Published in final edited form as:

Neuroimage. 2014 September ; 98: 386–394. doi:10.1016/j.neuroimage.2014.04.060.

A Three-way Parallel ICA Approach to Analyze Links among Genetics, Brain Structure and Brain Function

Victor M. Vergara¹, Alvaro Ulloa^{1,2}, Vince D. Calhoun^{1,2}, David Boutte¹, Jiayu Chen^{1,2}, and Jingyu Liu^{1,2}

¹The Mind Research Network, 1101 Yale Blvd. NE, Albuquerque, New Mexico, 87106

²Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico, 87106

Abstract

Multi-modal data analysis techniques, such as the Parallel Independent Component Analysis (pICA), are essential in neuroscience, medical imaging and genetic studies. The pICA algorithm allows the simultaneous decomposition of up to two data modalities achieving better performance than separate ICA decompositions and enabling the discovery of links between modalities. However, advances in data acquisition techniques facilitate the collection of more than two data modalities from each subject. Examples of commonly measured modalities include genetic information, structural magnetic resonance imaging (MRI) and functional MRI. In order to take full advantage of the available data, this work extends the pICA approach to incorporate three modalities in one comprehensive analysis. Simulations demonstrate the three-way pICA performance in identifying pairwise links between modalities and estimating independent components which more closely resemble the true sources than components found by pICA or separate ICA analyses. In addition, the three-way pICA algorithm is applied to real experimental data obtained from a study that investigate genetic effects on alcohol dependence. Considered data modalities include functional MRI (contrast images during alcohol exposure paradigm), gray matter concentration images from structural MRI and genetic single nucleotide polymorphism (SNP). The three-way pICA approach identified links between a SNP component (pointing to brain function and mental disorder associated genes, including BDNF, GRIN2B and NRG1), a functional component related to increased activation in the precuneus area, and a gray matter component comprising part of the default mode network and the caudate. Although such findings need further verification, the simulation and in-vivo results validate the three-way pICA algorithm presented here as a useful tool in biomedical data fusion applications.

Keywords

Data fusion; ICA; functional MRI; structural MRI; gray matter; SNP

Corresponding Author: Victor M. Vergara E. The Mind Research Network, 1101 Yale Blvd. NE, Albuquerque, New Mexico 87106
vvergara@mrn.org Phone: 505-4000-5241 or 505-272-5028 Fax: 505-272-8002.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

It is becoming increasingly common in neuroscience, medical imaging and genetic studies to acquire several data modalities from each subject. In these situations, the different data modalities might be related to each other constituting a set of complementary observations of the same phenomenon. The incorporation of all modalities in one comprehensive analysis has the potential to enhance discovery and improve significance when compared to methods that analyze each modality separated. An example is the influence of genetic factors on brain structure (Thompson et al., 2001), important in understanding normal brain development and mental disorders. Associations between these two data modalities are relevant in the study of Alzheimer disease (Stein et al., 2010) and have been studied using the multi-modal method known as Parallel Independent Component Analysis (pICA) (Meda et al., 2012). Other modalities of interest for this type of studies include genetic information, structural magnetic resonance imaging (sMRI), functional MRI (fMRI), and behavioral assessments. The need for multi-modal methods is evident since they are capable of leveraging potentially useful information from complementary modalities.

Multi-modal methods can be divided into hypothesis-driven and data-driven approaches. Hypothesis-driven approaches such as structural equation modeling (Bollen, 1989) allow simultaneous testing of multiple connections between brain MRI data and genetic variations (Baaré et al., 2001), which would be missed using an unimodal method. However, these approaches may suffer when important associations are not included in the initial hypothesis. Unlike hypothesis-driven methods, data-driven approaches rely on the structure of latent variables to reveal associations between modalities. Data-driven approaches based on the Independent Component Analysis (ICA) are beginning to be widely applied (Calhoun and Adali, 2009; Calhoun et al., 2006; Calhoun et al., 2009; Groves et al., 2011) to identify links between brain imaging networks from multiple modalities (e.g. structure and function). Another variant described by (Sui et al., 2010; Sui et al., 2011) combines canonical correlation analysis (CCA) (Hardoon et al., 2004) and ICA to form a multimodal fusion framework. By using CCA to “prefuse” the data, this technique can be extended to many modalities (Takane et al., 2008) without convergence issues. The framework used by (Martinez-Montes et al., 2004) applies partial least squares (PLS) to fuse electroencephalography (EEG) and fMRI data. A similar pairwise fusion algorithm was proposed in (Yang et al., 2010) that combines ICA with a support vector machine (SVM) to fuse fMRI and genetic data. A useful review of these and other existing methods for multimodal data fusion can be found in (Sui et al., 2012). Although all of these methods offer attractive features, they have either not been widely applied to more than two modalities, or they do not explicitly incorporate the information provided by all modalities in one comprehensive data decomposition. This is often a limitation when genetic, structural and functional data are collected together.

The pICA algorithm (Liu et al., 2008b) builds upon the blind matrix factorization techniques used in ICA to simultaneously extract latent statistically independent components and jointly identify mutual relationship between modalities. It allows the integration of data from different contexts and feature spaces. This flexibility is of great advantage, especially when

working with MRI and genetic data modalities which usually have very different number of variables and data ranges. The pICA algorithm analyzes two data modalities concurrently, weights both modalities equally, process each data type according to their own characteristics and allows for the extraction of different number of embedded components. We propose to further generalize the pICA concept by including three modalities. Simulation results demonstrate the three-way pICA ability to uncover connections between all three modalities and resolve components within each modality better than the pICA and ICA decompositions. In addition, we applied our algorithm to an experimental data set consisting of fMRI, collected during an alcohol exposure paradigm (Claus et al., 2011), sMRI and genetic information from 315 subjects participating in an alcohol study. Results obtained from three-way pICA, pICA and ICA are compared to further understand the advantages of the three-way pICA in processing three modalities rather than two or one.

2. Methods

We first discuss the mathematical underpinnings of the three-way pICA algorithm, then detail our simulation design and finally describe how data was collected from a 315 subject's cohort and for each of the three different modalities considered in this work. In the following, we will form one observation matrix for each data modality where subjects are arranged as rows and variables (voxels or loci) as columns. This way, the number of rows of any observation matrix is the same and the other matrix dimension is modality dependent.

2.1. Three-way pICA

The standard ICA problem (Comon, 1994) factorizes a matrix of observations $\mathbf{X}$ into a matrix $\mathbf{A}$ of loading coefficients and a matrix $\mathbf{S}$ representing statistically independent components in the measurement space. Elements in $\mathbf{A}$ accounts for the individual contribution that each component in $\mathbf{S}$ overturn to $\mathbf{X}$. The canonical procedure suggested by (Bell and Sejnowski, 1995) to perform this factorization is to find the matrix $\mathbf{A}^\dagger$ that maximizes some objective function $\mathcal{J}(\mathbf{A}^\dagger \mathbf{X})$ measuring statistical independence, where $\mathbf{A}^\dagger$ is the Moore-Penrose pseudo inverse of $\mathbf{A}$ and $\mathbf{S} = \mathbf{A}^\dagger \mathbf{X}$. The extension of the ICA problem to include two modalities, say $\mathbf{X}^{(1)} = \mathbf{A}^{(1)} \mathbf{S}^{(1)}$ and $\mathbf{X}^{(2)} = \mathbf{A}^{(2)} \mathbf{S}^{(2)}$, consists on constraining the ICA optimization to increase the correlation strength between selected columns of loading matrices $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$ (Liu et al., 2008b).

We propose a three-way pICA consisting of three ICA factorizations $\mathbf{X}^{(1)} = \mathbf{A}^{(1)} \mathbf{S}^{(1)}$, $\mathbf{X}^{(2)} = \mathbf{A}^{(2)} \mathbf{S}^{(2)}$ and $\mathbf{X}^{(3)} = \mathbf{A}^{(3)} \mathbf{S}^{(3)}$, each with individual cost functions, $\mathcal{J}_1(\mathbf{A}^{(1)\dagger} \mathbf{X}^{(1)})$, $\mathcal{J}_2(\mathbf{A}^{(2)\dagger} \mathbf{X}^{(2)})$ and $\mathcal{J}_3(\mathbf{A}^{(3)\dagger} \mathbf{X}^{(3)})$ applied on a different data modality. Additionally, three cost functions $\mathcal{G}_{ij}(\mathbf{A}_i^{(1)}, \mathbf{A}_j^{(2)})$, $\mathcal{G}_{jk}(\mathbf{A}_j^{(2)}, \mathbf{A}_k^{(3)})$, and $\mathcal{G}_{ki}(\mathbf{A}_k^{(3)}, \mathbf{A}_i^{(1)})$, measuring similarities among the i^{th} , j^{th} and k^{th} columns of loading matrices $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$ and $\mathbf{A}^{(3)}$ respectively, are required to assess the connection strength between modalities. The three-way pICA problem can be solved by maximizing the combination of all cost functions, i.e. finding

$$\underset{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}}{\operatorname{argmax}} \left\{ \mathcal{J}_1 + \mathcal{J}_2 + \mathcal{J}_3 + f_{i,j,k} \right\} \quad (1)$$

where $f_{i,j,k}=f(\mathcal{G}_{ij},\mathcal{G}_{jk},\mathcal{G}_{ki})$ is an additional aggregation function measuring the combined connection strength of the columns i, j, k from corresponding matrices $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$, respectively.

Following the development of the pICA, based on Infomax ICA, the entropy $H(\cdot)$ and the squared correlation $\rho^2(\cdot)$ are used to implement cost functions $\mathcal{J}$ and $\mathcal{G}$ respectively. In addition, we propose to set the function $f_{i,j,k}$ as the largest mean value of the squared pairwise column correlations ρ_{ij}^2 , ρ_{jk}^2 and ρ_{ki}^2 over all possible combinations of columns from $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, and $\mathbf{A}^{(3)}$. The three-way pICA implemented for further simulations and data analyses in this manuscript is specifically based on the following maximization

$$\underset{\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \mathbf{A}^{(3)}}{\operatorname{argmax}} \left\{ \sum_{p=1}^3 H_p \left(\mathbf{A}^{(p)\dagger} \mathbf{X}^{(p)} \right) + \left(\alpha_1 \rho_{ij}^2 + \alpha_2 \rho_{jk}^2 + \alpha_3 \rho_{ki}^2 \right) \right\} \quad (2)$$

where the coefficients α_1 , α_2 , and α_3 can be set to $1/3$ (mean value) or adjusted to regulate the relative importance of each correlation term. Gradient update equations for this optimization is similar to what have already been presented in (Liu et al., 2008b), consisting of a two-step update task, one corresponding to entropies and the other to correlations. The entropies for each modality, indexed by p , can be updated by the following equation,

$$\Delta \mathbf{W}^{(p)} = \lambda_p (\mathbf{I} + (\mathbf{I} - 2\mathbf{Y}^{(p)})\mathbf{U}^{(p)}) \quad (3)$$

$$\mathbf{Y}^{(p)} = (\mathbf{I} + e^{-\mathbf{U}^{(p)}})^{-1}; \mathbf{U}^{(p)} = \mathbf{W}^{(p)} \mathbf{X}^{(p)} + \mathbf{W}_0^{(p)}; \mathbf{W}^{(p)} = \mathbf{A}^{(p)\dagger}$$

The second set of updates improves similarities across modalities measured by the correlation terms. These updates are based on the Pearson correlation gradient (Strickert et al., 2008). This linkage between the three modalities is what separates this multi-modal algorithm from other data fusion techniques. First, determine columns i, j and k that correspond to the largest arithmetic mean $f_{i,j,k}$. This can be achieved by evaluating $f_{i,j,k}$ for all possible column combinations and finding the largest value. Once the columns with the strongest correlation are known, update the found columns sequentially and in the following order:

$$\text{define } \mathbf{A}_{u,v}^{(U,V)} = \frac{2\rho(\mathbf{A}_u^{(U)}, \mathbf{A}_v^{(V)})}{\sigma(\mathbf{A}_u^{(U)})\sigma(\mathbf{A}_v^{(V)})} \left\{ (\mathbf{A}_v^{(V)} - \bar{\mathbf{A}}_v^{(V)}) + \frac{\text{cov}(\mathbf{A}_u^{(U)}, \mathbf{A}_v^{(V)}) (\bar{\mathbf{A}}_u^{(U)} - \mathbf{A}_u^{(U)})}{\sigma^2(\mathbf{A}_u^{(U)})} \right\}$$

$$\Delta \mathbf{A}_i^{(1)} = \lambda_1 \eta_1 (\mathbf{A}_{i,j}^{(1,2)} + \mathbf{A}_{i,k}^{(1,3)})$$

$$\Delta \mathbf{A}_j^{(2)} = \lambda_2 \eta_2 \left(\mathbf{A}_{j,i}^{(2,1)} + \mathbf{A}_{j,k}^{(2,3)} \right) \quad (4)$$

$$\Delta \mathbf{A}_k^{(3)} = \lambda_3 \eta_3 \left(\mathbf{A}_{k,i}^{(3,1)} + \mathbf{A}_{k,j}^{(3,2)} \right)$$

Here λ_q are the individual learning rates (where q indexes each of the three modalities) also used to adaptively adjust the impact on the entropy maximization Eq. (3). The coefficients η_q are the step sizes calculated using the Wolfe conditions (Wright and Nocedal, 1999). The function $\cdot$ is the arithmetic mean, $\sigma(\cdot)$ is the standard deviation and $\sigma^2(\cdot)$ is the variance.

The additional updates in this second step may violate the ICA convergence assumptions and can potentially overestimate connections between components. However, adverse convergence behaviors can be minimized and eradicated as $\lambda_q \rightarrow 0$. In fact, the algorithm degenerate to three independent one mode ICAs if $\lambda_q = 0$. Thus, we can use the learning rates λ_q to control for these problems. We monitor the entropy change H_p for each modality online and adaptively anneal the corresponding λ_q if H_p is negative, to ensure that our convergence assumptions are not violated. Stepwise measurements of the entropies H_p reveal information about the impact that additional loading updates from Eq. (4) have on the overall cost function. When annealing is required, the relevant learning rate λ_q is multiplied by an annealing factor $\delta < 1$ (in our implementation defaulted to 0.9) such that $\lambda_q \leftarrow \delta \lambda_q$. These online adjustments guarantee the convergence of the three-way pICA.

The basic algorithm flow is shown in Fig. 1 and in general the algorithm proceeds as follows:

1. Use a standard ICA algorithm to maximize each independence cost function $\mathcal{J}_p(\mathbf{A}^{(p)\dagger} \mathbf{X}^{(p)})$. In this case we use the Infomax algorithm (Bell & Sejnowski, 1995) with entropy cost function $H_p(\mathbf{A}^{(p)\dagger} \mathbf{X}^{(p)})$.
2. At each ICA step find the strongest related triplet $\{\mathbf{A}_i^{(1)}, \mathbf{A}_j^{(2)}, \mathbf{A}_k^{(3)}\}$ among all possible column combinations. In this case we use the mean statistics for the triplet relationship $f_{i,j,k}$ and the squared correlation for pairwise relationship performance.
3. Sequentially update the columns $\{\mathbf{A}_i^{(1)}, \mathbf{A}_j^{(2)}, \mathbf{A}_k^{(3)}\}$ using updates in Eq.(4). Adjust learning rates λ_q and step sizes η_q as necessary to assure convergence.
4. If the ICA of a given modality p converged, no further improvement on $\mathcal{J}_p$ is possible, stop updates for that modality. The algorithm stops when all modalities converge.

The Three Way pICA algorithm was implemented using the software for numerical computation MATLAB (<http://www.mathworks.com/products/mat/>) and the code is made available for public use through <http://mialab.mrn.org/>. Simulations and the in-vivo data example were also processed using this environment.

2.2. Simulation design

In this simulation, a comparison among three-way pICA, pICA and single ICA was conducted considering fMRI, sMRI and single nucleotide polymorphism (SNP) modalities. Data for the three different modalities were simulated from sources that emulate real data behavior and loading matrices with normally distributed elements. The number of sources for all modalities was set to 10. The observation matrices result from the linear combination of loadings and sources plus a noise term for a signal-to-noise ratio (SNR) of 10dB. The simulated connections among the three modalities included two connections varying from weak to strong (fMRI with SNP and sMRI with SNP) and one fixed weak connection (not statistically significant $p>0.05$) between sMRI and fMRI. Correlation significance was computed using the default ρ to t transformation and t -student distribution build in MATLAB functions.

Ten simulated sources were generated for each MRI modality using SimTB (Erhardt et al., 2012), a MATLAB toolbox for the generation of sources and datasets, with similar properties and dimensionalities as those found in fMRI and sMRI sources. Simulated MRI images consisted of 200K voxels for structural and 50K voxels for functional data. Simulated SNP data consisted of 200K markers generated using discrete values (-1, 0, and 1) (Liu et al., 2008a). A set of 125 markers in each simulated SNP source represented the relevant risk alleles of interest. There was no overlap among the 10 simulated sources. The statistics of simulated sources was controlled by adding zero mean Gaussian distributed values to allow for a maximum kurtosis of 3.5.

Loading matrices were built considering a total population of 300 samples, expressed as rows, and ten columns (the number of sources selected for each modality). A multivariate normal distribution with zero mean and identity covariance matrix was used to generate the three loading matrices. The connection strength between modalities was set by randomly replacing one column from each loading matrix with specific designed modality linkage columns. The correlations between linked columns of SNP vs. sMRI and SNP vs. fMRI (the SNP linkage column is the same in both cases) were varied between 0.2 and 0.7. In contrast, the correlation between sMRI and fMRI linked columns was kept constant with a value of 0.1.

SNP observations were subject-wise quantized to integer values -1, 0 and 1 resembling SNP genotype data. Simulated observations were utilized to feed the ICA algorithms aiming at reconstructing the original sources. In the case of pICA and three-way pICA, we also intended to recover the preset correlation strength between modalities. Three-way pICA and regular ICA delivered one set of answers for each modality. However, the outcome from pICA consisted of two sets of results per modality since three pICAs were required for the three pair-wise tests.

2.3. Subjects and data collection

SNP, sMRI and fMRI data were collected from 315 subjects between the ages of 21 and 56 (32 ± 9.8) participating in a study that investigated genetic effects on alcohol dependence (Claus et al., 2011). Male subjects total 220 while the number of females totals

95. Subjects had a regular pattern of alcohol binge drinking (five or more drinks per episode for men, four or more drinks for women). These binge-drinking episodes had to occur at least five times within the past month. Besides binge drinking, subjects do not suffer of other ailments nor have history of severe brain injury or brain-related medical problems and no symptoms of psychosis during a diagnostic interview. Participants provided written informed consent for the collection of samples and subsequent analyses.

All participants underwent functional and structural MRI scans and also provided saliva for DNA extraction and genotyping. Genotyping was performed at the Mind Research Network using the Illumina Infinium Human 1M-Duo assay covering 1,199,187 SNP loci. BeadStudio was used to make the final genotype calls. Then, the PLINK software package (Purcell et al., 2007) (<http://pngu.mgh.harvard.edu/purcell/plink/>) was used to perform a series of quality control procedures (Anderson et al., 2010) including missing data control, heterozygosity control, relatedness control, Hardy-Weinberg Equilibrium, minor allele frequency (> 0.05) and population stratification using the EIGENSTRAT method (Price et al., 2006). This resulted in 717,129 SNP loci. The number of SNPs was further reduced by removing SNPs at intergenic regions and proceeding with linkage disequilibrium pruning using $r^2=0.5$. After this additional data reduction 117,590 SNP loci remained and were used in further analysis.

A T1-weighted structural MRI was acquired using a MP-RAGE sequence on a 3-Tesla scanner. Image registration, bias correction, tissue classification, and spatial normalization to the MNI space were performed using voxel based morphometry (VBM) (Ashburner and Friston, 2005) in SPM5 (<http://www.fil.ion.ucl.ac.uk/spm>). Unmodulated normalized parameters were used for segmentation as previously applied in a large VBM study (Segall et al., 2009) to segment the brain into white matter, grey matter (GM), and cerebral spinal fluid probabilistic maps. At any given voxel the GM map indicates the percentage of GM content at that voxel.

The functional MRI data were collected using a well-established alcohol cue exposure paradigm used to assess brain functional changes related to craving for alcohol (Claus et al., 2011). Participants were exposed to the taste of alcohol (subject's preferred alcoholic beverage) versus the taste of a control (litchi juice). Functional MRI images were collected during pseudo-randomized alcohol and control taste trials. Each trial consisted of a delivery period, followed by a washout period to allow the liquid taste to dissipate before the next trial. A Siemens trio 3T MRI scanner was used with configuration of TR = 2 s, TE = 29ms, flip angle = 75°, voxel size = 3.75 mm × 3.75 mm × 4.55 mm. The contrast images between alcohol trials versus control trials were extracted using FSL (Smith et al., 2004) (<http://www.fmrib.ox.ac.uk/fsl/>), after standard motion correction, normalization to the Montreal Neurological Institute space and smoothing with an 8mm Gaussian model preprocessing steps.

3. Results

3.1. Simulations

Fig. 2 shows performance results for the single ICA, pICA and the three-way pICA algorithms under varying connection strengths. Given that our simulated sample size is equal to 300, correlations smaller than 0.11 will not be significant at the 5% level. A threshold line with value $\rho_{crit} = 0.11$ was included in Fig. 2 in order to better interpret the connection strengths shown in the plots. The simulated correlation between the two connected modalities, SNP vs. sMRI and SNP vs. fMRI, varies between 0.2 and 0.7 as displayed in the x -axis of all plots. In the case of pICA and three-way pICA, the annealing factor δ was set to 0.9. The three-way pICA finds the optimal component triplet, one component from each modality, and corresponding loading matrix columns exhibiting the strongest connection among all possible combinations.

Fig. 2(a–c) shows source estimation accuracy for each of the three modalities. The accuracy of the SNP component is measured in terms of the sensitivity and specificity based on the 125 markers loci defined for the original sources. The specificity of ICA, pICA and three-way pICA are indistinguishable of each other (larger than 0.98) and thus not shown in the plot, however their sensitivities are noticeably different. The accuracy of structural and functional components is calculated by the correlation between the true simulated sources and the corresponding estimated components. The three algorithms show similar performance when all links are weak as in the cases of correlation between linked modalities of 0.2, but pICA and three-way pICA improve their sources estimation along with the strength of the connection between modalities. The two SNP components derived from the two pICAs in Fig. 2(a) exhibit similar accuracies since the simulation was designed with strong connections between the SNP and the two MRI modalities. Since the link between sMRI versus fMRI is weak, there is a difference of pICA results in Fig. 2(b–c). All pICA components resulting from a connection to the genetic modality have similar, but somewhat lower, accuracy as the three-way pICA. All pICA components obtained from sMRI versus fMRI have lower accuracy and similar to what regular ICA delivers.

Fig. 2(d–f) displays the intermodality correlations for each ICA method and among all three modality pairs. The performance for all three decompositions is similar for weak intermodality linkages as in Fig. 2(e) where the fMRI-to-sMRI connection was designed with a low and non-significant correlation. Correlations are better estimated by pICA because the algorithm is able to combine information from two modalities. Again, three-way pICA results are closest to the simulated connection between modalities (ground truth) than the other two algorithms as it combines information from the three available modalities.

3.2. Application to in-vivo data

Since the number of true sources from in-vivo data is in principle unknown, it is necessary to estimate an adequate number of components before proceeding with any ICA factorization. The numbers of components estimated for each modality were 8 from functional MRI, 13 from structural MRI and 10 from SNP data. These component numbers were obtained using corresponding consistency maps (Chen et al., 2012) and confirmed as a

good estimate using ICASSO (Himberg and Hyvarinen, 2003). Data from the three modalities was fed to the three-way pICA, pICA and regular ICA. The annealing factor δ was set to 0.9 for pICA and three-way pICA. All voxels outside the brain or with missing measurements were excluded. The three-way pICA components were compared to those obtained from regular ICA and pICA.

Fig. 3(a–c) shows the fMRI, sMRI and SNP components found by the three-way pICA as the strongest connected triplet. Fig. 3(a–b) are thresholded by $|Z| > 2.0$ to show the main brain regions, and Fig. 3(c) displays the top one hundred SNPs in red circles. Fig. 3(d) displays the pair-wise correlation between loading matrix columns corresponding to each of the selected components. All p-values in Fig. 3 are corrected for Bonferoni multi-comparison at the 5% significance level. The average correlation over all three modalities is 0.27.

In order to understand the SNP component, the set of one hundred genes, corresponding to the SNPs with the largest ICA component weights, were subject to a Pathway Analysis (PA) using the Interactive Pathway Analysis (www.ingenuity.com). The top identified genetic network is associated with function related to neurological diseases and nervous system development. The top canonical pathways point to neuronal regulation, signaling and growth. The main contributing genes to such function include BDNF (Brain-Derived Neurotrophic Factor), BDNF-AS (BDNF antisense RNA), GABRA5 (Gamma-aminobutyric acid A receptor, alpha 5), GRIN2B (Glutamate [NMDA] receptor subunit epsilon-2), NRG1 (Neuregulin 1), HLA-G (major histocompatibility complex, class I, G), DAB1 (Disabled, drosophila, homolog of 1) and ALCAM (Activated Leukocyte Cell Adhesion Molecule). All these genes have demonstrated close relation with neurodevelopmental functions. BDNF (member of the nerve growth factor family) regulates neuronal survival, growth and differentiation (Yang et al., 2013) and particularly plays a vital role in regulating dopaminergic neurons (Cellerino et al., 1998; Zuccato et al., 2001). GABA is the major inhibitory neurotransmitter in the brain, and GABRA5 has been associated with unipolar recurrent major depression (Oru et al., 1997). Glutamate is the major excitatory transmitter in the brain, and the expression of gene GRIN2B has been related to excitatory postsynaptic potentials and thus modulates the efficiency of synaptic plasticity (Chen et al., 1999). Neuregulins and their receptors are essential for neuronal and organ development (Cahill et al., 2013; Law et al., 2004). Together with the ErbB4 receptor, NGR1 constitute a major player in regulating various signaling pathways (Yao et al., 2013). The protein encoded by DAB1 is thought to be a Reelin's signal transducer that regulates neuronal positioning in the developing brain. In mice, DAB1 plays a central role in neuronal layering in the cerebral cortex, hippocampus, and cerebellum (Howell et al., 1997). Two other genes, HLA-G and ALCAM are related to the immune system and modulate general inflammatory response. HLA-G expression in brain specimens from patients with multiple sclerosis implies HLA-G as a contributor to the mechanisms regulating immune reactivity in the central nervous system (Wiendl et al., 2005). ALCAM fulfills important functions in the recruitment of leukocytes into the brain (Cayrol et al., 2008). A list of all one hundred genes used in the Pathway analysis is provided in the Supplementary Tab.1.

The brain regions revealed by the structural component indicates gray matter reduction in Superior Frontal Gyrus (SFG), anterior and posterior cingulate, part of the Default Mode

Network (DMN) (Buckner et al., 2008), the Caudate (an important dopaminergic area regulating reward and emotion), inferior frontal gyrus, dorsolateral PreFrontal Cortex (dlPFC) (Brodmann Areas 8, 9, 10 and 46, important for the executive control function), cerebellar Pyramis (aiding in coordination and movement), gray matter increase in superior and middle temporal gyri. In contrast to the structural component, the functional component contained only positive activated regions. The largest and highest activated region in the functional component is the Precuneus, often identified as part of the DMN (Raichle and Snyder, 2007), followed by adjacent superior parietal and postcentral gyrus areas. Talairach coordinates for the structural and function components discussed above can be seen in Tab. 1 and Tab. 2. The Talairach coordinates for other sMRI and fMRI components are included in Supplementary Tab. 2 and Supplementary Tab. 3.

When applied to the real data, ICA and pICA resolved spatially similar (but not identical) sets of components and with no significant correlation among modalities, except for one pair of fMRI vs. sMRI pICA results (correlation of 0.2 with corrected p-value of 0.02 after multiple comparison correction). There was no major content difference with respect to the fMRI components obtained from three-way pICA, pICA and ICA. This consistency of fMRI components was evaluated visually and through components correlation (>0.95 in all cases). In the case of the SNP components with the strongest link to the fMRI and sMRI components, important genes related to neurological development such as GRIN2B and NRG1 consistently appeared regardless of the ICA method used. However, it is worth to point out that some important neurodevelopmental genes, such as BDNF and BDNF-AS, were detected only by the three-way pICA method. Most of the brain regions within the sMRI component correlating to the SNP component were also consistent across ICA methods except for the Caudate and the dlPFC areas. The sMRI components from ICA and SNP vs. sMRI pICA included SFG, cingulate, temporal and pyramis areas. The sMRI component from fMRI vs. sMRI pICA highlighted the caudate and dlPFC among other brain areas. The three-way pICA sMRI component included all relevant brain areas detected by ICA and pICA delivering the strongest intermodality connections of all three methods. The sMRI results from all ICA methods in consideration are displayed in Supplementary Fig. 1.

4. Discussion

Results in Fig. 2 show a gradual improvement starting from ICA and ending with the more accurate performance obtained from the three-way pICA. At first glance, it is clear that ICA is not designed to accommodate information from complementary modalities, a task better performed by pICA and three-way pICA. The pICA algorithm is designed to pick up the most correlated components from two modalities. In the case of considering three modalities, pICA has to be performed three times (for three possible modality pairs) and resolves two sets of components per data modality with each set emphasizing different modality connections. In this scenario, it is possible for two sets of components of the same modality, but from different pICAs, to differ greatly. Thus, when sorting and matching components from the three pICA results, we may encounter no match. On the other hand, the three-way pICA resolves only one set of components per modality. Since the three-way pICA components characterize the coupled connection strength among three modalities, these components might differ from those components obtained with three pICAs. For

instance, pICA can select the most correlated fMRI and sMRI components, but since only two modalities are processed at one time, these components might not be related to the SNP modality. In contrast, three-way pICA will include fMRI, sMRI and SNP information in the process and identify an optimal three-way linkage.

On each of the individual modalities in the simulated results, seen in Fig. 2(a–c), the three-way pICA components are more correlated with simulated true sources than the other two methods. This confirms the three-way pICA capability in using extra information, provided by cross modal relationships, to guide the optimization search. In addition to the encouraging three-way pICA results for strongly connected modalities, weakly connected modalities do not impair the three-way pICA faculty to resolve independent components. This can be confirmed by considering the weakest linkage case of our simulations, a correlation between modalities of 0.2, because the three ICA methods deliver similar accuracy performance for component estimation. It is important to mention that while the algorithm optimizes the overall three-way connection, three-way pICA did not overestimate the weakly simulated two-way connection between sMRI and fMRI modalities. This fact can be seen in Fig. 2(e) where all three decomposition methods show comparable results.

We need to mention that several conditions were explored when designing the simulations for Fig. 2. In the ideal case of high kurtosis sources all ICA methods were able to recover the sources with high accuracy and robustness against the SNR level (except for SNR of zero or smaller). However, real data does not always fit this ideal condition. At low kurtosis levels, simulated by limiting the kurtosis to be less than 3.5, the three-way pICA showed better performance as illustrated in Fig. 2 and exhibited great robustness for SNR values above 7dB. Below 7dB SNR we could observe some performance degradation for all three ICA decompositions. Empirically, we suggest applying three-way pICA to data with at least 7dB SNR. The influence of the number of samples was tested and a lower limit was experimentally determined at about 200 samples for 200k SNPs. This lower limit indicates that three-way pICA is limited to genetic data sizes of 1kSNP/sample, considering that we simulated the SNP sources of 125 markers and individual marker's mean effect size of 6.0 ± 2.0 % variance explained for sMRI or fMRI components. However, it is safer to stay below the 1kSNP/sample limit in real data applications. In our real data example, we reduced the number of SNPs down to 117k markers (about 0.3kSNP/sample), which is well within the capability of the three-way pICA algorithm. The current version of the three-way pICA only constrains one component triplet (one for each modality). As a proof-of-concept, in this paper we only reported and discussed the most correlated triplet. Future versions of the algorithm might allow for more than one triplet. Another setting closely related to convergence time is the adjustment of the annealing factor δ . If the annealing factor is too high, or close to one, then the convergence of three-way pICA is compromised. If the annealing factor is too low then the algorithm will stop before finding the optimal strongest correlations among modalities. In practice it is convenient to start with a high value, around 0.9, and decrease δ if the algorithm doesn't converge within 512 steps.

When applying the three-way pICA algorithm to 315 alcohol binge drinkers' functional images, gray matter concentration and SNP data, we identified a significant three-way connection that could be interpreted as genetic influences on gray matter and functional

activation in brain regions. Although the same fMRI brain region was identified by all ICA methods in consideration, the complete picture of a three-way relationship with the other two modalities is uncovered only by the three-way pICA, thanks to the constraints posed through the intermodality correlations. Notice that a subset of SNPs only appears in the three-way pICA results, including important neurodevelopmental genes such as BDNF and BDNF-AS. This result could be interpreted as the three-way pICA's ability to highlight genetic information relevant to the two other MRI modalities. The discovery of extra SNP information increases the significance and improves the detection of intermodality linkage. A similar effect can be observed when comparing the different structural MRI components. The structural components found by the three-way pICA and regular ICA (see Supplementary Fig. 1(a–b)), associated with the same fMRI component, are approximately the same. However, the three-way pICA detected brain regions (the Caudate and dlPFC) ignored by regular ICA, but important for the three-way connection among considered modalities.

As observed in the real data application results (see supplementary Fig. 1), the two pICAs of fMRI vs. sMRI and SNP vs. sMRI resolved two distinct sMRI components. The first sMRI component is significantly associated to the fMRI component, but the second one is not significantly correlated to the SNP component. Notice that negative areas in the sMRI component derived from the pICA fMRI vs. sMRI (Supplementary Fig. 1(d)), corresponding to Caudate and dlPFC (BA9), are not present in the sMRI component derived from pICA SNP vs. sMRI (Supplementary Fig. 1(c)). However, the three-way pICA result (in Supplementary Fig. 1(a)) is more comprehensive, including brain areas integrated from regular ICA and the two pICAs. This is a real data example of the possible discrepancies that might be found when using pICA for more than two modalities. We can see that the three-way pICA resolves only one sMRI component that includes all relevant features for the three-way data decomposition.

As seen in Fig. 3(c), the genetic component does not exhibit dominant weights on any particular chromosome regions, but they are distributed across the genome. This is not surprising since the structural component includes regions associated to distinct brain networks, including default mode, executive control, reward/emotion regulation and motion regulation. Most likely, many genes expressed in the brain have some level of contribution to this component. Similarly, many genes involved in neurotransmission, basic cell function, and neurodevelopment likely contribute to brain function even though mainly from the precuneus region. Most importantly, the pairwise correlations between the genetic effects and the functional and structural component are also consistent with previously reported genetic influences on the default mode network (Glahn et al., 2010) where heritability of the default mode network connectivity were reported at 0.424 ± 0.17 .

This overall genetic contribution was confirmed by the PA that associated several genes in the SNP component with neurological diseases including autism, schizophrenia, Parkinson's, motor neuropathy and hypertrophy of dopaminergic neurons. Found pathways highlight a set of genes important to neurological development among which we can mention BDNF-AS, BDNF, GRIN2B and NRG1. The gene Brain-Derived Neurotrophic Factors (BDNF) and the gene BDNF-AS, which inhibits the effects of BDNF (Modarresi et

al., 2012), are known to modulate neurotransmitter activity, cellular growth and have a role in neuronal plasticity. Through intraocular injections of BDNF, BDNF have been linked to hypertrophy (an increase of 41% in the soma area) of dopaminergic neurons, with intense tyrosine hydroxylase (TH) immunoreactivity, in the retina of rodents (Cellerino et al., 1998). Similarly, TH immunoreactive neurons can be found in the frontal lobe and the striatum (Dubach, 1994), regions highlighted by the structural component. In relation with schizophrenia, the expression of gene GRIN2B is altered in the temporal lobe (Bowden et al., 2008). The NRG1 mRNA is expressed in many regions through the brain and clearly detected in the prefrontal cortex and cerebellum (Law et al., 2004).

In relation to the link between structural and functional MRI, there is no obvious spatial overlap between the gray matter concentration and functional activation. However, the three-way pICA results suggest a relationship between brain structural changes in several regions of the DMN and activation in the precuneus region. Some studies provide indication for the existence of this relationship. A circuitry involving the precuneus, medial frontal and posterior cingulate cortex was related to an increase in alcohol odor sensitivity (Bragulat et al., 2008). More directly, deactivation on the precuneus have been connected with a higher availability of dopamine active transporters in the striatum (Tomasi et al., 2009). A decrease in dopamine receptors, thus a decrease on dopamine availability, located in the striatal region was associated with alcohol dependence (Martinez et al., 2005). Although more evidence is needed and given that our sample was drawn from an alcohol study, the negative structural component regions found by the three-way pICA, specially caudate and prefrontal, could reflect decreased dopamine availability in the mesocorticolimbic system of alcoholics which in turn influences an increased activation in the precuneus region. Although such relations need to be further verified, obtained results serve as a proof of concept since the three-way pICA was able to extract very interesting and promising information regarding the relationship among brain function, structure and genetics, which might not be observed with separate ICA of each modality and pICA on pairs of modalities.

5. Conclusion

Increased interest in data fusion of MRI and genetic data as well as increased prevalence of intermediate phenotype imaging paradigms (Glahn et al., 2007) led us to develop this extension of the pICA algorithm. In particular, linking genetic contributions to changes in brain function and structure has been of special interest to many researchers (Hariri and Weinberger, 2003), since the completion of the Human Genome Project. Compared to pICA, the three-way pICA algorithm allows one to jointly analyze three modalities. Simulation indicates that three-way pICA algorithm is able to extract components more correlated with the ground truth, represented by the simulated sources, than pICA or ICA and also reproduce the underlying connection between modalities. In the absence of intermodality linkage, three-way pICA can also find the components without overestimation of intermodality connections. Although the real data application is very limited, as a proof-of concept, our application to sMRI, fMRI and SNP data from the alcohol study samples exemplifies these valuable properties. Results suggest that three-way pICA has the ability of uncovering important interrelated features from the three modalities. The parallel of brain regions and genetic content in our results with other alcohol related findings in the literature

confirms the three-way pICA as a proficient tool for the concurrent analysis of brain structure, function and genetic data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful for the help of the MRN Genetics Lab provided in collecting the genetic data. This work was supported by NIH grants R33DA027626 to JL, 1RC1MH089257 to VC.

References

- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. *Nature Protocols*. 2010; 5:1564–1573.
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage*. 2005; 26:839–851. [PubMed: 15955494]
- Baaré WF, Pol HEH, Boomsma DI, Posthuma D, de Geus EJ, Schnack HG, van Haren NE, van Oel CJ, Kahn RS. Quantitative genetic modeling of variation in human brain morphology. *Cerebral Cortex*. 2001; 11:816–824. [PubMed: 11532887]
- Bell AJ, Sejnowski TJ. An Information Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*. 1995; 7:1129–1159. [PubMed: 7584893]
- Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.
- Bowden NA, Scott RJ, Tooney PA. Altered gene expression in the superior temporal gyrus in schizophrenia. *Bmc Genomics*. 2008; 9:199. [PubMed: 18445270]
- Bragulat V, Dziedzic M, Talavage T, Davidson D, O'Connor SJ, Kareken DA. Alcohol sensitizes cerebral responses to the odors of alcoholic drinks: an fMRI study. *Alcoholism: Clinical and Experimental Research*. 2008; 32:1124–1134.
- Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network. *Annals of the New York Academy of Sciences*. 2008; 1124:1–38. [PubMed: 18400922]
- Cahill ME, Remmers C, Jones KA, Xie Z, Sweet RA, Penzes P. Neuregulin1 signaling promotes dendritic spine growth through kalirin. *Journal of neurochemistry*. 2013; 126:625–635. [PubMed: 23742124]
- Calhoun VD, Adali T. Feature-Based Fusion of Medical Imaging Data. *Ieee Transactions on Information Technology in Biomedicine*. 2009; 13:711–720. [PubMed: 19273016]
- Calhoun VD, Adali T, Giuliani NR, Pekar JJ, Kiehl KA, Pearlson GD. Method for Multimodal analysis of independent source differences in schizophrenia: Combining gray matter structural and auditory oddball functional data. *Human Brain Mapping*. 2006; 27:47–62. [PubMed: 16108017]
- Calhoun VD, Liu J, Adali T. A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage*. 2009; 45:S163–S172. [PubMed: 19059344]
- Cayrol R, Wosik K, Berard JL, Dodelet-Devillers A, Ifergan I, Kebir H, Haqqani AS, Kreymborg K, Krug S, Moumdjian R. Activated leukocyte cell adhesion molecule promotes leukocyte trafficking into the central nervous system. *Nature immunology*. 2008; 9:137–145. [PubMed: 18157132]
- Cellerino A, Pinzón-Duarte G, Carroll P, Kohler K. Brain-derived neurotrophic factor modulates the development of the dopaminergic network in the rodent retina. *The Journal of neuroscience*. 1998; 18:3351–3362. [PubMed: 9547243]
- Chen J, Calhoun VD, Liu J. ICA order selection based on consistency: application to genotype data. *Conf Proc IEEE Eng Med Biol Soc*. 2012; 2012:360–363. [PubMed: 23365904]
- Chen N, Luo T, Raymond LA. Subtype-dependence of NMDA receptor channel open probability. *The Journal of neuroscience*. 1999; 19:6844–6854. [PubMed: 10436042]
- Claus ED, Ewing SWF, Filbey FM, Sabbineni A, Hutchison KE. Identifying neurobiological phenotypes associated with alcohol use disorder severity. *Neuropsychopharmacology*. 2011; 36:2086–2096. [PubMed: 21677649]

- Comon P. Independent component analysis, a new concept. *Signal processing*. 1994; 36:287–314.
- Dubach, M. Telencephalic dopamine cells in monkeys, humans and rats. In: Reiner, A., editor. *Phylogeny and development of catecholamine systems in the CNS of vertebrates*. Cambridge; New York: University of Cambridge Press; 1994. p. 488pxvi
- Erhardt EB, Allen EA, Wei Y, Eichele T, Calhoun VD. SimTB, a simulation toolbox for fMRI data under a model of spatiotemporal separability. *Neuroimage*. 2012; 59:4160–4167. [PubMed: 22178299]
- Glahn D, Winkler A, Kochunov P, Almasy L, Duggirala R, Carless M, Curran J, Olvera R, Laird A, Smith S. Genetic control over the resting brain. *Proceedings of the National Academy of Sciences*. 2010; 107:1223–1228.
- Glahn DC, Thompson PM, Blangero J. Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. *Human Brain Mapping*. 2007; 28:488–501. [PubMed: 17440953]
- Groves AR, Beckmann CF, Smith SM, Woolrich MW. Linked independent component analysis for multimodal data fusion. *Neuroimage*. 2011; 54:2198–2217. [PubMed: 20932919]
- Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*. 2004; 16:2639–2664. [PubMed: 15516276]
- Hariri AR, Weinberger DR. Imaging genomics. *British Medical Bulletin*. 2003; 65:259–270. [PubMed: 12697630]
- Himberg J, Hyvarinen A. Icasso: software for investigating the reliability of ICA estimates by clustering and visualization. *Neural Networks for Signal Processing, 2003*. 2003:259–268. NNSP'03. 2003 IEEE 13th Workshop on. IEEE.
- Howell BW, Hawkes R, Soriano P, Cooper JA. Neuronal position in the developing brain is regulated by mouse disabled-1. *Nature*. 1997; 389:733–737. [PubMed: 9338785]
- Law A, Shannon Weickert C, Hyde T, Kleinman J, Harrison P. Neuregulin-1 (NRG-1) mRNA and protein in the adult human brain. *Neuroscience*. 2004; 127:125–136. [PubMed: 15219675]
- Liu J, Bixler JN, Calhoun VD. A multimodality ICA study-integrating genomic single nucleotide polymorphisms with functional neuroimaging data. *Bioinformatics and Biomedicine Workshops*. 2008a:151–157. 008. BIBMW 2008. IEEE International Conference on. IEEE.
- Liu JY, Demirci O, Calhoun VD. A parallel independent component analysis approach to investigate genomic influence on brain function. *Ieee Signal Processing Letters*. 2008b; 15:413–416. [PubMed: 19834575]
- Martinez-Montes E, Valdes-Sosa PA, Miwakeichi F, Goldman RI, Cohen MS. Concurrent EEG/fMRI analysis by multiway Partial Least Squares. *Neuroimage*. 2004; 22:1023–1034. [PubMed: 15219575]
- Martinez D, Gil R, Slifstein M, Hwang D-R, Huang Y, Perez A, Kegeles L, Talbot P, Evans S, Krystal J. Alcohol dependence is associated with blunted dopamine transmission in the ventral striatum. *Biological psychiatry*. 2005; 58:779–786. [PubMed: 16018986]
- Meda SA, Narayanan B, Liu J, Perrone-Bizzozero NI, Stevens MC, Calhoun VD, Glahn DC, Shen L, Risacher SL, Saykin AJ. A large scale multivariate parallel ICA method reveals novel imaging-genetic relationships for Alzheimer's disease in the ADNI cohort. *Neuroimage*. 2012; 60:1608–1621. [PubMed: 22245343]
- Modarresi F, Faghihi MA, Lopez-Toledano MA, Fatemi RP, Magistri M, Brothers SP, van der, Brug MP, Wahlestedt C. Inhibition of natural antisense transcripts in vivo results in gene-specific transcriptional upregulation. *Nature biotechnology*. 2012; 30:453–459.
- Oru L, Verheyen G, Fura I, Ivezi S, Jakovljevi M, Raeymaekers P, Van Broeckhoven C. Positive association between the GABRA5 gene and unipolar recurrent major depression. *Neuropsychobiology*. 1997; 36:62–64. [PubMed: 9267853]
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*. 2006; 38:904–909. [PubMed: 16862161]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007; 81:559–575.

- Raichle ME, Snyder AZ. A default mode of brain function: a brief history of an evolving idea. *Neuroimage*. 2007; 37:1083–1090. [PubMed: 17719799]
- Segall JM, Turner JA, van Erp TGM, White T, Bockholt HJ, Gollub RL, Ho BC, Magnotta V, Jung RE, McCarley RW, Schulz SC, Lauriello J, Clark VP, Voyvodic JT, Diaz MT, Calhoun VD. Voxel-based Morphometric Multisite Collaborative Study on Schizophrenia. *Schizophrenia Bulletin*. 2009; 35:82–95. [PubMed: 18997157]
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004; 23:S208–S219. [PubMed: 15501092]
- Stein JL, Hua X, Morra JH, Lee S, Hibar DP, Ho AJ, Leow AD, Toga AW, Sul JH, Kang HM. Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. *Neuroimage*. 2010; 51:542–554. [PubMed: 20197096]
- Strickert M, Schleif F-M, Seiffert U, Villmann T. Derivatives of Pearson correlation for gradient-based analysis of biomedical data. *Inteligencia Artificial: revista iberoamericana de inteligencia artificial*. 2008; 12:37–44.
- Sui J, Adali T, Pearlson G, Yang H, Sponheim SR, White T, Calhoun VD. A CCA+ ICA based model for multi-task brain imaging data fusion and its application to schizophrenia. *Neuroimage*. 2010; 51:123–134. [PubMed: 20114081]
- Sui J, Adali T, Yu QB, Chen JY, Calhoun VD. A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*. 2012; 204:68–81. [PubMed: 22108139]
- Sui J, Pearlson G, Caprihan A, Adali T, Kiehl KA, Liu JY, Yamamoto J, Calhoun VD. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA plus joint ICA model. *Neuroimage*. 2008; 57:839–855. [PubMed: 21640835]
- Takane Y, Hwang H, Abdi H. Regularized Multiple-Set Canonical Correlation Analysis. *Psychometrika*. 2008; 73:753–775.
- Thompson PM, Cannon TD, Narr KL, Van Erp T, Poutanen V-P, Huttunen M, Lönqvist J, Standertskjöld-Nordenstam C-G, Kaprio J, Khaledy M. Genetic influences on brain structure. *Nature neuroscience*. 2001; 4:1253–1258.
- Tomasi D, Volkow ND, Wang R, Telang F, Wang G-J, Chang L, Ernst T, Fowler JS. Dopamine transporters in striatum correlate with deactivation in the default mode network during visuospatial attention. *PLoS One*. 2009; 4:e6102. [PubMed: 19564918]
- Wiendl H, Feger U, Mittelbronn M, Jack C, Schreiner B, Stadelmann C, Antel J, Brueck W, Meyermann R, Bar-Or A. Expression of the immune-tolerogenic major histocompatibility molecule HLA-G in multiple sclerosis: implications for CNS immunity. *Brain*. 2005; 128:2689–2704. [PubMed: 16123145]
- Wright, S.; Nocedal, J. Numerical optimization. New York: Springer; 1999. p. 37-41.
- Yang HH, Liu JY, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Frontiers in Human Neuroscience*. 2010;4. [PubMed: 20198130]
- Yang J-L, Lin Y-T, Chuang P-C, Bohr VA, Mattson MP. BDNF and Exercise Enhance Neuronal DNA Repair by Stimulating CREB-Mediated Production of Apurinic/Apyrimidinic Endonuclease 1. *Neuromolecular medicine*. 2013:1–14.
- Yao J-J, Sun J, Zhao Q-R, Wang C-Y, Mei Y-A. Neuregulin-1/ErbB4 signaling regulates Kv4. 2-mediated transient outward K⁺ current through the Akt/mTOR pathway. *American Journal of Physiology-Cell Physiology*. 2013; 305:C197. [PubMed: 23703525]
- Zuccato C, Ciammola A, Rigamonti D, Leavitt BR, Goffredo D, Conti L, MacDonald ME, Friedlander RM, Silani V, Hayden MR. Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science*. 2001; 293:493–498. [PubMed: 11408619]

Highlights

- We propose an independent component analysis to analyze three data modalities.
- The algorithm uncovers hidden relationships among analyzed data modalities.
- Our algorithm outperforms one-modality ICA as shown by simulation results.
- Results from genetic, fMRI and sMRI data reveal connections among these modalities.

Figure 1.
Block diagram of the three-way pICA.

Figure 2.

Component accuracy within modalities and correlation between modalities are shown for the three-way pICA, pICA and ICA. Two out of three intermodality links are significant. The correlation between linked modalities (SNP vs. sMRI and SNP vs. fMRI) varies from 0.2 to 0.7 and corresponds to the x -axis in all plots. The link between sMRI vs. fMRI is statistically non-significant and kept constant at 0.1 as displayed in (e). Given the simulated sample size of 300, correlations below the critical value $\rho_{crit} = 0.11$ are not significant at the 5% level. Plots (a-c) show component accuracy for SNP, sMRI and fMRI respectively. Plots (d) and

(f) show the estimated correlation between modalities. In all plots the three-way pICA algorithm is more accurate at extracting the underlying components and estimating the intermodality correlations than other methods.

Figure 3.

This figure shows the top associated structural MRI, functional MRI and genetic SNP components resulting from the three-way pICA analysis of real data. Structural and functional MRI images were Z-scored and plotted with a threshold greater than $|Z| \geq 2.0$. The SNP component was also Z-scored. The top 100 SNPs, highlighted in red, were used for further analysis. Plot (c) also highlights four neurodevelopmental genes discussed in the text. All correlations are significant and the p-values have been Bonferoni corrected at the 5% significance level.

Table 1

Talairach coordinates for the structural MRI component in Fig. 3(a).

Struct ural Talairach Table			
Area	Brodmann Area	volume(cc) L/R	random effects: Max Value (x, y, z) L/R
<i>Negative</i>			
Superior Frontal Gyrus	6, 8, 9, 10	2.1/2.0	4.2 (0, 20, 49)/4.1 (2, 20, 49)
Inferior Frontal Gyrus	9, 46, 11, 47	1.5/1.9	5.9 (−34, 7, 27)/4.4 (36, 9, 27)
Middle Frontal Gyrus	9, 46, 11, 10, 6	0.8/0.7	5.0 (−34, 9, 31)/4.3 (36, 11, 27)
Pyramis	*	0.3/0.0	5.0 (−6, −85, −23)/−999.0 (0, 0, 0)
Caudate	*	1.6/1.4	4.1 (−12, 18, 8)/4.3 (14, 20, 8)
Anterior Cingulate	24, 32, 33	1.0/3.0	3.0 (−12, 21, −3)/3.2 (2, 33, 0)
Posterior Cingulate	29, 30	1.4/0.6	3.9 (−14, −52, 15)/2.7 (2, −46, 12)
<i>Positive</i>			
Superior Temporal Gyrus	13, 22, 39	1.8/3.4	5.5 (−40, −55, 19)/7.0 (44, −48, 21)
Middle Temporal Gyrus	19, 21, 22, 37, 39	3.1/0.5	5.3 (−40, −57, 21)/4.7 (44, −55, 25)

Table 2

Talairach coordinates for the functional MRI component in Fig. 3(b).

Functional Talairach Table			
Area	Brodmann Area	volume(cc) L/R	random effects: Max Value (x, y, z) L/R
All Positive			
Precuneus	7, 19, 39	8.9/10.7	5.7 (−3, −58, 61)/5.9 (3, −55, 61)
Postcentral Gyrus	1, 2, 3, 5, 7, 40	3.8/4.6	5.4 (−3, −46, 66)/5.6 (6, −49, 66)
Superior Parietal Lobule	5, 7	3.8/4.5	4.9 (−3, −64, 56)/5.4 (12, −67, 56)