Multivariate Methods for Fusion of Multimodal Imaging and Genetic Data

Vince D. Calhoun, Ph.D.
Executive Science Officer &
Director, Image Analysis & MR Research
The Mind Research Network

Distinguished Professor, Electrical and Computer Engineering
(Primary), Biology, Computer Science, Psychiatry, & Neurosciences
The University of New Mexico
Overview

• Motivation
• Joint ICA
  • ERP/FMRI
  • Multitask-fMRI
  • CC-ICA/Biomarker Identification
• mCCA
  • sMRI/fMRI/ERP
• mCCA+jICA
  • DTI+fMRI
  • N-way/CASH
• Parallel ICA
  • SNPs
• Parallel Group ICA
• Conclusions
Multimodal Data Collection

Brain Function (spatiotemporal, task-related)

Brain Structure (spatial)

Genetic Data (spatial/chromosomal)

Covariates (Age, etc.)

Task 1-N

EEG

fMRI

T1/T2

SNP

DTI

Gene Expression

Other*
A Role For Data Fusion?:

- EEG: Lower Temporal Resolution, Higher Spatial Resolution
- fMRI: Higher Temporal Resolution, Lower Spatial Resolution
Possible approaches for joint analyses

• **Voxel-based**
  - Correlation [Worsely 1998]
  - *Straightforward, but difficult to visualize*

• **Region-based**
  - Interregional correlation [Horwitz, et al, 1984]
  - Multiple regression and extensions [e.g., Kalman filters, Buchel & Friston, 1998]
  - *Useful for model testing, does not take into account all brain regions*

• **Transformation-based**
  - A natural set of tools for this problem include those that transform data matrices into a smaller set of modes or components
  - Singular value decomposition [Friston et al., 1993; Friston et al., 1996]
  - Partial Least Squares [McIntosh, Bookstein, et al, 1996]
  - Canonical Variates Analysis [Strother et al, 1995]
In a non-joint analysis, we maximize the likelihood functions for each modality separately…

Resulting in two unmixing parameters, that then have to somehow be fused together

In contrast: for a joint analysis we maximize the joint likelihood function, resulting in a single fused unmixing parameter

\[ w_{1}^{*} = \arg \max_{w_1} \log p \left( x^{(E)}; w_1 \right) \]

\[ w_{2}^{*} = \arg \max_{w_2} \log p \left( x^{(F)}; w_2 \right) \]

\[ w^{*} = \arg \max_{w} \log p \left( x^{(E)}, x^{(F)}; w \right) \]
Why Multivariate?

Cross-validation performance using the top M SNP’s selected via two different methods as the basis for an N-factor principle component model.
Why Independence?

Uncorrelated: $E\{y_1, y_2\} = E\{y_1\} E\{y_2\}$

Independent: $p(y_1, y_2) = p(y_1) p(y_2)$

$\Rightarrow E\{h(y_1) h(y_2)\} = E\{h(y_1)\} E\{h(y_2)\}$

PCA finds directions of maximal variance (using second order statistics)

ICA finds directions which maximize independence (using higher order statistics)
Why Features?

• **What is a feature?**
  - Lower dimensional data containing information of interest
  - Examples: An image of activation amplitudes, A gray matter segmentation image, fractional anisotropy image

• **Advantages**
  - Less-computationally complex/easier to model
  - Takes advantages of existing analytic approaches
  - Can be used to examine inter-relationships between multiple data types at the subject level
Feature-based ICA

Meta-level ICA

A Family of Multivariate Methods

ICA algorithms are based on cost functions which utilize the higher order statistical information.

- Explicit computation of higher-order statistics
  - Cumulants — e.g., JADE (Cardoso and Souloumiac, 1993)
  - Kurtosis — e.g., FastICA (Hyvärinen 1999)

- Implicit computation of higher-order statistics
  - Mutual information
  - Maximum likelihood/Information maximization — e.g., Infomax (Bell and Sejnowski, 1995)
  - Nonlinear decorrelations
  - Maximization of non-Gaussianity — e.g., FastICA

and computed using matrix diagonalizations, or iteratively using natural/relative gradient, fixed-point, and Newton variate updates.
Joint ICA

Generative Model:

$$\begin{bmatrix} x_{Task1} \\ x_{Task2} \end{bmatrix} = A \times \begin{bmatrix} s_{Task1} \\ s_{Task2} \end{bmatrix}$$

Update Equation:

$$\Delta W = \eta \left\{ I - 2 y^{F1} (u^{F1})^T - 2 y^{F2} (u^{F2})^T \right\} W$$

sMRI/fMRI:


fMRI/fMRI:


EEG/fMRI:

Overview

- **Motivation**
- **Joint ICA**
  - ERP/FMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- **mCCA**
  - sMRI/fMRI/ERP
- **mCCA+jICA**
  - DTI+fMRI
  - N-way/CASH
- **Parallel ICA**
  - SNPs
- **Parallel Group ICA**
- **Conclusions**
Preprocess

Feature Extraction

fMRI

Feature Extraction

EEG

Normalize

jICA

Normalize

Noise Removal

Visualization

Peak of “Target”-locked fMRI signal

Position Cz of “Target”-locked EEG signal

Subject 1

Subject 2

Subject N

Subject 1

Subject 2

Subject N
Joint ERP/fMRI Components

ERP (temporal) Components: \( T = \begin{bmatrix} t_1 & K & t_N \end{bmatrix} \)

FMRI (spatial) Components: \( S = \begin{bmatrix} s_1 & K & s_N \end{bmatrix} \)

ERP Timecourse Snapshot: \( M_E(v) = T \times |S|^T(v) \)

Linked EEG/fMRI Results

Overview

- Motivation
- Joint ICA
  - ERP/FMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
- Parallel Group ICA
- Conclusions
Preprocessing
Timing/motion correction
Spatial normalization
Spatial smoothing

Feature Extraction
GLM Analysis

Normalize
Normalize

Joint Feature Matrix

AOD “targets”
SB-WM “recognition”

jICA

8 components (MDL)

Component Selection
Test loading parameters
(patients differ from controls)
Joint Histograms: fMRI

Overview

• Motivation
• Joint ICA
  • ERP/FMRI
  • Multitask-fMRI
  • CC-ICA/Biomarker Identification
• mCCA
  • sMRI/fMRI/ERP
• mCCA+jICA
  • DTI+fMRI
  • N-way/CASH
• Parallel ICA
  • SNPs
• Parallel Group ICA
• Conclusions
Evaluating Multiple Features

- **Auditory Oddball fMRI Task**
  - Target-related activity
  - Novel-related activity

- **Sternberg fMRI Task**
  - Recognition-related activity

- **MPRAGE**
  - Gray matter segmentation

Identifying the “best” combination

- Kullback-Leibler divergence:

\[ D(s \parallel u) = \int p_s(\xi) \ln \left( \frac{p_s(\xi)}{p_u(\xi)} \right) d\xi \]

\[ s, u \in \mathbb{R}^2 \in \{ F_{SB}, F_{AOD_N}, F_{AOD_T}, S_{GM} \} \]

Optimal Selection of Discriminative Features

\[ C = -E[\ln f(y)] + \lambda \cdot \sum T_i^2 \]

Overview

- Motivation
- Joint ICA
  - ERP/FMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
- Parallel Group ICA
- Conclusions
Multi-set Canonical Correlation Analysis

Three-way MCCA Results....

<table>
<thead>
<tr>
<th>Modalities</th>
<th>three modalities</th>
<th>two modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( t )</td>
<td>( \alpha )</td>
</tr>
<tr>
<td>fMRI</td>
<td>3.4526</td>
<td>0.0015</td>
</tr>
<tr>
<td>sMRI</td>
<td>2.8604</td>
<td>0.001</td>
</tr>
<tr>
<td>EEG</td>
<td>3.6079</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Overview

- Motivation
- Joint ICA
  - ERP/FMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
- Parallel Group ICA
- Conclusions
mCCA and jICA are complementary approaches

- Identify both feature-common and feature-distinct connection
- Make full use of the cross-information of two data sets, in order to separate sources more accurately.
- Use mCCA to link features in the two datasets, then use jICA to decomposes the spatial maps
- Addresses limitation of both mCCA & jICA
mCCA+jICA

Comparison of jICA, mCCA, & mCCA+jICA

Two-way fusion of fMRI/DTI in BP & SZ

Two modalities

FMRI : Auditory oddball (target) contrast maps (AOD)
DTI : Fractional anisotropy (FA)

Three groups including participants

54 schizophrenia (SZ, age 37±12, 22 females)
48 bipolar disorder (BP, age 37±14, 26 females)
62 healthy controls (HC, age 38±17, 30 females)

mean FA maps                                   mean AOD maps

J. Sui, G. D. Pearlson, T. Adali, K. A. Kiehl, A. Caprihan, J. Liu, J. Yamamoto, and V. D. Calhoun,
"Discriminating Schizophrenia and Bipolar Disorder by Fusing FMRI and DTI in A Multimodal CCA+Joint
Results from Schizophrenia, Bipolar, and Healthy Individuals

If the two ICs have the same frame color in two modalities, they are joint ICs.

Variation of functional spatial maps

Back-reconstruction used to estimate group specific maps

Extension of mCCA+jICA to N-way fusion

Preprocessing->Feature

X1 fMRI
X2 DTI
X3 sMRI

Multiset-CCA

A1 x C1
A2 x C2
A3 x C3

Joint ICA

Concatenation

S1 S2 S3

Independent sources

E(A_i, A_j^T) = \text{diag}(\{r_i^{(1)}, r_i^{(2)}, ..., r_i^{(M)}\})

\Lambda_{i,j} = E(A_i A_j^T) = \text{diag}(\{r_i^{(1)}, r_i^{(2)}, ..., r_i^{(M)}\})
i, j \in [1, 2, 3], M \text{ is No. of Component}

Correlation of mixing matrices after multi-set CCA

Group Differences HC vs SZ

<table>
<thead>
<tr>
<th>Group</th>
<th>Number (sex)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HC</td>
<td>116 (44 female)</td>
</tr>
<tr>
<td>SZ</td>
<td>97 (24 female)</td>
</tr>
</tbody>
</table>

Imaging: fMRI, sMRI, DTI
Site effects regressed out

Pair-wise correlation of mixing coefficients

<table>
<thead>
<tr>
<th>IC No.</th>
<th>fMRI-DTI</th>
<th>fMRI-sMRI</th>
<th>DTI-sMRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.195</td>
<td>0.251</td>
<td>0.407</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.085</td>
<td>0.135</td>
</tr>
<tr>
<td>3</td>
<td>0.069</td>
<td>0.097</td>
<td>0.038</td>
</tr>
<tr>
<td>4</td>
<td>0.023</td>
<td>0.094</td>
<td>0.079</td>
</tr>
<tr>
<td>5</td>
<td>0.025</td>
<td>0.002</td>
<td>0.101</td>
</tr>
<tr>
<td>6</td>
<td>0.15</td>
<td>0.11</td>
<td>0.247</td>
</tr>
<tr>
<td>7</td>
<td>0.116</td>
<td>0.065</td>
<td>0.188</td>
</tr>
<tr>
<td>8</td>
<td>0.135</td>
<td>0.003</td>
<td>0.093</td>
</tr>
<tr>
<td>9</td>
<td>0.078</td>
<td>0.101</td>
<td>0.494</td>
</tr>
</tbody>
</table>
Overview

• Motivation
• Joint ICA
  • ERP/FMRI
  • Multitask-fMRI
  • CC-ICA/Biomarker Identification
• mCCA
  • sMRI/fMRI/ERP
• mCCA+jICA
  • DTI+fMRI
  • N-way/CASH
• Parallel ICA
  • SNPs
• Parallel Group ICA
• Conclusions
S. M. Plis, J. Sui, T. Lane, S. Roy, V. Clark, V. Potluru, R. Huster, A. Michael, S. Sponheim, M. P. Weisend, and V. D. Calhoun, "High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia," NeuroImage, in press.
Coregulation analysis via spectra of a hypercube (CASH)

Given a set of \( n \) random variables \( X = \{X_1, X_2, \ldots X_n\} \) the task is to partition their joint probability density into \( k \leq m \) factors \( \{F_1, F_2, \ldots F_k\} \). The assignment of random variables to factors should satisfy the following conditions:

1) The assignment forms a partition of \( X \).
2) Variables in a factor are maximally dependent.
3) Factors are maximally independent.

Let us denote by \( X_i^F \) the \( i^{th} \) random variable assigned to factor \( F \), and by \( \mathcal{D} \) the criterion used to evaluate statistical dependence (\( \mathcal{D} \), for example, can be mutual information or multiinformation). Now we express condition 2 as:

\[
\max J_2 = \Sigma_{i=1}^{k} \mathcal{D}(F_i) = \Sigma_{i=1}^{k} \mathcal{D}(X_1^{F_i}, \ldots X_{|F_i|}^{F_i}), \tag{1}
\]

and condition 3 as:

\[
\min J_1 = \mathcal{D}(F_1, F_2, \ldots F_k). \tag{2}
\]

The cumulative objective function for the proposed factoring of the joint distribution of the random variables is then

\[
\min J = \frac{J_1^\alpha}{J_2^{1-\alpha}}, \tag{3}
\]

S. M. Plis, J. Sui, T. Lane, S. Roy, V. Clark, V. Potluru, R. Huster, A. Michael, S. Sponheim, M. P. Weisend, and V. D. Calhoun, "High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia," NeuroImage, in press.
S. M. Plis, J. Sui, T. Lane, S. Roy, V. Clark, V. Potluru, R. Huster, A. Michael, S. Sponheim, M. P. Weisend, and V. D. Calhoun, "High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia," NeuroImage, in press.
S. M. Plis, J. Sui, T. Lane, S. Roy, V. Clark, V. Potluru, R. Huster, A. Michael, S. Sponheim, M. P. Weisend, and V. D. Calhoun, "High-order interactions observed in multi-task intrinsic networks are dominant indicators of aberrant brain function in schizophrenia," NeuroImage, in press.
Simulation Framework for Data Fusion

- Copulas: a form of bivariate (or multivariate) distributions with uniformly distributed marginals
- Sampling from a bivariate copula leads to two sets of values (each uniformly distributed) which are associated in the way described by the specific copula
- Using the probability integral (or inverse cumulative distribution, icdf) we can assign any desired marginal distribution

Overview

• Motivation

• Joint ICA
  • ERP/FMRI
  • Multitask-fMRI
  • CC-ICA/Biomarker Identification

• mCCA
  • sMRI/fMRI/ERP

• mCCA+jICA
  • DTI+fMRI
  • N-way/CASH

• Parallel ICA
  • SNPs

• Parallel Group ICA

• Conclusions
• **Genetic:** single nucleotide polymorphism (SNP)

• **Genetic:** copy number variation (CNV)

• **Epigenetic:** methylation
Parallel ICA: Two Goals

Data1: (fMRI)  Data2: (SNP)

Identify Hidden Features

\[
\text{MAX : } \{ H(Y_1) + H(Y_2) \}, \quad \langle \text{Infomax} \rangle
\]

Subject to: \( \arg \max g\{W_1, W_2 \mid s_1, s_2 \} \), \( g(\cdot) = \text{Correlation}(A_1, A_2)^2 = \frac{\text{Cov}(a_{1i}, a_{2j})^2}{\text{Var}(a_{1i}) \times \text{Var}(a_{2j})} \)

Simulation

Simulation: Designed to provide a more complete understanding of Parallel ICA while applied to genomic SNP array studies. We specified the parameters for each component and input them into PLINK, an open-source whole genome association analysis toolset [http://pngu.mgh.harvard.edu/purcell/plink/].

Conditions: sample size effect, case to control ratio, SNP array size effect, case-related SNP’s vs. total SNP’s, odds ratio, connection strength between genotype and phenotype effects

Simulation results suggest that parallel ICA, in general, is able to extract more accurately the components and connections than a correlation test, in particular for weak linkages. Results also indicate that the ratio of sample size to SNP size should be at least 0.02. However, when the data have a low odds ratio or cases vs. controls ratio, the correlation test provides results reliably, though with lower accuracy.

Initial Proof of Concept: SNP/fMRI Fusion

Data Description: 20 Sz & 43 Healthy controls
fMRI: one image per subject (Target activation in AOD task)
SNP: one array per subject (384 SNP genotypes - -> 367 SNPs)

controls

Control vs Patient
p<0.001

SNP

<table>
<thead>
<tr>
<th>SNP</th>
<th>Z score</th>
<th>Gene Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rs1466163</td>
<td>-4.08</td>
<td>AADC: aromatic L-amino acid decarboxylase</td>
</tr>
<tr>
<td>Rs2429511</td>
<td>3.97</td>
<td>ADRA2A: alpha-2A adrenergic receptor gene</td>
</tr>
<tr>
<td>Rs3087454</td>
<td>-3.09</td>
<td>CHRNA7: alpha 7 nicotinic cholinergic receptor</td>
</tr>
<tr>
<td>Rs821616</td>
<td>2.96</td>
<td>DISC1: disrupted in schizophrenia 1</td>
</tr>
<tr>
<td>Rs885834</td>
<td>-2.78</td>
<td>CHAT: choline acetyltransferase</td>
</tr>
<tr>
<td>Rs1355920</td>
<td>-2.77</td>
<td>CHRNA7: cholinergic receptor, nicotinic, alpha 7</td>
</tr>
<tr>
<td>R4765623</td>
<td>2.73</td>
<td>SCARB1: scavenger receptor class B, member 1</td>
</tr>
<tr>
<td>Rs4784642</td>
<td>-2.71</td>
<td>GNAO1: guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O</td>
</tr>
<tr>
<td>Rs2071521</td>
<td>2.58</td>
<td>APOC3: apolipoprotein C-III</td>
</tr>
<tr>
<td>Rs7520974</td>
<td>2.55</td>
<td>CHRM3: muscarinic-3 cholinergic receptor</td>
</tr>
</tbody>
</table>

Larger scale study of schizophrenia

- Schizophrenia patients and healthy controls
  - MCIC data: Boston, Iowa, Minnesota and New Mexico
  - Genome-wide 1M SNP data - [biallelic coding (AA, AB, or BB)]
  - fMRI sensorimotor task- Block design motor response to auditory stimulation
- SNP data
  - Subject control: heterozygosity, nearest neighbor, 2nd degree or closer relatives, duplication
  - SNP control: missing genotyping ratio, minor allele frequency, Hardy-Weinberg equilibrium, linkage disequilibrium, etc.
  - Population stratification correction: using PCA
  - Coding (0 for ‘AA’, 1 for ‘AB, and 2 for ‘BB’)
- fMRI data
  - SPM preprocessing (alignment, normalization, filter, GLM) and contrast image
  - Outlier subject excluded
  - Datasets: 208 subjects with SNP (777365 SNPs) and fMRI data (52322 voxels)

Resulting Linked Component

- fMRI component number = 8, SNP component number = 5
- One pair of linked components is identified, with p-value passing Bonferroni correction

<table>
<thead>
<tr>
<th>fMRI component index</th>
<th>SNP component index</th>
<th>$r_{\text{fMRI-SNP}}$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>-0.065</td>
<td>3.49E-01</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0.042</td>
<td>5.48E-01</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.099</td>
<td>1.51E-01</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>-0.138</td>
<td>4.54E-02</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0.141</td>
<td>4.16E-02</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>-0.128</td>
<td>6.44E-02</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.178</td>
<td>9.95E-03</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td><strong>0.282</strong></td>
<td><strong>3.39E-05</strong></td>
</tr>
</tbody>
</table>

- Bootstrap: Multiple runs of parallel-ICA with 5157 randomly selected SNPs. The median correlation was 0.16

### Identified fMRI component

<table>
<thead>
<tr>
<th>Brain region</th>
<th>Brodmann area</th>
<th>Volume</th>
<th>Max z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postcentral Gyrus</td>
<td>1, 2, 3, 5, 7, 40, 43</td>
<td>9.7/4.4</td>
<td>6.23(42,-32,60)/5.70(-39,-32,62)</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>4, 6</td>
<td>8.2/1.7</td>
<td>6.03(42,-12,56)/4.40(-59,-12,42)</td>
</tr>
<tr>
<td>Inferior Parietal Lobule</td>
<td>40</td>
<td>3.0/0.5</td>
<td>6.03(45,-35,57)/4.88(-45,-32,57)</td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>:6, 32</td>
<td>1.0/1.6</td>
<td>4.82(3,-3,53)/5.27(0,-3,53)</td>
</tr>
<tr>
<td>Superior Temporal Gyrus</td>
<td>21, 22, 38</td>
<td>1.4/0.5</td>
<td>3.24(62,-12,1)/2.93(-50,-3,-5)</td>
</tr>
</tbody>
</table>

Summary of SNP Results

- Conduct pathway analysis and functional annotation clustering based on identified 94 genes
  - IPA (Ingenuity Pathway Analysis) identifies “Schizophrenia of humans” as one of the top biofunctions, involving 11 genes
  - IPA also identifies a number of significant canonical pathways, four of which are related to neurotransmitter signaling
  - David’s Bioinformatics Resource reports the most significant cluster to be functionally related to synapse. A cluster annotated as “cell projection” is also identified

<table>
<thead>
<tr>
<th>Disease and disorder</th>
<th>Gene</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizophrenia of humans</td>
<td>BDNF, COMT, DISC1, DRD3, ERBB4, GAD1,</td>
<td>6.49E-09</td>
</tr>
<tr>
<td></td>
<td>GRIN2B, HTR7, NOTCH4, NRG1, NRG3</td>
<td></td>
</tr>
<tr>
<td>Neurotransmitter signaling pathway</td>
<td>GABA receptor signaling</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GABRA4, GABRG3, GAD1</td>
<td>2.13E-03</td>
</tr>
<tr>
<td>Dopamine receptor signaling</td>
<td>COMT, DRD3, PPP2R2C</td>
<td>7.66E-03</td>
</tr>
<tr>
<td>Neuregulin signaling</td>
<td>NRG1, NRG3, ERBB4</td>
<td>1.15E-02</td>
</tr>
<tr>
<td>Glutamate receptor signaling</td>
<td>GRIN2B, GRID2</td>
<td>3.74E-02</td>
</tr>
<tr>
<td>Functional annotation cluster</td>
<td>Synapse</td>
<td></td>
</tr>
<tr>
<td></td>
<td>GABRA4, GABRG3, GAD1, GRIN2B, GRID2,</td>
<td>5.20E-03</td>
</tr>
<tr>
<td></td>
<td>SHC4, OTOF, PSD3, CTBP2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cell projection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRD3, GAD1, GRIN2B, MYCBP2, DNAH11, WNT2,</td>
<td>9.70E-02</td>
</tr>
<tr>
<td></td>
<td>ESR1, CDH13, ALCAM, MYO5A</td>
<td></td>
</tr>
</tbody>
</table>

sMRI/SNP

Structural deficits in brain regions consistently implicated in previous schizophrenia reports, including frontal and temporal lobes and thalamus were related to SNPs from 16 genes, several previously associated with schizophrenia risk and/or involved in normal CNS development, including AKT, PI3k, SLC6A4, DRD2, CHRM2 and ADORA2A.

A. sMRI component –A (group difference)

B. sMRI component –B (linked, but no group difference)

Genetics and P3 ERP generation

- Subjects: 41 healthy subjects (24 female, 17 male)
- EEG collected during AOD task, target/novel ERPs extracted
- Blood sample collected, genotyped 384 SNPs from 222 genes 6 physiological systems.

ERP Topography & SNP Associations

Pathway Analysis

Parallel ICA w/ Reference

\[ X_{1/2} = A_{1/2} \cdot S_{1/2} \quad \rightarrow \quad S_{1/2} = W_{1/2} \cdot X_{1/2} \]

\[ Y_{1/2} = \frac{1}{1 + e^{-U_{1/2}}}, \quad U_{1/2} = W_{1/2} \cdot X + W_{10/20} \]

\[ F1 = \max \{ H(Y_i) \} = \max \left\{ -E \left[ \text{Inf}_{Y_i}(Y_i) \right] \right\} \]

\[ F2 = \max \left\{ H(Y_2) - \text{dist}^2(r, S_{2k}) \right\} = \max \left\{ -E \left[ \text{Inf}_{Y_2}(Y_2) \right] + \lambda \cdot \| r - (W_{2k}X_2)_{k} \|_2^2 \right\} \]

\[ F3 = \max \left\{ \sum_{i,j} \text{corr}(A_{1i}, A_{2j})^2 \right\} = \max \left\{ \sum_{i,j} \frac{\text{cov}(A_{1i}, A_{2j})^2}{\text{var}(A_{1i})\text{var}(A_{2j})} \right\} \]

---

pICA-R: Schizophrenia & ANK pathway

### Classification with SNP & fMRI

**Dataset**

<table>
<thead>
<tr>
<th>Subjects</th>
<th>20 Patients</th>
<th>20 Healthy Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI data</td>
<td>Auditory Oddball Task</td>
<td></td>
</tr>
<tr>
<td>Gene data</td>
<td>367 SNPs</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Number of Training Subjects</th>
<th>Number of Testing Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>28</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>4</td>
</tr>
</tbody>
</table>

**SVM Classification**

Overview

- Motivation
- Joint ICA
  - ERP/fMRI
  - Multitask-fMRI
  - CC-ICA/Biomarker Identification
- mCCA
  - sMRI/fMRI/ERP
- mCCA+jICA
  - DTI+fMRI
  - N-way/CASH
- Parallel ICA
  - SNPs
- Parallel Group ICA
- Conclusions
Parallel Group ICA: Concurrent EEG/fMRI

The SNC results for component pairs that are $r < 0.4$. The edges in red refer to components pairs that survived age adjustment. The correlation coefficient values are adjacent to the edges.

Parallel Group ICA: DTI/fMRI

L. Wu, A. Caprihan, and V. D. Calhoun, "Connectivity Patterns Revealed by Whole Brain Tractography Parcellation with Group ICA," in Proc. HBM, Salt Lake City, 2013.
Parallel Group ICA: DTI/fMRI

Decreased connectivity

Increased connectivity

P* < 0.05 FDR corrected
Age effect removal*
Cluster size > 5 voxels

Mialab Software

- http://mialab.mrn.org/software
- freeware, written in MATLAB (also offering compiled versions): over 11,000 unique downloads

- **Group ICA of fMRI Toolbox (GIFT)**
  - Single subject/Group ICA
  - MANCOVA testing framework
  - Source Based Morphometry
  - Model order estimation
  - ICASSO (clustering/stability)

- **Fusion ICA Toolbox (FIT)**
  - Parallel ICA, jICA
  - mCCA+jICA & much more!

- **Simulation Toolbox (SimTB)**
  - Flexible generation of fMRI-like data

- **COINS**
  - http://coins.mrn.org/dx
You do not need separate MCCA/CCA and IVA figures. IVA includes MCCA and CCA both. IVA takes all-order statistics into account and when it is used with the Gaussian model (hence taking one 2nd order stats) it reduces to MCCA---which includes CCA of course.

Then, you need two models for IVA as well, one for associations among components and one for associations among mixing matrix columns. For the latter, you will need to consider the transposed form of the data matrices---which you have for MCCA.

This will simplify the figures you have and also eliminate the unnecessary correlated-dependent distinction. Just consider IVA as a generalization of MCCA to higher-order (all-order) statistics and only consider dependence, which reduces to linear dependence (correlation) with Gaussian model---and equivalently to MCCA.
Default Mode

Abbreviations:
F: frontal lobe
P: parietal lobe
T: temporal lobe
O: occipital lobe
SLF: Superior longitudinal fasciculus
CGC: Cingulum
CST: Corticospinal tract
UF: Uncinate fasciculus
IFO: Inferior fronto-occipital fasciculus
ILF: Inferior longitudinal fasciculus
ATR: Anterior thalamic radiation.