: A comparison of single-subject and group approaches. *Human Brain Mapping*, 37, 1005–1025.
- Du, Y., & Fan, Y. (2013). Group information guided ica for fmri data analysis. *Neuroimage*, 69, 157–197.
- Du, Y., Fu, Z., Sui, J., Gao, S., Xing, Y., Lin, D., Salman, M., Abrol, A., Rahaman, M. A., Chen, J., Hong, L. E., Kochunov, P., Osuch, E. A., & Calhoun, V. D. (2020). NeuroMark: An automated and adaptive ICA based pipeline to identify reproducible fMRI markers of brain disorders. *NeuroImage: Clinical*, 28, 102375.
- Ebner, T. J., & Pasalar, S. (2008). Cerebellum predicts the future motor state. *The Cerebellum*, 7, 583–588.
- Eickhoff, S., Nichols, T. E., Horn, J. D. V., & Turner, J. A. (2016). Sharing the wealth: Neuroimaging data repositories. *NeuroImage*, 124, 1065–1068.
- Fedota, J. R., & Stein, E. A. (2015). Resting-state functional connectivity and nicotine addiction: prospects for biomarker development. *Annals of the New York Academy of Sciences*, 1349, 64.
- Gazula, H., Baker, B., Damaraju, E., Plis, S. M., Panta, S. R., Silva, R. F., & Calhoun, V. D. (2018). Decentralized analysis of brain imaging data: Voxel-based morphometry and dynamic functional network connectivity. *Frontiers in Neuroinformatics*, 12, 55.
- Gazula, H., Holla, B., Zhang, Z., Xu, J., Verner, E., Kelly, R., Jain, S., Bharath, R. D., Barker, G. J., Basu, D., Chakrabarti, A., Kalyanam, K., Kumaran, K., Singh, L., Kuriyan, R., Murthy, P., Benega, V., Plis, S. M., Sarwate, A. D., Turner, J. A., Schumann, G., & Calhoun, V. D. (2021). Decentralized multisite VBM analysis during adolescence shows structural changes linked to age, body mass index, and smoking: a COINSTAC analysis. *Neuroinformatics*.
- Gazula, H., Kelly, R., Romero, J., Verner, E., Baker, B. T., Silva, R. F., Imtiaz, H., Saha, D. K., Raja, R., Turner, J. A., et al. (2020). Coinstac: Collaborative informatics and neuroimaging suite toolkit for anonymous computation. *Journal of Open Source Software*, 5, 2166.
- Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., & FAGERSTROM, K. -O. (1991). The fagerström test for nicotine dependence: a revision of the fagerstrom tolerance questionnaire. *British Journal of Addiction*, 86, 1119–1127.
- Holmes, A. J., Hollinshead, M. O., O'keefe, T. M., Petrov, V. I., Fariello, G. R., Wald, L. L., Fischl, B., Rosen, B. R., Mair, R. W., Roffman, J. L., et al. (2015). Brain genomics superstruct project initial data release with structural, functional, and behavioral measures. *Scientific Data*, 2, 1–16.
- Jansen, J. M., van Holst, R. J., van den Brink, W., Veltman, D. J., Caan, M. W., & Goudriaan, A. E. (2015). Brain function during cognitive flexibility and white matter integrity in alcohol-dependent patients, problematic drinkers and healthy controls. *Addiction biology*, 20, 979–989.
- Lin, Q. -H., Liu, J., Zheng, Y. -R., Liang, H., & Calhoun, V. D. (2010). Semiblind spatial ica of fmri using spatial constraints. *Human Brain Mapping*, 31, 1076–1088.
- Ming, J., Verner, E., Sarwate, A., Kelly, R., Reed, C., Kahleck, T., Silva, R., Panta, S., Turner, J., Plis, S., et al. (2017). Coinstac: Decentralizing the future of brain imaging analysis. *F1000Research*, 6, 1512.
- Plis, S. M., Sarwate, A. D., Wood, D., Dieringer, C., Landis, D., Reed, C., Panta, S. R., Turner, J. A., Shoemaker, J. M., Carter, K. W., et al. (2016). Coinstac: a privacy enabled model and prototype for leveraging and processing decentralized brain imaging data. *Frontiers in Neuroscience*, 10, 365.
- Popa, L. S., & Ebner, T. J. (2019). Cerebellum, predictions and errors. *Frontiers in Cellular Neuroscience*, 12, 524.
- Pujol, J., Blanco-Hinojo, L., Batalla, A., López-Solà, M., Harrison, B. J., Soriano-Mas, C., Crippa, J. A., Fagundo, A. B., Deus, J., De la Torre, R., et al. (2014). Functional connectivity alterations in brain networks relevant to self-awareness in chronic cannabis users. *Journal of Psychiatric Research*, 51, 68–78.
- Saha, D. K., Calhoun, V. D., Du, Y., Fu, Z., Panta, S. R., Kwon, S., Sarwate, A., & Plis, S. M. (2021). Privacy-preserving quality control of neuroimaging datasets in federated environment. *bioRxiv*, (p. 826974).
- Saha, D. K., Calhoun, V. D., Panta, S. R., & Plis, S. M. (2017). See without looking: joint visualization of sensitive multi-site datasets. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI'2017)* (pp. 2672–2678). Melbourne, Australia.
- Salman, M. S., Wager, T. D., Damaraju, E., Abrol, A., Vergara, V. M., Fu, Z., & Calhoun, V. D. (2021). An approach to automatically label and order brain activity/component maps. *Brain Connectivity*.
- Sarwate, A. D., Plis, S. M., Turner, J. A., Arbabshirani, M. R., & Calhoun, V. D. (2014). Sharing privacy-sensitive access to neuroimaging and genetics data: a review and preliminary validation. *Frontiers in Neuroinformatics*, 8, 35.

- Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., Conrod, P., Dalley, J., Flor, H., Gallinat, J., et al. (2010). The imagen study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, 15, 1128.
- Sharma, E., Vaidya, N., Iyengar, U., Zhang, Y., Holla, B., Purushottam, M., Chakrabarti, A., Fernandes, G. S., Heron, J., Hickman, M., Desrivieres, S., Kartik, K., Jacob, P., Rangaswamy, M., Bharath, R. D., Barker, G., Orfanos, D. P., Ahuja, C., Murthy, P., Jain, S., Varghese, M., Jayarajan, D., Kumar, K., Thennarasu, K., Basu, D., Subodh, B. N., Kuriyan, R., Kurpad, S. S., Kalyanram, K., Krishnaveni, G., Krishna, M., Singh, R. L., Singh, L. R., Kalyanram, K., Toledano, M., Schumann, G., & Benegal, V. (2020). Consortium on vulnerability to externalizing disorders and addictions (cVEDA): A developmental cohort study protocol. *BMC Psychiatry*, 20.
- Shringarpure, S. S., & Bustamante, C. D. (2015). Privacy risks from genomic data-sharing beacons. *The American Journal of Human Genetics*, 97, 631–646.
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10, 557–570.
- Vergara, V. M., Liu, J., Claus, E. D., Hutchison, K., & Calhoun, V. (2017). Alterations of resting state functional network connectivity in the brain of nicotine and alcohol users. *NeuroImage*, 151, 45–54.
- White, T., Blok, E., & Calhoun, V. D. (2020). Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. *Human Brain Mapping*.
- Wilcox, C. E., Brett, M. E., & Calhoun, V. D. (2020). Objective markers for psychiatric decision-making: How to move imaging into clinical practice. *NeuroImage: Clinical*, 26, 102084.
- Zhang, Y., Vaidya, N., Iyengar, U., Sharma, E., Holla, B., Ahuja, C. K., Barker, G. J., Basu, D., Bharath, R. D., Chakrabarti, A., Desrivieres, S., Elliott, P., Fernandes, G., Gourisankar, A., Heron, J., Hickman, M., Jacob, P., Jain, S., Jayarajan, D., Kalyanram, K., Kartik, K., Krishna, M., Krishnaveni, G., Kumar, K., Kumaran, K., Kuriyan, R., Murthy, P., Orfanos, D. P., Purushottam, M., Rangaswamy, M., Kupard, S. S., Singh, L., Singh, R., Subodh, B. N., Thennarasu, K., Toledano, M., Varghese, M., Benegal, V., & Schumann, G. (2020). The consortium on vulnerability to externalizing disorders and addictions (c-VEDA): an accelerated longitudinal cohort of children and adolescents in India. *Molecular Psychiatry*, 25, 1618–1630.
- Zhu, X., Cortes, C. R., Mathur, K., Tomasi, D., & Momenan, R. (2017). Model-free functional connectivity and impulsivity correlates of alcohol dependence: a resting-state study. *Addiction Biology*, 22, 206–217.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Harshvardhan Gazula¹  · Kelly Rootes-Murdy² · Bharath Holla^{3,44} · Sunitha Basodi² · Zuo Zhang⁴ · Eric Verner² · Ross Kelly² · Pratima Murthy³ · Amit Chakrabarti⁵ · Debasish Basu⁶ · Subodh Bhagyalakshmi Nanjajya⁶ · Rajkumar Lenin Singh⁷ · Roshan Lourembam Singh⁸ · Kartik Kalyanram⁹ · Kamakshi Kartik⁹ · Kumaran Kalyanaraman¹⁰ · Krishnaveni Ghattu¹⁰ · Rebecca Kuriyan¹¹ · Sunita Simon Kurpad¹² · Gareth J Barker¹³ · Rose Dawn Bharath¹⁴ · Sylvane Desrivieres⁴ · Meera Purushottam¹⁵ · Dimitri Papadopoulos Orfanos¹⁶ · Eesha Sharma¹⁷ · Matthew Hickman¹⁸ · Mireille Toledano¹⁹ · Nilakshi Vaidya^{20,21} · Tobias Banaschewski²² · Arun L. W. Bokde²³ · Herta Flor^{24,25} · Antoine Grigis¹⁶ · Hugh Garavan²⁶ · Penny Gowland²⁷ · Andreas Heinz²⁸ · Rüdiger Brühl²⁹ · Jean-Luc Martinot³⁰ · Marie-Laure Paillère Martinot^{30,31} · Eric Artiges^{30,32} · Frauke Nees^{21,22,24} · Tomás Paus^{33,34} · Luise Poustka³⁵ · Juliane H. Fröhner³⁶ · Lauren Robinson³⁷ · Michael N. Smolka³⁶ · Henrik Walter²⁸ · Jeanne Winterer^{28,38} · Robert Whelan³⁹ · IMAGEN Consortium · Jessica A. Turner^{2,40} · Anand D. Sarwate⁴¹ · Sergey M. Plis^{2,42} · Vivek Benegal³ · Gunter Schumann^{4,21,43} · Vince D. Calhoun^{2,40}

¹ Martinos Center for Biomedical Imaging, MGH and Harvard Medical School, Boston, USA

² Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA

³ Department of Psychiatry, National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India

⁴ Centre for Population Neuroscience and Precision Medicine (PONS), Institute of Psychiatry, Psychology & Neuroscience, SGDP Centre, King's College London, London SE5 8AF, United Kingdom

⁵ ICMR-Center on Non-communicable Diseases, Kolkata, India

⁶ Department of Psychiatry, Postgraduate Institute of Medical Education & Research, Chandigarh, India

⁷ Department of Psychiatry, Regional Institute of Medical Sciences, Imphal, Manipur, India

⁸ Department of Psychology, Regional Institute of Medical Sciences, Imphal, Manipur, India

⁹ Rural Health Centre, Rishi Valley, Andhra Pradesh, India

¹⁰ Epidemiology Research Unit, CSI Holdsworth Memorial Hospital, Mysore, India

¹¹ Division of Nutrition, St John's Research Institute, Bengaluru, India

¹² Department of Psychiatry & Department of Medical Ethics, St. John's Medical College & Hospital, Bangalore, India

- ¹³ Department of Neuroimaging, Institute of Psychology, Psychiatry & Neuroscience, King's College London, London, UK
- ¹⁴ Department of Neuroimaging and Interventional Radiology, NIMHANS, Bengaluru, India
- ¹⁵ Molecular Genetics Laboratory, NIMHANS, Bengaluru, India
- ¹⁶ NeuroSpin, CEA, Université Paris-Saclay, Paris, France
- ¹⁷ Department of Child & Adolescent Psychiatry, NIMHANS, Bengaluru, India
- ¹⁸ Bristol Medical School, University of Bristol, Bristol, UK
- ¹⁹ MRC Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
- ²⁰ Present Address: Centre for Addiction Medicine, NIMHANS, Bengaluru, India
- ²¹ Centre for Population Neuroscience and Precision Medicine (PONS), Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin Berlin, Berlin, Germany
- ²² Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim 68159, Germany
- ²³ Discipline of Psychiatry, School of Medicine and Trinity College Institute of Neuroscience, Trinity College Dublin, Dublin, Ireland
- ²⁴ Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany
- ²⁵ Department of Psychology, School of Social Sciences, University of Mannheim, Mannheim 68131, Germany
- ²⁶ Departments of Psychiatry and Psychology, University of Vermont, Burlington, Vermont 05405, USA
- ²⁷ Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom
- ²⁸ Department of Psychiatry and Psychotherapy CCM, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany
- ²⁹ Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Berlin, Germany
- ³⁰ Institut National de la Santé et de la Recherche Médicale, INSERM U A10 "Trajectoires développementales en psychiatrie"; Université Paris-Saclay, Ecole Normale supérieure Paris-Saclay, CNRS, Centre Borelli, Gif-sur-Yvette, France
- ³¹ AP-HP, Sorbonne Université, Department of Child and Adolescent Psychiatry, Pitié-Salpêtrière Hospital, Paris, France
- ³² Psychiatry Department, EPS Barthélémy Durand, Etampes, France
- ³³ Department of Psychiatry, Faculty of Medicine and Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada
- ³⁴ Departments of Psychiatry and Psychology, University of Toronto, Toronto, Ontario, Canada
- ³⁵ Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, Göttingen 37075, Germany
- ³⁶ Department of Psychiatry and Neuroimaging Center, Technische Universität Dresden, Dresden, Germany
- ³⁷ Department of Psychological Medicine, Section for Eating Disorders, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, UK
- ³⁸ Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany
- ³⁹ School of Psychology and Global Brain Health Institute, Trinity College Dublin, Dublin, Ireland
- ⁴⁰ Department of Psychology, Georgia State University, Atlanta, GA, USA
- ⁴¹ Department of Electrical and Computer Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ, USA
- ⁴² Department of Computer Science, Georgia State University, Atlanta, GA, USA
- ⁴³ Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China
- ⁴⁴ Integrative Medicine, National Institute of Mental Health and Neurosciences (NIMHANS), Bengaluru, India