



Estimating multivariate similarity between neuroimaging datasets with sparse canonical correlation analysis: an application to perfusion imaging

Maria J. Rosa^{1*}, Mitul A. Mehta¹, Emilio M. Pich², Celine Risterucci², Fernando Zelaya¹, Antje A. T. S. Reinders³, Steve C. R. Williams¹, Paola Dazzan^{3,4}, Orla M. Doyle^{1†} and Andre F. Marquand^{1,5†}

OPEN ACCESS

Edited by:

Pedro Antonio Valdes-Sosa,
Centro de Neurociencias de Cuba,
Cuba

Reviewed by:

Nelson Jesús Trujillo-Barreto,
University of Manchester, UK
Felix Carbonell,
Biospective Inc., Canada

*Correspondence:

Maria J. Rosa,
Centre for Neuroimaging Sciences,
Institute of Psychiatry (PO89),
De Crespigny Park, London
SE5 8AF, UK
maria.rosa@kcl.ac.uk

† Joint senior authors.

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 21 July 2015

Accepted: 23 September 2015

Published: 13 October 2015

Citation:

Rosa MJ, Mehta MA, Pich EM, Risterucci C, Zelaya F, Reinders AATS, Williams SCR, Dazzan P, Doyle OM and Marquand AF (2015) Estimating multivariate similarity between neuroimaging datasets with sparse canonical correlation analysis: an application to perfusion imaging. *Front. Neurosci.* 9:366. doi: 10.3389/fnins.2015.00366

¹ Centre for Neuroimaging Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, ² F. Hoffmann-La Roche Ltd., Basel, Switzerland, ³ Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK, ⁴ National Institute for Health Research Mental Health Biomedical Research Centre, South London and Maudsley National Health Service Foundation Trust, King's College London, London, UK, ⁵ Department of Cognitive Neuroscience, Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands

An increasing number of neuroimaging studies are based on either combining more than one data modality (inter-modal) or combining more than one measurement from the same modality (intra-modal). To date, most intra-modal studies using multivariate statistics have focused on differences between datasets, for instance relying on classifiers to differentiate between effects in the data. However, to fully characterize these effects, multivariate methods able to measure similarities between datasets are needed. One classical technique for estimating the relationship between two datasets is canonical correlation analysis (CCA). However, in the context of high-dimensional data the application of CCA is extremely challenging. A recent extension of CCA, sparse CCA (SCCA), overcomes this limitation, by regularizing the model parameters while yielding a sparse solution. In this work, we modify SCCA with the aim of facilitating its application to high-dimensional neuroimaging data and finding meaningful multivariate image-to-image correspondences in intra-modal studies. In particular, we show how the optimal subset of variables can be estimated independently and we look at the information encoded in more than one set of SCCA transformations. We illustrate our framework using Arterial Spin Labeling data to investigate multivariate similarities between the effects of two antipsychotic drugs on cerebral blood flow.

Keywords: multivariate analysis, sparse canonical correlation analysis, pharmacological MRI, Arterial Spin Labeling, resting cerebral blood flow, intra-modal data, repeated measures, antipsychotics

Introduction

Acquiring multiple neuroimaging datasets from the same subject is becoming common practice (Sui et al., 2012). An increasing number of studies are now based on either combining more than one imaging modality [*inter-modal* studies; e.g., Ystad et al. (2011)] or combining more

than one measurement from the same modality (*intra-modal* studies; e.g., Elsenbruch et al., 2012). Both provide important benefits: inter-modal studies allow the complementary information from different data modalities to be used whereas intra-modal studies can increase power (e.g., through the use of repeated-measures designs) with respect to single modal studies. To date, intra-modal studies using multivariate methods have focused almost exclusively on differences between datasets, for instance relying on classification algorithms to differentiate between effects in the data.

In the context of pharmacological studies using functional Magnetic Resonance Imaging (fMRI), which commonly employ repeated-measures (i.e., *intra-modal*) designs, multivariate analyses have exclusively focused on *discriminating* the distributed effects of different drug interventions. Several studies have employed Blood Oxygen-Level Dependent (BOLD)-fMRI for this purpose based on brain activity patterns derived from participants performing a working memory task (Marquand et al., 2011), on functional connectivity based on resting fMRI (Sripada et al., 2013) or in response to drug infusion (Doyle et al., 2013a). Other studies have used Arterial Spin Labeling (ASL) to discriminate the effects of different drugs on regional cerebral blood flow (rCBF) (Chen et al., 2011; Marquand et al., 2012; Doyle et al., 2013b; Paloyelis et al., 2014). This approach offers the advantage over BOLD that the derived measures are quantitative and can be readily compared between scanning sessions.

Although these studies have significantly contributed to our knowledge of the differential effects of pharmacological interventions across brain regions, it is important to identify not only *differences* but also *similarities* between drug effects (Nathan et al., 2014). This is particularly important in drug discovery, where the effects of novel pharmacological compounds are routinely compared to those of existing drugs (Ashburn and Thor, 2004; Wild et al., 2012; Bruns et al., 2015). For these reasons, and in the larger context of *intra-modal* studies, it would be highly advantageous to develop methods to assess the relationship between sets of measurements (e.g., repeated measures), taking advantage of the multivariate and high-dimensional nature of the data. One classical multivariate technique for estimating the linear relationship between two sets of variables from the same samples (e.g., subjects) is canonical correlation analysis (CCA) (Hotelling, 1936). CCA aims to find a set of linear transformations of the variables in each set of measurements that are maximally correlated with each other. In other words, CCA finds basis vectors (canonical vectors) for each set, such that the correlation between the projections of the data onto these basis vectors is maximized. To find more than one set of basis vectors, CCA can be re-applied to the residual data obtained after subtracting the effect of the first set of canonical vectors from the original data. This way we can look for multivariate relationships in the same data that have not been explained away by the first set of canonical vectors. If the data are normally distributed, CCA is equivalent to maximizing the mutual information between the two datasets (Bach and Jordan, 2003). From a slightly different perspective, CCA can be considered as a method to predict one of the datasets from

the other while accounting for the correlations between features (Breiman and Friedman, 1997; Klami et al., 2013).

In the context of high-dimensional neuroimaging studies, the number of variables (voxels) greatly exceeds the number of subjects, rendering the use of CCA for finding image-to-image correspondences challenging. One way to deal with this issue is to use a kernel-based version of CCA, kernel CCA (KCCA, Hardoon et al., 2004). An advantage of KCCA is that it can capture not only linear but also non-linear dependencies between datasets (Cristianini and Shawe-Taylor, 2000; Hardoon et al., 2004). KCCA has been successfully applied to neuroimaging data (Hardoon et al., 2007, 2009), but similarly to classical CCA, this technique yields non-sparse canonical vectors [i.e., the canonical vectors have non-zero coefficients for every variable (e.g., voxel)]. This is not ideal since the main goal in neuroimaging analyses is often to identify brain regions or networks that are most relevant to the question being studied.

With this goal in mind we seek a sparse solution of CCA, Sparse CCA (SCCA). This can be achieved by regularizing CCA with a penalty that imposes sparsity on the entries of the canonical vectors, as proposed by Witten et al. (2009). SCCA finds a sparse linear combination of the variables in one set of measurements that maximally correlates with a sparse linear combination of the variables in the other set of measurements. The corresponding patterns of non-zero coefficients can be interpreted in terms of networks of brain regions that show similar effects across datasets. Due to the penalization term, SCCA can be applied to high-dimensional/small sampled datasets. A sparse version of KCCA has also been developed but sparsity is imposed only on one of the datasets (Hardoon and Shawe-Taylor, 2011).

In neuroimaging there are several lines of related work. SCCA has been used to find multivariate relationships between high-dimensional brain data and low-dimensional genetic, clinical, and/or cognitive measurements (Boutte and Liu, 2010; Le Floch et al., 2012; Lin et al., 2013a, 2014; Avants et al., 2014; McMillan et al., 2014). To our knowledge, the only work to date where SCCA has been used to look at image-to-image multivariate similarities between whole brain datasets is Avants et al. (2010). The authors used SCCA to relate information encoded in different brain imaging modalities aiming to map the relationship between whole-brain changes in cortical thickness (measured with structural MRI) and white matter integrity (measured with diffusion tensor imaging) in dementia patients relative to healthy controls. The authors chose to use a fixed degree of sparsity that yielded a solution with 50% of the variables in each set of measurements equal to zero. Multi-modal data fusion is another important application for CCA and other related methods (see Biessmann et al., 2011 and Sui et al., 2012 for reviews). These methods include variants of partial least squares (PLS), such as multi-way PLS (Martínez-Montes et al., 2004) and variants of independent component analysis (ICA) such as linked ICA (Groves et al., 2011).

This work makes three main contributions. First, the applications of CCA to repeated measures designs and to ASL are novel. Second, we extend the SCCA framework developed by Witten and Tibshirani (2009) to estimate the optimal

subset of variables for each imaging set independently, using permutation. This is important because there is no guarantee that each set of imaging measurements (each “view”) has the same properties. For example, one view may have many more activated voxels than the other or may have different smoothness. This is particularly important in drug studies, because different drugs can have very different influences on BOLD activity or rCBF. Optimizing a separate parameter for each view provides the model with additional flexibility to accommodate such differences. A simple simulation is provided to illustrate this effect in the Supplementary Material. Third, we explicitly optimize model parameters controlling sparsity, which is in contrast to most other applications that use a fixed sparsity level (e.g., Avants et al., 2010; McMillan et al., 2014). Finally, we map the information encoded in more than one set of SCCA components, which is in contrast to other applications that only examine the first component. We illustrate the feasibility and utility of our framework using Arterial Spin Labeling data from a neuropharmacological study, with the aim of investigating multivariate similarities between the effects of two antipsychotic drugs (haloperidol and aripiprazole) on resting rCBF in the healthy human brain.

Materials and Methods

Canonical Correlation Analysis (CCA)

CCA was first introduced by Hotelling (1936) and is a classical multivariate statistical technique for finding linear relationships between two sets of variables. We begin with two datasets represented by the matrices X_1 and X_2 , having dimensions $n \times p_1$ and $n \times p_2$ respectively, where n is the number of samples (e.g., subjects) and p_1 and p_2 are the number of variables (e.g., voxels) in set X_1 and X_2 , respectively. CCA seeks linear transformations of X_1 and X_2 that are maximally correlated with each other:

$$\begin{aligned} \max_{u,v,q} q &= u^T X_1^T X_2 v \\ \text{subject to } u^T X_1^T X_1 u &= 1 \text{ and } v^T X_2^T X_2 v = 1. \end{aligned}$$

CCA assumes that the columns of X_1 and X_2 are standardized to have mean of zero and standard deviation of one. The vectors u and v , with dimensions $p_1 \times 1$ and $p_2 \times 1$, respectively, are known as the *canonical vectors* (or weights); the vectors $X_1 u$ and $X_2 v$, with dimensions $n \times 1$, are known as the *canonical variables*; and q is called the *canonical correlation*. A schematic representation of CCA is presented in the Supplementary Material. The solution to the above equation can be obtained analytically (Hotelling, 1936). However, when the number of variables exceeds the number of samples, which is often the case in the context of neuroimaging data, CCA cannot be applied directly, because the analytical solutions require inverting covariance matrices that are rank deficient (singular) in this setting. To circumvent this issue, various authors (Waaijenborg et al., 2008; Parkhomenko et al., 2009; Witten et al., 2009) have proposed a regularized version of CCA, called sparse canonical correlation analysis (SCCA). In this paper, we extend the approach developed by Witten et al. (2009).

Sparse Canonical Correlation Analysis (SCCA)

Similarly to CCA, SCCA looks for a linear combination of the variables in X_1 that is maximally correlated with a linear combination of the variables in X_2 . However, in SCCA two penalty function terms, P_1 and P_2 , are introduced to regularize the solution of u and v :

$$\begin{aligned} \max_{u,v,q} q &= u^T X_1^T X_2 v \\ \text{subject to } \|u\|_2^2 &\leq 1, \|v\|_2^2 \leq 1 \text{ and } P_1(u) \leq c_1, P_2(v) \leq c_2. \end{aligned}$$

P_1 and P_2 are in general chosen to yield sparse u and v vectors in that some elements (here, voxel weights) are zero. The constants c_1 and c_2 , are regularization parameters, and can be optimized from the data. For sparse norms, they control the amount of sparsity (number of zero elements) in the solution of u and v . The terms $\|u\|_2^2 = 1$ and $\|v\|_2^2 = 1$ result from replacing the covariance matrix of each dataset by its diagonal, known as “diagonal penalized CCA” (Witten et al., 2009). This simplification has been shown to produce good results in high dimensional classification problems, even when correlation between variables is evident (Dudoit et al., 2002; Tibshirani et al., 2003). In addition, assuming that X_1 and X_2 have been standardized, as described before, the solution of SCCA is unique, i.e., the vectors u and v are unique even when the number of variables in each dataset, p_1 and p_2 , greatly exceed the number of samples, n . In this paper, we use the Least Absolute Shrinkage and Selection Operator (LASSO) penalty function (Tibshirani, 1996) for both u and v vectors: $P_1(u) = \|u\|_1$, and $P_2(v) = \|v\|_1$ where $\|u\|_1$ and $\|v\|_1$ are the L1-norms of vector u and v , respectively (i.e., the sum of the absolute value of the elements of each vector). The estimation procedure for SCCA proposed by Witten et al. (2009) is based on a penalized matrix decomposition and is summarized in the Supplementary Material. We note that although other penalty functions are possible (such as the total variation penalty), the LASSO penalty performs well in practice for neuroimaging data. In practice, the solutions obtained by SCCA are less sparse than for other problems such as logistic regression because of the additional L2 penalty implicitly imposed by the penalized matrix decomposition (see Witten et al., 2009). Additionally, we restrict the elements of u and v to be nonnegative (i.e., $u_i, v_i = 0$), so that the solution of SCCA can be interpreted as a sparse weighted average of the variables in X_1 that maximally correlates with a sparse weighted average of the variables in X_2 (Witten and Tibshirani, 2009). It is important to note here that SCCA does not provide a variable-wise (e.g., voxel-wise) measure of similarity between the two datasets. It provides a single measure of similarity (the canonical correlation) and a vector of weights, representing the basis vectors u and v , for each dataset.

Selection of Regularization Parameters

We use a permutation-based approach for choosing the regularization parameters from the data. We adapt the permutation-based framework proposed by Witten and Tibshirani (2009) in order to allow the regularization parameter for each set of variables to be optimized independently of the other set. In other words, we allow a different value for

c_1 and c_2 . In the original implementation, this is not allowed and the same parameter setting must be used for both. As noted, this is important because X_1 and X_2 may not have the same properties (e.g., number of activated voxels). This does, however, entail estimating two model hyper parameters in place of one. A didactic simulation illustrating this is presented in the Supplementary Material. Since we use SCCA with the L1-norm penalty function, as mentioned above, these parameters control how sparse the solution is, i.e., how many elements of u and v are zero. The permutation-based algorithm proceeds as follows:

1. Given a value of c_1 and c_2 , we independently permute the samples of the datasets, X_1 and X_2 , to create two new datasets, \hat{X}_1 and \hat{X}_2 . We then apply SCCA with parameters c_1 and c_2 as described before, to obtain vectors \hat{u} and \hat{v} .
2. We calculate the canonical correlation for the permuted data: $\hat{q} = \text{Corr}(\hat{X}_1 \hat{u}, \hat{X}_2 \hat{v})$.
3. We repeat steps 1 and 2 k times, each time by permuting the data again and calculating a new \hat{q} .
4. We repeat steps 1 to 3 but use a different choice of parameters c_1 and c_2 (a grid-search approach is used to exhaustively cover all combinations of c_1 and c_2).
5. For each pair of parameters c_1 and c_2 we then use the Fisher transformation to convert the correlation values, \hat{q} , into random variables that are approximately normally distributed and compute a z -statistic: $z_{c_1 c_2} = (\text{Fisher}(\hat{q}) - \text{mean}(\text{Fisher}(\hat{q}))) / \text{std}(\text{Fisher}(\hat{q}))$, where q is the canonical correlation obtained using the original data and parameters c_1 and c_2 . We then choose c_1 and c_2 that yield the highest z -statistic.
6. Finally, we re-estimate SCCA using the original data and the chosen parameters c_1 and c_2 from the last step to obtain the final u and v vectors.

It is common practice in machine learning and statistics to estimate a model using the entire dataset once model selection has been performed using the permutation-based approach described above or other method, such as cross-validation (Burbidge et al., 2001; Lin et al., 2013b). The permutation-based approach is preferable to cross-validation when the number of samples is small, since it avoids the need to split the data into even smaller sets for training and testing (Witten and Tibshirani, 2009). For this reason and the fact that our approach is purely unsupervised, we do not separate the data into training and test sets in this study.

Determination of Significance

To determine the significance of the estimated canonical correlation values (i.e., to measure how unlikely it is to obtain a particular value of q or higher if there was no relationship between the two datasets) we use the same permutation-based approach described above. This involves counting how many times the canonical correlations obtained with the permuted data, \hat{q} , are equal or higher to the canonical correlation obtained with the original data, q , and dividing by the number of permutations. This calculation yields an estimate of the p -value for the overall significance of the correlation.

Multiple Canonical Vectors

The algorithm for SCCA described above can be extended to output more than one set of canonical vectors. The procedure to obtain the i^{th} set of canonical vectors simply involves applying the algorithm to the residual data obtained from removing the effect of the i^{th} -1 set of canonical vectors from the data. This technique is known in linear algebra as matrix deflation. For example, the second set of canonical vectors is obtained by applying SCCA to the dataset:

$$X_1^T X_2' \leftarrow X_1^T X_2 - (u^T X_1^T X_2 v) uv^T,$$

where u and v are the first set of canonical vectors. It is important to note here that the different sets of vectors are not orthogonal given the presence of the penalty terms in the main SCCA equation, although alternative approaches to deflation can be employed to enforce orthogonality (Monteiro et al., 2014).

Pharmacological MRI Data

We illustrate SCCA with Arterial Spin Labeling (ASL) data published in Handley et al. (2013). This study focused on univariate differences between the effects of two antipsychotic drugs (haloperidol and aripiprazole) on resting cerebral blood flow (rCBF) in the human brain. The details of the study can be found in Handley et al. (2013), but for completeness, the main experimental details are reported briefly here.

Participants

Twenty healthy right-handed English speaking Caucasian males, aged 18–33 (mean 23 years SD 4.5), participated in the experiment. Mean IQ and body mass index were within the normal range (Handley et al., 2013). Participants were nonsmoking, university students with no recent or current drug or medication use and had no exposure to psychotropic medication, nor a history of personal or familial psychiatric diagnosis. Written informed consent was obtained from all participants and the study was approved by the Human Research Ethics Committee of the Institute of Psychiatry, London, and conducted in compliance with the Declaration of Helsinki.

Design

A crossover, randomized within-subject, double blinded, placebo-controlled design was used. Participants received a single oral dose of haloperidol (3 mg), aripiprazole (10 mg), or placebo, presented in identical capsules, in a randomized order on three visits. Antipsychotic dose and time of administration (3 h and 30 min) before scanning were intended to achieve a striatal D2 receptor occupancy level comparable across the two compounds (Handley et al., 2013). A minimum of 14 days separated each visit to allow for drug washout. No alcohol or medications were used for 24 h, nor caffeine for 6 h, before scanning.

Data Acquisition

Participants were scanned with their eyes open and images were acquired in a General Electric Signa HDX 1.5 T scanner at the Centre for Neuroimaging Sciences, Institute of Psychiatry.

Regional CBF measurements were made using a pulsed-continuous Arterial Spin Labeling technique (pCASL, Dai et al., 2008). The pulse sequence parameters can be found in Handley et al. (2013).

After labeling, images were acquired with a three-dimensional Fast Spin Echo (FSE) spiral multishot readout. To minimize blurring, the spiral acquisition for each slice was very short (4 ms), and the required resolution was achieved with eight interleaves (TE 32 ms/TR = 5500 ms; ETL = 64). Images were acquired at a $48 \times 64 \times 48$ matrix on an $18 \times 24 \times 18$ cm field of view. Images were reconstructed to a 256×256 matrix, resulting in a nominal spatial resolution of 1×1 mm in plane. Sixty slices of 3 mm thickness were obtained. Three pairs of tagged-untagged images were collected. Following the three ASL control-label pair averages, images were acquired with the same imaging sequence but with inversion recovery preparation instead of ASL. These images were used to quantify blood flow from the ASL, as described in Handley et al. (2013). Two additional structural, high spatial resolution images were acquired for co-registration and normalization, including a T2 weighted fast spin echo and fluid-attenuated inversion-recovery fast spin echo scan to exclude the presence of any brain pathology.

Data Analysis

The pre-processing of the ASL data was performed according to Handley et al. (2013). The rCBF images were processed using Statistical Parametric Mapping 8 (SPM8, www.fil.ion.ucl.ac.uk/spm). For each participant: (i) the extra-cerebral signal from the T2 scan was removed using the “Brain Extraction Tool” (BET) of the Functional Software Library (FSL, Smith, 2002). The skull stripped T2 volume and its corresponding binary mask were then coregistered to the rCBF map; (ii) the coregistered, binary, brain-only mask, was multiplied by the rCBF map to remove extra-cerebral signal from this scan. The skull stripped T2 and rCBF maps were then coregistered back to the space of the original T2 scan; (iii) the original T2 scan was normalized to the MNI based T2 template provided in SPM, and the transformation matrix was applied to

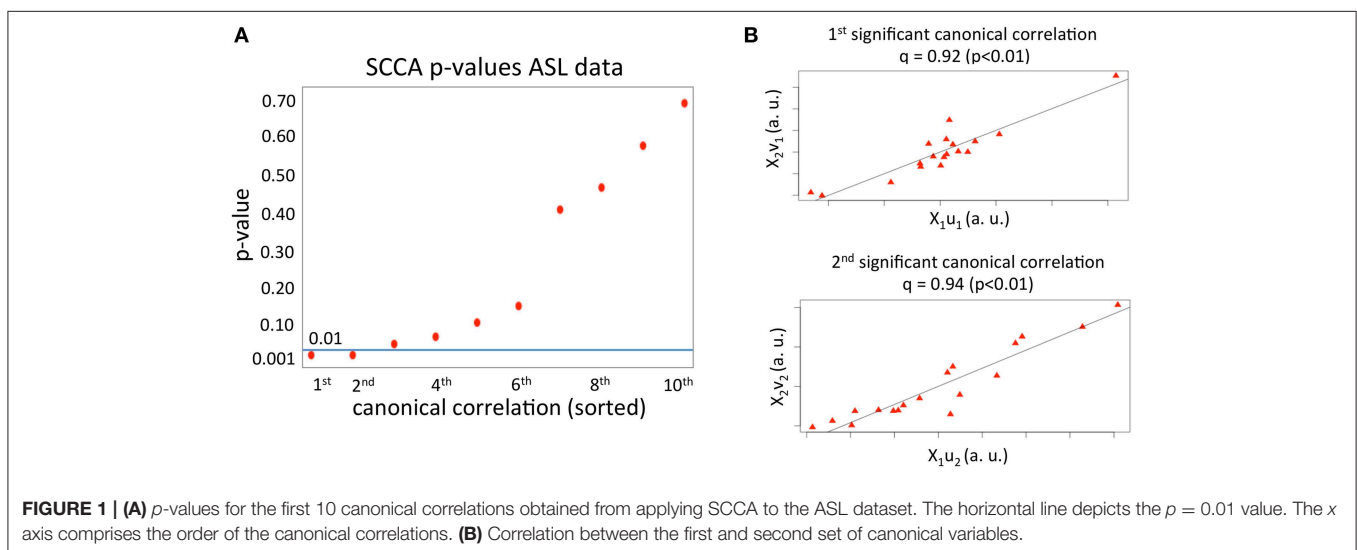
the rCBF map and the T2 scan. After normalization, the data were downsampled to $3 \times 3 \times 3$ mm voxels and then smoothed using a 10 mm Gaussian smoothing kernel.

In addition, we mean scaled each ASL scan (i.e., subtracted the mean of each image) to remove the session effect and subtracted the placebo scan from the haloperidol and aripiprazole images for each subject individually. This way we look for similarities in the mean effects of haloperidol and aripiprazole relative to placebo. The resulting images were then masked with a gray matter template (obtained by thresholding the gray matter template provided by SPM) and used as input to SCCA. Finally, SCCA analyses were based on 18 subjects per group (drug) as the data of two subjects had to be discarded because these subjects developed nausea and vomiting with aripiprazole.

Results

We first applied SCCA to the haloperidol ($n = 18$ subjects and $p_1 = 37702$ variables) and aripiprazole data ($n = 18$ subjects and $p_2 = 37702$ variables) to find the first 10 canonical correlations and sets of canonical vectors, as described above. As can be seen in **Figure 1A**, only two correlations had a $p < 0.01$ (p -value estimated using 1000 permutations). The first significant canonical correlation was 0.92 (p -value = 0.003) and the second significant correlation was 0.94 (p -value = 0.003). The value of the regularization parameters, $c_{1,2}$, that account for how sparse the solution vectors are was varied from 0.3 to 0.9 in steps of 0.1. This range was chosen based on pilot runs, and covered a wide range of sparsity values while enforcing sufficient regularization to avoid the trivial solution (Hardoon et al., 2004). The optimal parameters ($c_1 = 0.4$ and $c_2 = 0.7$) were chosen based on the first set of canonical vectors and using 1000 permutations, as described in the Materials and Methods section.

The first two sets of significant canonical variables are plotted in **Figure 1B**. As can be seen, the correlation between canonical variables (linear transformation of the original samples) is very high, meaning that in the space spanned by the first and second



significant set of canonical vectors, the mean effects of the two drugs on rCBF are highly correlated.

The first set of canonical vectors with significant correlation is shown in **Figure 2**. As can be seen, the first set of canonical vectors is more sparse for haloperidol relative to aripiprazole but with many overlapping regions between the drugs. In both cases, most clusters of non-zero variables were distributed across areas of the frontal cortex (especially medially), limbic system and the striatum. The second set of canonical vectors with significant correlation were characterized by clusters of non-zero variables distributed across areas of the temporal cortex, insula and the middle part of the cingulum (**Figure 3**). We also present an alternative visualization of the weights in the Supplementary Material. As can be seen in **Figure 2** and in the Supplementary Material, the rCBF signature obtained from the first set of canonical vectors shows a similar pattern of non-zero coefficients (network) for haloperidol and aripiprazole on rCBF across the brain. Both drugs seem to affect a network of regions that mainly comprises frontal regions (such as the orbito-frontal cortex and olfactory cortex), the striatum structures (caudate and putamen), the amygdala, anterior cingulate cortex, straight gyrus, the insula, and the temporal poles. One of the main differences between the networks highlighted for the two drugs seems to be the extent of non-zero weight found in frontal regions, which is much larger for aripiprazole (**Figure 2**).

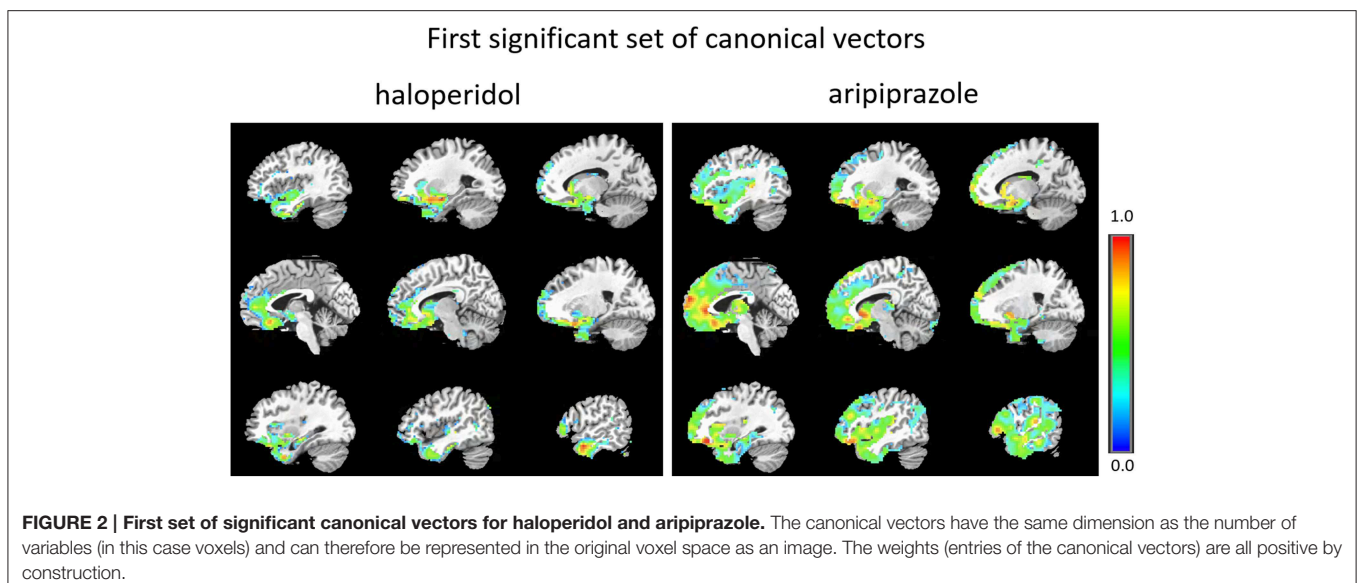
As can be seen, the second set of vectors (**Figure 3** and Supplementary Material) highlights similar networks for the two drugs, comprising mostly temporal regions (temporal poles, superior, and middle temporal cortex), and Heschl's gyrus, the insula, the amygdala, parahippocampal regions, the frontal operculum and the supra-marginal cortex.

Discussion

Here we illustrated the utility of SCCA for investigating multivariate linear relationships between high-dimensional

intra-modal neuroimaging datasets acquired for the same group of subjects. We applied the technique to pharmacological ASL data acquired using a repeated measures design to study the effects of two antipsychotic drugs [haloperidol—first-generation (FGA), and aripiprazole—third-generation, (Mailman and Murthy, 2010); on rCBF of healthy volunteers]. We found similar distributed rCBF networks that maximize the correlation between the two drugs. In particular, the first set of canonical vectors highlighted a network comprising mainly the striatum (caudate and putamen), anterior cingulate cortex, orbital frontal cortex, amygdala, straight gyrus (rectus), olfactory cortex, and temporal poles for both antipsychotics. The second set of canonical vectors highlighted a network comprising mostly temporal regions, the insula, frontal operculum, the supra-marginal cortex and Heschl gyrus. Differences between the networks obtained from the first set of canonical vectors highlight a larger number of frontal regions involved in the effect of aripiprazole compared to haloperidol. The second set of vectors suggests a stronger response in the cingulum again in aripiprazole compared to haloperidol. These differences could be related to the fact that haloperidol is principally a dopamine D2, D3, and D4 receptor antagonist (Okuyama et al., 1997), while aripiprazole on the other hand is a partial agonist at D2 receptors, a serotonin 5-HT1a partial agonist, and an antagonist at serotonin 5-HT2a receptors (Pae et al., 2008).

It is not straightforward to compare these results with previous univariate analyses of the same data, because the SCCA coefficients have a different interpretation to mass-univariate general linear model (GLM) coefficients: SCCA coefficients are multivariate and describe the weight applied to each brain region to maximize the overall correlation between the inputs. In contrast, GLM coefficients describe focal differences between brain regions in a univariate sense. Nevertheless, many features of the networks detected do seem to overlap with previous findings (Handley et al., 2013). Handley et al. (2013) have shown that both haloperidol and aripiprazole have focal effects on



genetic polymorphisms (Grellman et al., 2015). For future work, it may be interesting to extend Bayesian methods to provide sparsity over the features, perhaps using a similar approach to the Bayesian LASSO (Park, 2005) but at the present time we favor the sparse CCA approach presented here.

Sparse CCA has many other potential uses in neuroimaging in addition to the application reported here. The canonical variables can be compared between groups (e.g., between patients and control subjects) when there are more than one group of subjects for whom more than one dataset was acquired (Sui et al., 2012). Sparse CCA can also be used as a feature selection method in a discriminative framework. In other words, SCCA can be used to discriminate two or more classes based on variables that maximize the multivariate correlation between two sets of measurements obtained for each class. This may be useful to find an optimal subspace for classification based on multiple measurements where class differences are more salient.

One limitation of the methodological approach presented here is the fact that even though we constrain the entries of the canonical vectors to be non-negative in order to facilitate their interpretation (in terms of a weighted average of the variables in each dataset), these coefficients are part of a multivariate model and, consequently, regional inferences cannot be performed directly on SCCA maps (i.e., canonical vectors should not be thresholded for local inference). This issue is not specific to SCCA and affects classical CCA, KCCA, and most multivariate-based predictive models used in neuroimaging, such as commonly used linear classifiers. The only way to mitigate this problem to some extent, without relying on *post-hoc* permutation tests and corresponding analytical approximations (Gaonkar and Davatzikos, 2013), is to enforce sparsity within the model, as was done in this work.

To our knowledge this is the first study to present multivariate similarity measures of the effects of two antipsychotic drugs on human rCBF. In general, these measures can be potentially

useful to characterize the effects of any two different tasks or conditions (such as mental and neurological disorders) on human brain activity. In drug studies, similarities are particularly important in drug discovery, where the effects of novel pharmacological compounds are routinely compared to those of existing drugs.

To conclude, we have shown that SCCA is a powerful technique to investigate multivariate similarities between different sets of measurements. In particular, we have demonstrated its feasibility when applied to the challenging question of finding meaningful image-to-image correspondences between neuroimaging data acquired from the same group of subjects. Using ASL data from two different classes of drugs, the feasibility was clear, even in the case of small sample sizes, typical of pharmacological imaging studies.

Acknowledgments

This work was supported by a research fellowship funded by F Hoffman La Roche awarded to SCRW. OD gratefully acknowledges support from the EPSRC grant EP/L000296/1. AM gratefully acknowledges support from the King's College London Centre of Excellence in Medical Engineering, funded by the Wellcome Trust and EPSRC (grant number WT088641/Z/09/Z) and the Language in Interaction project funded by the Dutch Organization for Scientific Research (NWO) under the Gravitation programme (grant number 024.001.006). We also thank Dr David Alsop for facilitating the 3D pseudo continuous ASL pulse sequence used in this work.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnins.2015.00366>

References

- Ashburn, T. T., and Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3, 673–683. doi: 10.1038/nrd1468
- Avants, B. B., Cook, P. A., Ungar, L., Gee, J. C., and Grossman, M. (2010). Dementia induces correlated reductions in white matter integrity and cortical thickness: a multivariate neuroimaging study with sparse canonical correlation analysis. *Neuroimage* 50, 1004–1016. doi: 10.1016/j.neuroimage.2010.01.041
- Avants, B. B., Libon, D. J., Rascovsky, K., Boller, A., McMillan, C. T., Massimo, L., et al. (2014). Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. *Neuroimage* 84, 698–711. doi: 10.1016/j.neuroimage.2013.09.048
- Bach, F. R., and Jordan, M. I. (2003). Kernel independent component analysis. *J. Mach. Learn. Res.* 3, 1–48. doi: 10.1109/icassp.2003.1202783
- Barouche, W. H., Tewelde, B., Schlyer, D., Redvanly, C., Shea, C., and Axeloff, D. (1994). Effects of haloperidol challenge on regional cerebral glucose utilization in normal human subjects. *Am. J. Psychiatry* 151, 681–686. doi: 10.1176/ajp.151.5.681
- Bartlett, E. J., Brodie, J. D., Simkowitz, P., Dewey, S. L., Rusinek, H., Volkow, N. D., et al. (1996). Time-dependent effects of a haloperidol challenge on energy metabolism in the normal human brain. *Psychiatry Res.* 60, 91–99. doi: 10.1016/0165-1781(96)02711-4
- Biessmann, F., Pils, S., Meinecke, F. C., Eichele, T., and Miller, K. R. (2011). Analysis of multimodal neuroimaging data. *IEEE Rev. Biomed. Eng.* 4, 26–58. doi: 10.1109/RBME.2011.2170675
- Boutte, D., and Liu, J. (2010). “Sparse canonical correlation analysis applied to fMRI and sparse data fusion,” in *2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (Hong Kong), 422–426.
- Breiman, L., and Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. *J. R. Stat. Soc. B Stat. Methodol.* 59, 3–54. doi: 10.1111/1467-9868.00054
- Bruns, A., Mueggler, T., Kuennecke, B., Risterucci, C., Prinssen, E. P., Wettstein, J. G., et al. (2015). “Domain gauges”: a reference system for multivariate profiling of brain fMRI activation patterns induced by psychoactive drugs in rats. *Neuroimage* 112, 70–85. doi: 10.1016/j.neuroimage.2015.02.032
- Burbidge, R., Trotter, M., Buxton, B., and Holden, S. (2001). Drug design by machine learning: support vector machines for pharmaceutical data analysis. *Comput. Chem.* 26, 5–14. doi: 10.1016/S0097-8485(01)00094-8
- Chen, X., Liu, H., and Carbonell, J. G. (2012). “Structured sparse canonical correlation analysis,” in *International Conference on Artificial Intelligence and Statistics* (La Palma), 199–207.
- Chen, Y., Wan, H. I., O'Reardon, J. P., Wang, D. J. J., Wang, Z., Korczykowski, M., et al. (2011). Quantification of cerebral blood flow as biomarker of drug

- effect: arterial spin labeling phMRI after a single dose of oral citalopram. *Clin. Pharmacol. Ther.* 89, 251–258. doi: 10.1038/clpt.2010.296
- Cristianini, N., and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press.
- Dai, W., Garcia, D., de Bazelaire, C., and Alsop, D. C. (2008). Continuous flow-driven inversion for arterial spin labeling using pulsed radio frequency and gradient fields. *Magn. Reson. Med.* 60, 1488–1497. doi: 10.1002/mrm.21790
- Doyle, O. M., Ashburner, J., Zelaya, F. O., Williams, S. C. R., Mehta, M. A., and Marquand, A. F. (2013b). Multivariate decoding of brain images using ordinal regression. *Neuroimage* 81, 347–357. doi: 10.1016/j.neuroimage.2013.05.036
- Doyle, O. M., de Simoni, S., Schwarz, A. J., Brittain, C., O'Daly, O. G., Williams, S. C. R., et al. (2013a). Quantifying the attenuation of the ketamine pharmacological magnetic resonance imaging response in humans: a validation using antipsychotic and glutamatergic agents. *J. Pharmacol. Exp. Ther.* 345, 151–160. doi: 10.1124/jpet.112.201665
- Dudoit, S., Fridlyand, J., and Speed, T. P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* 97, 77–87. doi: 10.1198/016214502753479248
- Elsenbruch, S., Kotsis, V., Benson, S., Rosenberger, C., Reidick, D., Schedlowski, M., et al. (2012). Neural mechanisms mediating the effects of expectation in visceral placebo analgesia: an fMRI study in healthy placebo responders and nonresponders. *Pain* 153, 382–390. doi: 10.1016/j.pain.2011.10.036
- Gaonkar, B., and Davatzikos, C. (2013). Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. *Neuroimage* 78, 270–283. doi: 10.1016/j.neuroimage.2013.03.066
- Grellman, C., Bitzer, S., Neumann, J., Westive, L. T., Andreassen, O. A., Villringer, A., et al. (2015). Comparison of variants of canonical correlation analysis and partial least squares for combined analysis of MRI and genetic data. *Neuroimage* 107, 289–310. doi: 10.1016/j.neuroimage.2014.12.025
- Grosnick, L., Klingenberg, B., Katovich, K., Knutson, B., and Taylor, J. E. (2013). Interpretable whole-brain prediction analysis with GraphNet. *Neuroimage* 72, 304–321. doi: 10.1016/j.neuroimage.2012.12.062
- Groves, A. R., Beckmann, C. F., Smith, S. M., and Woolrich, M. W. (2011). Linked independent component analysis for multimodal data fusion. *Neuroimage* 54, 2198–2217. doi: 10.1016/j.neuroimage.2010.09.073
- Handley, R., Zelaya, F. O., Reinders, A. A. T., Marques, T. R., Mehta, M. A., O'Gorman, R., et al. (2013). Acute effects of single-dose aripiprazole and haloperidol on resting cerebral blood flow (rCBF) in the human brain. *Hum. Brain Mapp.* 34, 272–282. doi: 10.1002/hbm.21436
- Hardoon, D. R., Ettinger, U., Mourão-Miranda, J., Antonova, E., Collier, D., Kumari, V., et al. (2009). Correlation-based multivariate analysis of genetic influence on brain volume. *Neurosci. Lett.* 450, 281–286. doi: 10.1016/j.neulet.2008.11.035
- Hardoon, D. R., Mourao-Miranda, J., Brammer, M., and Shawe-Taylor, J. (2007). Correlation-based multivariate analysis of genetic influence on brain volume. *Neuroimage* 37, 1250–1259. doi: 10.1016/j.neuroimage.2007.06.017
- Hardoon, D. R., and Shawe-Taylor, J. (2011). Sparse canonical correlation analysis. *Mach. Learn.* 83, 331–353. doi: 10.1007/s10994-010-5222-7
- Hardoon, D. R., Szedmak, S., and Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* 16, 2639–2664. doi: 10.1162/0899766042321814
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* 28, 321–377. doi: 10.1093/biomet/28.3-4.321
- Kim, E., Kwon, J. S., Shin, Y. W., Lee, J. S., Kang, W. J., Jo, H. J., et al. (2008). Taq1A polymorphism in the dopamine D2 receptor gene predicts brain metabolic response to aripiprazole in healthy male volunteers. *Pharmacogenet. Genom.* 18, 91–97. doi: 10.1097/FPC.0b013e3282f3ef8c
- Klami, A., Virtanen, S., and Kaski, S. (2013). Bayesian canonical correlation analysis. *J. Mach. Learn. Res.* 14, 965–1003. Available online at: <http://dl.acm.org/citation.cfm?id=2502581.2502612>
- Lahti, A. C., Holcomb, H. H., Weiler, M. A., Medoff, D. R., and Tamminga, C. A. (2003). Functional effects of antipsychotic drugs: comparing clozapine with haloperidol. *Biol. Psychiatry* 53, 601–608. doi: 10.1016/S0006-3223(02)01602-5
- Lahti, A. C., Weiler, M. A., Holcomb, H. H., Tamminga, C. A., and Cropsey, K. L. (2009). Modulation of limbic circuitry predicts treatment response to antipsychotic medication: a functional imaging study in schizophrenia. *Neuropsychopharmacology* 34, 2675–2690. doi: 10.1038/npp.2009.94
- Lahti, A. C., Weiler, M. A., Medoff, D. R., Tamminga, C. A., and Holcomb, H. H. (2005). Functional effects of single dose first-and second-generation antipsychotic administration in subjects with schizophrenia. *Psychiatry Res. Neuroimaging* 139, 19–30. doi: 10.1016/j.pscychresns.2005.02.006
- Le Floch, É., Guillemot, V., Frouin, V., Pinel, P., Lalanne, C., Trinchera, L., et al. (2012). Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse Partial Least Squares. *Neuroimage* 63, 11–24. doi: 10.1016/j.neuroimage.2012.06.061
- Lin, D., Calhoun, V. D., and Wang, Y. P. (2014). Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Med. Image Anal.* 18, 891–902. doi: 10.1016/j.media.2013.10.010
- Lin, D., Zhang, J., Li, J., Calhoun, V. D., Deng, H. W., and Wang, Y. P. (2013b). Group sparse canonical correlation analysis for genomic data integration. *BMC Bioinformatics* 14:245. doi: 10.1186/1471-2105-14-245
- Lin, D., Zhang, J., Li, J., Calhoun, V., and Wang, Y. P. (2013a). “Identifying genetic connections with brain functions in schizophrenia using group sparse canonical correlation analysis,” in *2013 IEEE 10th International Symposium on Biomedical Imaging (ISBI)* (San Francisco, CA: IEEE), 278–281.
- Mailman, R. B., and Murthy, V. (2010). Third generation antipsychotic drugs: partial agonism or receptor functional selectivity? *Curr. Pharm. Des.* 16, 488. doi: 10.2174/138161210790361461
- Marquand, A. F., de Simoni, S., O'Daly, O. G., Williams, S. C., Mourão-Miranda, J., and Mehta, M. A. (2011). Pattern classification of working memory networks reveals differential effects of methylphenidate, atomoxetine, and placebo in healthy volunteers. *Neuropsychopharmacology* 36, 1237–1247. doi: 10.1038/npp.2011.9
- Marquand, A. F., O'Daly, O. G., de Simoni, S., Alsop, D. C., Maguire, R. P., Williams, S. C., et al. (2012). Dissociable effects of methylphenidate, atomoxetine and placebo on regional cerebral blood flow in healthy volunteers at rest: a multi-class pattern recognition approach. *Neuroimage* 60, 1015–1024. doi: 10.1016/j.neuroimage.2012.01.058
- Martinez-Montes, E., Valdes-Sosa, P. A., Miwakeichi, F., Goldman, R. I., and Cohen, M. S. (2004). Concurrent EEG/fMRI analysis by multiway partial least squares. *Neuroimage* 22, 1023–1034. doi: 10.1016/j.neuroimage.2004.03.038
- McMillan, C. T., Toledo, J. B., Avants, B. B., Cook, P. A., Wood, E. M., Suh, E., et al. (2014). Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiol. Aging* 35, 1473–1482. doi: 10.1016/j.neurobiolaging.2013.11.029
- Meinshausen, N., and Bühlmann, P. (2010). Stability selection. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 72, 417–473. doi: 10.1111/j.1467-9868.2010.00740.x
- Michel, V., Gramfort, A., Varoquaux, G., Eger, E., and Thirion, B. (2011). Total variation regularization for fMRI based prediction of behaviour. *IEEE Trans. Med. Imaging* 30, 1328–1340. doi: 10.1109/TMI.2011.2113378
- Miller, D. D., Andreasen, N. C., O'Leary, D. S., Watkins, G. L., Boles Ponto, L. L., and Hichwa, R. D. (2001). Comparison of the effects of risperidone and haloperidol on regional cerebral blood flow in schizophrenia. *Biol. Psychiatry* 49, 704–715. doi: 10.1016/S0006-3223(00)01001-5
- Monteiro, J. M., Rao, A., Ashburner, J., Shawe-Taylor, J., and Mourao-Miranda, J. (2014). “Leveraging clinical data to enhance localisation of brain atrophy,” in *4th NIPS Workshop on Machine Learning and Interpretation in Neuroimaging (MLINI)* (Montreal, QC).
- Nathan, P. J., Phan, K. L., Harmer, C. J., Mehta, M. A., and Bullmore, E. T. (2014). Increasing pharmacological knowledge about human neurological and psychiatric disorders through functional neuroimaging and its application in drug discovery. *Curr. Opin. Pharmacol.* 14, 54–61. doi: 10.1016/j.coph.2013.11.009
- Okuyama, S., Chaki, S., Yoshikawa, R., Suzuki, Y., Ogawa, S. I., Imagawa, Y., et al. (1997). *In vitro* and *in vivo* characterization of the dopamine D4 receptor, serotonin 5-HT_{2A} receptor and alpha-1 adrenoceptor antagonist (R)-(+)-2-amino-4-(4-fluorophenyl)-5-[1-[4-(4-fluorophenyl)-4-oxobutyl] pyrrolidin-3-yl] thiazole (NRA0045). *J. Pharmacol. Exp. Ther.* 282, 56–63.
- Pae, C. U., Serretti, A., Patkar, A. A., and Masand, P. S. (2008). Aripiprazole in the treatment of depressive and anxiety disorders: a review of current evidence. *CNS Drugs* 22, 367–388. doi: 10.2165/00023210-200822050-00002

- Paloyelis, Y., Doyle, O. M., Zelaya, F. O., Maltezos, S., Williams, S. C., Fotopoulou, A., et al. (2014). A spatiotemporal profile of *in vivo* cerebral blood flow changes following intranasal oxytocin in humans. *Biol. Psychiatry*. doi: 10.1016/j.biopsych.2014.10.005. [Epub ahead of print].
- Park, T. (2005). The Bayesian Lasso. *J. Am. Stat. Assoc.* 103, 681–686. doi: 10.1198/016214508000000337
- Parkhomenko, E., Tritchler, D., and Beyene, J. (2009). Sparse canonical correlation analysis with application to genomic data integration. *Stat. Appl. Genet. Mol. Biol.* 8, 1–34. doi: 10.2202/1544-6115.1406
- Ryali, S., Chen, T., Supekar, K., and Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage* 59, 3852–3861. doi: 10.1016/j.neuroimage.2011.11.054
- Sabatini, U., Boulanouar, K., Fabre, N., Martin, F., Carel, C., Colonnese, C., et al. (2000). Cortical motor reorganization in akinetic patients with Parkinson's disease: a functional MRI study. *Brain* 123, 394–403. doi: 10.1093/brain/123.2.394
- Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062
- Sripada, C. S., Kessler, D., Welsh, R., Angstadt, M., Liberzon, I., Phan, K. L., et al. (2013). Distributed effects of methylphenidate on the network structure of the resting brain: a connectomic pattern classification analysis. *Neuroimage* 81, 213–221. doi: 10.1016/j.neuroimage.2013.05.016
- Sui, J., Adali, T., Yu, Q., Chen, J., and Calhoun, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *J. Neurosci. Methods* 204, 68–81. doi: 10.1016/j.jneumeth.2011.10.031
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Methodol.* 58, 267–288.
- Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2003). Class prediction by nearest shrunken centroids, with applications to DNA microarrays. *Stat. Sci.* 18, 104–117. doi: 10.1214/ss/1056397488
- Waaajenborg, S., Verselewe de Witt Hamer, P. C., and Zwinderman, A. H. (2008). Quantifying the association between gene expressions and DNA-markers by penalized canonical correlation analysis. *Stat. Appl. Genet. Mol. Biol.* 7, 1544–6115. doi: 10.2202/1544-6115.1329
- Wild, D. J., Ding, Y., Sheth, A. P., Harland, L., Gifford, E. M., and Lajiness, M. S. (2012). Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research. *Drug Discov. Today* 17, 469–474. doi: 10.1016/j.drudis.2011.12.019
- Witten, D. M., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 10, 515–534. doi: 10.1093/biostatistics/kxp008
- Witten, D. M., and Tibshirani, R. J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data. *Stat. Appl. Genet. Mol. Biol.* 8, 1–27. doi: 10.2202/1544-6115.1470
- Ystad, M., Hodneland, E., Adolfsdottir, S., Haász, J., Lundervold, A. J., Eichele, T., et al. (2011). Cortico-striatal connectivity and cognition in normal aging: a combined DTI and resting state fMRI study. *Neuroimage* 55, 24–31. doi: 10.1016/j.neuroimage.2010.11.016

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Rosa, Mehta, Pich, Risterucci, Zelaya, Reinders, Williams, Dazzan, Doyle and Marquand. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.