In Search of Multimodal Neuroimaging Biomarkers of Cognitive Deficits in Schizophrenia

Supplemental Information

Table S1. The substance use history and medication information of the subjects

Measure	HC	SZ	p^a
Number	50	47	
Age	36.7 ± 12.6	35.3 ± 12.6	0.6
Gender	20 F / 30 M	6 F / 41 M	0.01
Education	13.8 ± 1.6	12.7 ± 2.2	0.001
Smokers (yes/no)	11/39	15/32	0.28
Alcohol use history (yes/no)	10/40	14/33	0.24
Cannabis use history (yes/no)	2/48	8/39	0.03
Stimulant use history (yes/no)	0/50	5/42	0.02
Opioid use history (yes/no)	0/50	2/45	0.14
Cocaine use history(yes/no)	0/50	1/46	0.51
History of heart disease, stroke, hypertension, diabetes and dyslipidemia (yes/no)	11/39	7/40	0.24
Olanzapine equivalent	NA	13.5 ± 9.4	
Antipsychotic (yes/no)	NA	42/5	
Antipsychotic (1^{st} /2^{nd} generation)	NA	5/37	
Clozapine (yes/no)	NA	2/45	
Mood stabilizer (yes/no)	NA	1/46	
Antidepressant (yes/no)	NA	8/39	
Benzodiazepine (yes/no)	NA	9/38	
Anticholinergic (yes/no)	NA	3/44	
Beta-blocker (yes/no)	NA	3/44	

[a]The *p* values represent the statistical significance between controls (HC) and schizophrenia patients (SZ). All subjects were not active substance use disorder (6 month minimum before enrollment, except for nicotine). Though there are more patients having substance use history than controls, there is no significant correlation between the substance use history with either the MCCB composite or the loadings of the identified CV8.

MCCA Method

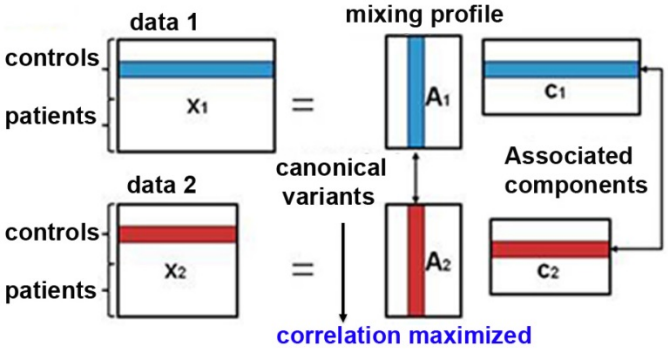

Figure S1. Optimization strategy of two-way CCA

Multi-set canonical correlation analysis (MCCA) (1, 2) is an extension of traditional two-way CCA. As shown in Figure S1, two-way multimodal CCA allows a different mixing matrix for each modality and is used to find the transformed coordinate system that maximizes the inter-subject covariation across two data sets (3). Each dataset is decomposed into a set of components (such as spatial areas for fMRI/sMRI) and their corresponding mixing profiles, which are called canonical variants (CVs). The canonical variants have varying levels of weights for different subjects and are linked if they modulate similarly across subjects. After decomposition, the correlation between CVs is maximized, which correlate with each other only on the same indices.

Likewise, MCCA aims to find linear transforms that simplify the correlation structure among *N* datasets. In MCCA, any pair-wise modality combination among *N* modalities is optimized similarly as in two-way CCA, but by multiple stages, please see more details in (2). In each stage, a linear combination is found such that correlations among the canonical variants are maximized. Finally, a group of corresponding components with heterogeneous correlation values are generated, which only correlate with each other on the same indices (same columns of mixing matrix of each dataset). Therefore, MCCA is primarily designed to find associations in

multiple datasets/modalities, which is suitable to identify linked relationships between measures of interest and co-varied patterns in multimodal brain imaging data.

Choice of Multivariate Fusion Methods

Multimodal joint analysis has proven to be more informative and powerful in understanding brain activity and disorders. The existing multivariate fusion methods have different optimization priorities and limitations. Among them, joint independent component analysis (jICA) (4), MCCA (5), and MCCA+jICA (6) are three data-driven multivariate methods that need no priors and have been successfully used in several mental disorder applications (7, 8). Each of them presents a different view in interpreting and connecting the multiple datasets based on their own hypotheses. In Table S2, we listed these 3 fusion methods with their feasible combinations, optimization assumptions and purpose of the analysis, please see more details in (9), which may serve as a guideline on method selection based on a given research and data. In conclusion, the use of data fusion is a powerful technique, which can help better elucidate the relationship among multiple modalities and facilitate new discoveries in brain disorders. Selecting which fusion model to use should be done carefully based on the research purpose. All these methods are available for use in Fusion ICA Toolbox (http://mialab.mrn.org/software/fit/index.html).

Table S2. Hypothesis of 3 data-driven fusion methods and how to choose among them

Methods	Combinations	Optimization Assumptions	Analyzing Purposes
Joint ICA	fMRI-sMRI fMRI-DTI fMRI-EEG	Multiple datasets share the same mixing matrix, while the independence among joint components is maximized	∨ To examine a common inter-subject co-variation among *N* modalities and to find the linked source maps
MCCA	fMRI-EEG fMRI-sMRI-EEG fMRI-sMRI-DTI	Pair-wise modality correlation between the corresponding mixing profiles are maximized	∨ To detect flexible co-occurring associations among *N* modalities ∨ To identify linked relationships between measures of interest and multimodal data
MCCA+jICA	fMRI-DTI fMRI-sMRI-DTI	Assume the decomposed components from each modality were correlated (highly or weakly) between subject-mixing profiles, while the spatial independence is also maximized.	∨ To achieve both flexible modality associations (high or low correlations) as well as accurate source separation. ∨ To explore multimodal co-occurring and modality-unique alterations between groups

Supplemental References

1. Kettenring JR (1971): Canonical analysis of several sets of variables. *Biometrika*. 58:433-451.

2. Li YO, Adali T, Wang W, Calhoun VD (2009): Joint blind source separation by multi-set canonical correlation analysis. *IEEE Trans Signal Process*. 57:3918-3929.

3. Correa NM, Li YO, Adali T, Calhoun VD (2008): Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE J Sel Top Signal Process*. 2:998-1007.

4. Calhoun VD, Adali T, Liu J (2006): A feature-based approach to combine functional MRI, structural MRI and EEG brain imaging data. *Proceedings of the 28th IEEE EMBS Annual International Conference*. New York City, USA, pp 3672-3675.

5. Correa NM, Eichele T, Adali T, Li YO, Calhoun VD (2010): Multi-set canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI. *NeuroImage*. 50:1438-1445.

6. Sui J, He H, Pearlson GD, Adali T, Kiehl KA, Yu Q, *et al.* (2012): Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia. *NeuroImage*. 2:119-132.

7. Sui J, Yu Q, He H, Pearlson GD, Calhoun VD (2012): A selective review of multimodal fusion methods in schizophrenia. *Front Hum Neurosci*. 6:27.

8. Kohavi R (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI'95 Proceedings of the 14th international joint conference on artificial intelligence*. 2:1137-1143.

9. Sui J, Adali T, Yu Q, Chen J, Calhoun VD (2012): A review of multivariate methods for multimodal fusion of brain imaging data. *J Neurosci Methods*. 204:68-81.