

Correlation Based Multivariate Analysis of the Genetic Influence on Brain Volume

David R. Hardoon^{†*}, Ulrich Ettinger[‡],
Janaina Mourão-Miranda[‡], Elena Antonova[‡], David Collier[°],
Veena Kumari^{*}, Steven C.R. Williams[‡], Michael Brammer[‡]

[†]Computational Statistics & Machine Learning Centre
Dept. of Computer Science
University College London
London, WC1E 6BT
E-mail: D.Hardoon@cs.ucl.ac.uk

[‡]Centre for Neuroimaging Sciences
Institute of Psychiatry, De Crespigny Park
Kings College London

[°]Social, Genetic and Developmental Psychiatry Centre
Institute of Psychiatry, De Crespigny Park
King's College London

^{*}Department of Psychology
Institute of Psychiatry, De Crespigny Park
King's College London

9th July 2008

Abstract

Considerable research effort has focused on achieving a better understanding of the genetic correlates of individual differences in volumetric and morphological brain measures. The importance of these efforts is underlined by evidence suggesting that brain changes in a number of neuropsychiatric disorders are at least partly genetic in origin. The currently used methods to study these relationships are mostly based on single-genotype univariate analysis techniques. These methods are limited as multiple genes are likely to interact with each other in their influences on brain structure and function. In this paper we present a feasibility study where we show that by using kernel correlation analysis, with a new genotypes representation, it is possible to analyse the relative associations of several genetic polymorphisms interaction with brain structure. The implementation of the method is demonstrated on genetic and structural magnetic resonance imaging data acquired from a group of 16 healthy subjects by showing the multivariate genetic influence on grey and white matter.

Keywords: Canonical Correlation Analysis, Kernel Methods, Multivariate Genetic Analysis, Multivariate sMRI Analysis

1 Introduction

Macroscopic features of brain morphology and volume are known to be highly heritable (Peper, Brouwer, Boomsma, Kahn & Pol, 2007). Given the observation of volumetric and structural brain changes in a number of neuropsychiatric conditions (Honea, Crow, Passingham & Mackay, 2005; Kempton, Geddes, Ettinger, Williams & Grasby, 2008), considerable research effort has focussed on elucidating these changes in vivo using structural magnetic resonance imaging (MRI). The importance of these efforts is underlined by evidence suggesting that brain changes in a number of disorders are at least partly genetic in origin, as suggested by family and twin studies (Bearden, van Erp, Thompson, Toga & Cannon, 2007; Ettinger, Picchioni, Landau, Matsumoto, van Haren, Marshall, Hall, Schulze, Touloupoulou, Davies, Ribchester, McGuire & Murray, 2007).

Recently, MRI studies have begun to identify specific molecular genetic candidates for these brain changes. In a typical genetic MRI

*Correspondence author

paradigm, participants are classified on the basis of DNA analysis according to a certain genotype (e.g. allelic groups AA, AB, BB; or A-allele carriers vs. non- carriers). Statistical analysis, such as a t-test or analysis of variance (ANOVA), of MRI data is then carried out to identify differences between the groups which are assumed to reflect the influence of the investigated genotype. These single-gene association studies have been useful in characterising the likely mechanisms of risk genes on brain structure and function and, therefore, the pathophysiology of neuropsychiatric conditions (Ho, Wassink, O’Leary, Sheffield & Andreasen, 2005; McIntosh, Baig, Hall, Job, Whalley, Lymer, Moorhead, Owens, Miller, Porteous, & Johnstone, 2007).

However, the analysis of single-gene effects is limited as complex phenotypes are likely influenced by multiple genetic polymorphisms and given that genes are likely to interact in their influences on brain structure and function. The complexity of statistical analysis of such interactions calls for the development of novel statistical techniques that can identify which combination of genotypes can best explain statistical variance in brain structure and function.

Machine learning/pattern recognition methods are increasingly being used to analyze fMRI data. The most commonly employed method, the Support Vector Machine (SVM), a supervised method, associates properties of the imaging data with simple specific categorical labels (e.g., $-1, 1$ indicating experimental conditions 1 and 2). The aim is then to find the hyperplane that gives optimal separation between data belonging to the two classes, using data from the whole brain in each subject, for examples see Mourao-Miranda, Bokde, Born, Hampel & Stetter (2005); Mourao-Miranda, Reynaud, McGlone, Calvert & Brammer (2006). In some situations, the use of categorical labels may not be optimal and in an earlier study (Hardoon, Mourao-Miranda, Brammer & Shawe-Taylor, 2007), we introduced a new unsupervised fMRI analysis method based on Kernel Canonical Correlation Analysis (KCCA) to overcome this problem. KCCA replaces the simple categorical labels used in SVM ($+1/-1$) with a label vector for each stimulus containing details of the features of that stimulus, e.g a simple image label of pleasant/unpleasant is replaced by a vector of image features.

In this paper we extend the unsupervised application of KCCA to the analysis of genotypic effects on brain structure and function. We show that by using KCCA, with a new orthogonalised representation to express the genotype labeling, it is possible to examine the

influences of the interactions of multiple genes on brain structure.

2 Genetic and sMRI Data

2.1 Participants

16 healthy participants (14 males) took part. All participants were right-handed, Caucasian, and psychiatrically, medically and neurologically healthy. Ages ranged from 18 to 36 years (mean=23.69, SD=4.33) and participants had spent an average (SD) of 16.81 (2.64) years in full-time education. Participants provided written, informed consent. The study had permission from the local research ethics committee.

2.2 Genetic Acquisition

DNA was obtained from either buccal swabs or whole blood. The cheek swab method typically provides excellent yield and quality of DNA; cheek swab samples were sent to the laboratory at the Social, Genetic, and Developmental Psychiatry Centre (SGDP) at the Institute of Psychiatry, where DNA was extracted from the samples using established procedures (Freeman, Smith, Curtis, Hockett, Mill & Craig, 2003). DNA was extracted from whole blood by isolation of the white cells followed by a modified phenol-chloroform procedure. After extraction the DNA was re-suspended in TRIS-EDTA (10mM tris pH 7.4, 0.1mM EDTA) buffer and quantified by spectroscopy before storage at -80°C . We selected genes that were found to be associated in the previous literature with brain volume. All samples were genotyped on an Illumina 317K genotyping array by Decode Genetics in Iceland, as part of a separate genetic project (SGENE, European Commission FP6). For the present analysis, SNPs were selected from the array genotype data only from the following genes, and other markers were not analysed:

- ASPM (rs3762271, rs10801589, rs12034362, rs1127661, rs12137359)
- IGF1 (rs10860862, rs6219, rs6214, rs978458, rs2288378, rs7136446, rs10735380, rs1019731, rs2162679, rs35766, rs35765, rs855211)
- IGF2 (rs734351)
- MAOA (rs909525, rs3027409, rs6609257, rs3027415, rs1799836)

- BDNF (rs925946, rs10501087, rs2203877, rs6265, rs11030104, rs10835211, rs7934165, rs12273363, rs908867, rs1491850)
- APOE (rs405509)
- SHH (rs1233556)
- Plexin B3 (rs4898439, rs762650, rs762651)
- MCPH1 (rs2920616, rs4840940, rs1057187, rs6995735, rs2442546, rs894888, rs17076812, rs1968586, rs1129703, rs1129706, rs2034143, rs2442502, rs2440399, rs12674488, rs2920689, rs2440445, rs2922806, rs2442473, rs2979666, rs2442632, rs3780088, rs3020213, rs1961222, rs2515464, rs2515466, rs2515477, rs2959812, rs4841224, rs2959809, rs2922876, rs2897911, rs2922873, rs4478599, rs2515493, rs921291, rs2515507, rs3020242, rs2922861, rs2922859, rs2442579, rs2442573, rs2442572, rs4841336, rs2442567, rs10100002, rs2959802, rs1257, rs2959799, rs2013938, rs1057090, rs2959797, rs2911968, rs2912065, rs2980654, rs2433146, rs1057091)

2.3 MRI Acquisition

Participants underwent MRI scanning using a General Electric Signa Advantage scanner at 1.5 Tesla. A 3-dimensional T1-weighted, coronal, spoiled gradient (SPGR) of the whole head was obtained. Acquisition followed realigning along the inter-hemispheric fissure and the AC-PC line. The sequence used an echo time of 5.1ms, a repetition time of 18ms, a flip angle of 20° and a field of view of 240×240×192mm for a resulting voxel dimension of 0.9375×0.9375×1.5 mm. Grey/white matter discrimination was achieved by means of an inversion time of 450ms.

2.4 MRI Preprocessing

First, each structural image was reoriented to the antero-posterior commissure line of the Montreal Neurological Institute (MNI) template. Second, structural data were preprocessed following the procedure used for voxel-based morphometry method with the SPM5 software¹. SPM5 implements a unified segmentation/normalisation framework, which is a single probabilistic model combining tissue classification approach in native space, non-uniformity correction, and

¹<http://www.fil.ion.ucl.ac.uk/spm/>

nonlinear registration to the standard (MNI) space in one procedure (Ashburner & Friston, 2005). Third, segmented and normalised images were modulated with Jacobian determinates, which involves scaling each image by the amount of contraction incurred during nonlinear warping. This step allows preserving the total amount of tissue in the modulated grey or white matter image as in the original image, yielding the estimate of grey or white matter volume. Finally, since normalization procedure in SPM5 is of much higher accuracy than in SPM2 and modulation step somewhat smoothes the images, the grey and white matter tissue segments were smoothed with 8 mm FWHM Gaussian kernel, which approximates smoothness of 12 mm obtained using SPM2 software². The grey and white matter probability images were resliced with $2 \times 2 \times 2$ voxel size, since higher resolution of $1 \times 1 \times 1$ would considerably increase the dimensionality of the classification problem, distributing any pattern in the data over a much broader area in term of search space for the classifier.

3 Multivariate Analysis Methodology

CCA is a technique, proposed by Hotelling (1936) for finding pairs of basis vectors that maximise the correlation of a set of paired variables. These pairs can be considered as two “views” of the same object. This technique is applicable in cases where each “view” contains, as a subspace, all “relevant” information plus some “irrelevant” information. CCA identifies a projection space containing the relevant subspaces in both views. This projection space is often referred to as the semantic space. In the following study we consider the SNPs of the genetic sequence and the segmented white or grey matter of structural MRI brain scans to be two “views” of the same object. CCA seeks a pair of linear transformations one for each of the paired variables such that when the variables are transformed the corresponding coordinates are maximally correlated.

Consider the linear combination

$$\begin{aligned}x &= \mathbf{w}'_a \mathbf{x} \\ y &= \mathbf{w}'_b \mathbf{y}.\end{aligned}$$

Let \mathbf{x} and \mathbf{y} be two random variables from a multi-dimensional distribution, with zero mean. The maximisation of the correlation between

²<http://dbm.neuro.uni-jena.de/vbm/segmentation/modulation/>

x and y corresponds to solving

$$\max_{\mathbf{w}_a, \mathbf{w}_b} \rho = \mathbf{w}'_a C_{\mathbf{ab}} \mathbf{w}_b$$

subject to $\mathbf{w}'_a C_{\mathbf{aa}} \mathbf{w}_a = \mathbf{w}'_b C_{\mathbf{bb}} \mathbf{w}_b = 1$. $C_{\mathbf{aa}}$ and $C_{\mathbf{bb}}$ are the non-singular “within-set” covariance matrices and $C_{\mathbf{ab}}$ is the “between-sets” covariance matrix.

We suggest using the kernel variant of CCA (Fyfe & Lai, 2001; Bach & Jordan, 2002) to overcome a potential limitation of the linearity of CCA, namely that useful descriptors may not be extracted from the data if they exist in terms of some non-linear relationship. The kernelising of CCA (KCCA) offers an alternate solution by first projecting the data into a higher dimensional feature space

$$\phi : \mathbf{x} = (x_1, \dots, x_n) \rightarrow \phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_N(\mathbf{x})) \quad (N \geq n)$$

before performing CCA in the new feature space. Given the kernel functions κ_a and κ_b let

$$\begin{aligned} K_{\mathbf{a}} &= \mathbf{X}_a \mathbf{X}'_a \\ K_{\mathbf{b}} &= \mathbf{X}_b \mathbf{X}'_b \end{aligned}$$

be the kernel matrices corresponding to the two representations of the data, where \mathbf{X}_a is the matrix whose rows are the vectors $\phi_a(\mathbf{x}_i)$, $i = 1, \dots, \ell$ from the first representation (SNPs) while \mathbf{X}_b is the matrix with rows $\phi_b(\mathbf{x}_i)$ from the second representation (structural MRI). The weights \mathbf{w}_a and \mathbf{w}_b can be expressed as a linear combination of the training examples

$$\begin{aligned} \mathbf{w}_a &= \mathbf{X}_a \boldsymbol{\alpha} \\ \mathbf{w}_b &= \mathbf{X}_b \boldsymbol{\beta}. \end{aligned}$$

Substituting into the primal CCA equation gives the optimisation

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \rho = \boldsymbol{\alpha}' \mathbf{K}_{\mathbf{a}} \mathbf{K}_{\mathbf{b}} \boldsymbol{\beta} \tag{1}$$

subject to $\boldsymbol{\alpha}' \mathbf{K}_{\mathbf{a}}^2 \boldsymbol{\alpha} = \boldsymbol{\beta}' \mathbf{K}_{\mathbf{b}}^2 \boldsymbol{\beta} = 1$. This is the dual form of the primal CCA optimisation problem given above, which can be cast as a generalised eigenvalue problem and for which the first k generalised eigenvectors can be found efficiently. Both CCA and KCCA can be formulated as an eigenproblem (Hardoon, Szedmak & Shawe-Taylor,

2004). Despite the potential benefit of moving into a non-linear kernel defined feature space we limit ourselves to a linear kernel. This self-imposed restriction allows us to compute the corresponding weights in the input space, i.e to compute the weightings of SNPs or voxels in grey or white matter following KCCA. One might ask what is the benefit of KCCA using a linear kernel over CCA. The answer is that the same results are achieved but that KCCA is significantly less computationally intensive as we only need to consider a kernel matrix of the size of the subjects whereas in CCA a covariance matrix of the size of sMRI voxels needs to be computed.

The naïve application of CCA in kernel space will be likely to produce perfect correlations between the two views. These correlations can therefore fail to distinguish between spurious features and those that capture the underlying semantics (Bach & Jordan, 2002; Kuss & Graepel, 2002; Hardoon et al., 2004; Shawe-Taylor & Cristianini, 2004; Fukumizu, Bach & Gretton, 2006). To control this issue, we require a regularisation on the derived directions to remove the spurious correlation found via the naive application. We use the regularisation formulation given in Hardoon et al. (2004), which has a lower and upper bound of $0 \leq \tau \leq 1$, where we define τ to be the regularisation parameter. The regularised KCCA optimisation is the same as given in equation (1) but now subject to

$$\begin{aligned}\boldsymbol{\alpha}'\mathbf{K}_a((1-\tau)\mathbf{K}_a+\tau\mathbf{I})\boldsymbol{\alpha} &= 1 \\ \boldsymbol{\beta}'\mathbf{K}_b((1-\tau)\mathbf{K}_b+\tau\mathbf{I})\boldsymbol{\beta} &= 1.\end{aligned}$$

Note that when using $\tau > 0$ it is important to a-posteriori re-normalise

$$\begin{aligned}\boldsymbol{\alpha} &= \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\mathbf{K}_a^2\boldsymbol{\alpha}} \\ \boldsymbol{\beta} &= \frac{\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{K}_b^2\boldsymbol{\beta}}\end{aligned}$$

so that the original KCCA³ conditions $\boldsymbol{\alpha}'\mathbf{K}_a^2\boldsymbol{\alpha} = \boldsymbol{\beta}'\mathbf{K}_b^2\boldsymbol{\beta} = 1$ hold, this will ensure that the corresponding ρ is in fact a correlation value.

4 Experiment

In the following section we elaborate on our experimental set-up, analyses and results. Our data set consisted of 16 subjects. Genotyping

³The KCCA Matlab package was obtained from <http://academic.davidroiardoon.com>

of the sample yielded 94 SNPs in each subject. Each SNP consists of 3 genotypes, i.e. a subject is for any given SNP we examined a homozygote for allele A, an AB heterozygote, or a homozygote for allele B. Each subject also completed a structural MRI scan of the brain as described above. In our analysis we masked the voxels in the structural MR images using grey and white matter masks thresholded at voxel intensity values of 0.2 for both grey and white tissue images. This value was derived by visual inspection of the individual images. The regularisation parameter, from Section 3, is heuristically fixed to $\tau = 0.03$ for the present preliminary study. The rigorous optimisation of this regularisation parameter will be addressed in a future study.

In the multivariate analysis we are unable to use the conventional univariate label representation of the SNP genotype of 1, 2 and 3 due to the fact that we do not know a-priori the direction of each of the SNPs vector, i.e. whether the genotype representation $1 < 2 < 3$ or $1 > 2 > 3$ is true throughout the subjects 94 SNPs, although the direction for each SNP across the subjects is the same. We overcome this issue by representing the SNP genotypes in a 3-dimensional orthogonalised label such that $1 \perp 2 \wedge 1 \perp 3 \wedge 2 \perp 3$. In other words we use the following representation of the genotypes

$$\begin{aligned} 1 &\rightarrow [0, 0, 1] \\ 2 &\rightarrow [0, 1, 0] \\ 3 &\rightarrow [1, 0, 0] \end{aligned}$$

as our corresponding label representation of the 3 genotypes.

Following the KCCA procedure, as detailed in Section 3, we wish to analyse the effect of all the SNP's genotypes have on the entire, masked, brain. The inherited structure of KCCA entails that we implicitly incorporate the effect of the voxels on each other. This would have also been true for the effect of the SNP's on each other but our orthogonalised representation entails that each SNP's mean weight value will be zero. In other words, what we examined was the relative influence of the three different genotypes within the SNPs on grey and white matter. The resulting output weights from the proposed technique represent, on the one hand, the changes in brain matter volume and on the other, the corresponding relative influence of the genotypes within the SNP's on the change in brain matter volume.

We wish to avoid overinterpreting our results due to the small sample size, therefore we confine ourselves to plot in Figure 1 representing

each SNP’s largest contributing (absolute) genotype value, this allows us to visualise the overall influence each SNP has on the grey and white matter. We normalise the 94 SNP vector so that it will be comparable between the grey and white plots.

Table 1: In the following table we list the SNP’s from Figure 1 that have a greater weight influence than 0.16 on grey and white matter. The SNP number corresponds to a fairly random numbering of the SNPs in the order as they are presented in the Method section.

SNP (grey)	SNP weight (grey)	SNP (white)	SNP weight (white)
(45) rs17076812	0.1664	(19) rs909525	0.1713
(47) rs1129703	0.1664	(21) rs6609257	0.1635
(48) rs1129706	0.1664	(64) rs2515477	0.1601
(49) rs2034143	0.1664	(80) rs2442572	0.1665
(51) rs2440399	0.1608	(84) rs2959802	0.1676
(52) rs12674488	0.1652	(86) rs2959799	0.1681
(72) rs2515493	0.1717	(87) rs2013938	0.1634
(80) rs2442572	0.1745	(88) rs1057090	0.1676
		(90) rs2911968	0.1676

5 Discussion

The current study presented a method for the analysis of complex gene-MRI associations. We aimed to explore whether KCCA could usefully be applied to genetic (SNPs in genes thought to be related to human brain volume) and neuroimaging (grey and white matter extracted from structural MRI) data to detect multivariate correlations between genetic and structural imaging data. The results of the study are shown in two contexts: (1) the impact of genetic variability at each SNP on structural variation in grey or white matter volume and (2) the individual weightings of voxels in grey and white matter in terms of how much their variability was correlated with SNP variability across all SNPs.

Figure 1 shows the SNP “loading” profiles with voxelwise volume variation of grey and white matter, e.g the relative correlations of

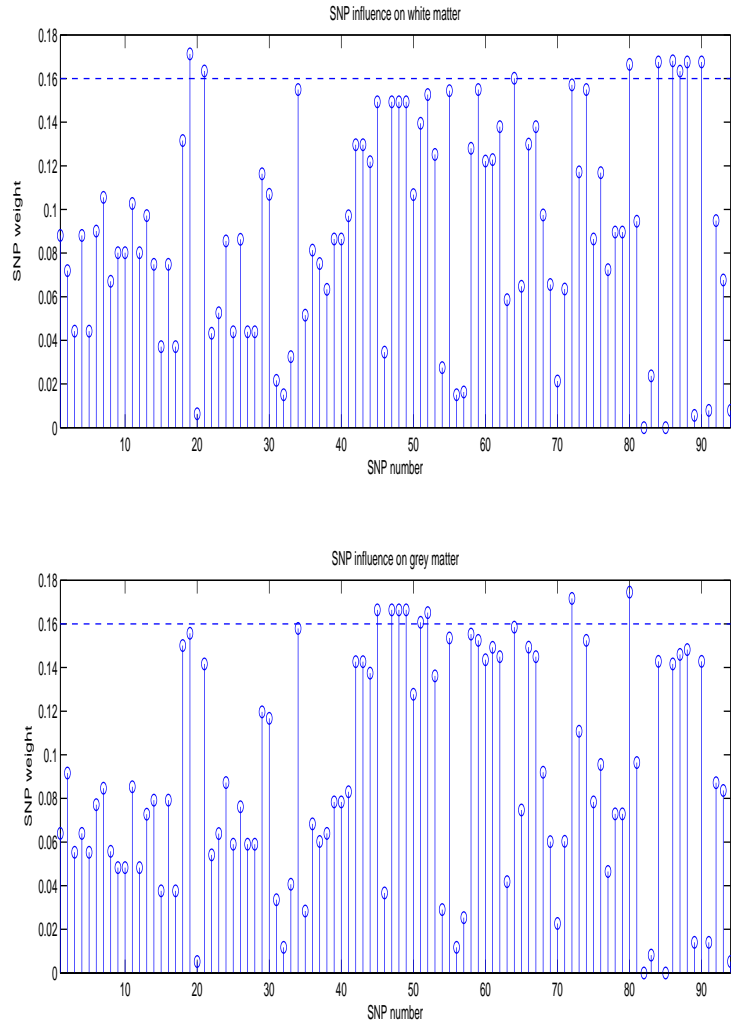


Figure 1: We plot the SNP's largest contributing (absolute) genotype value.

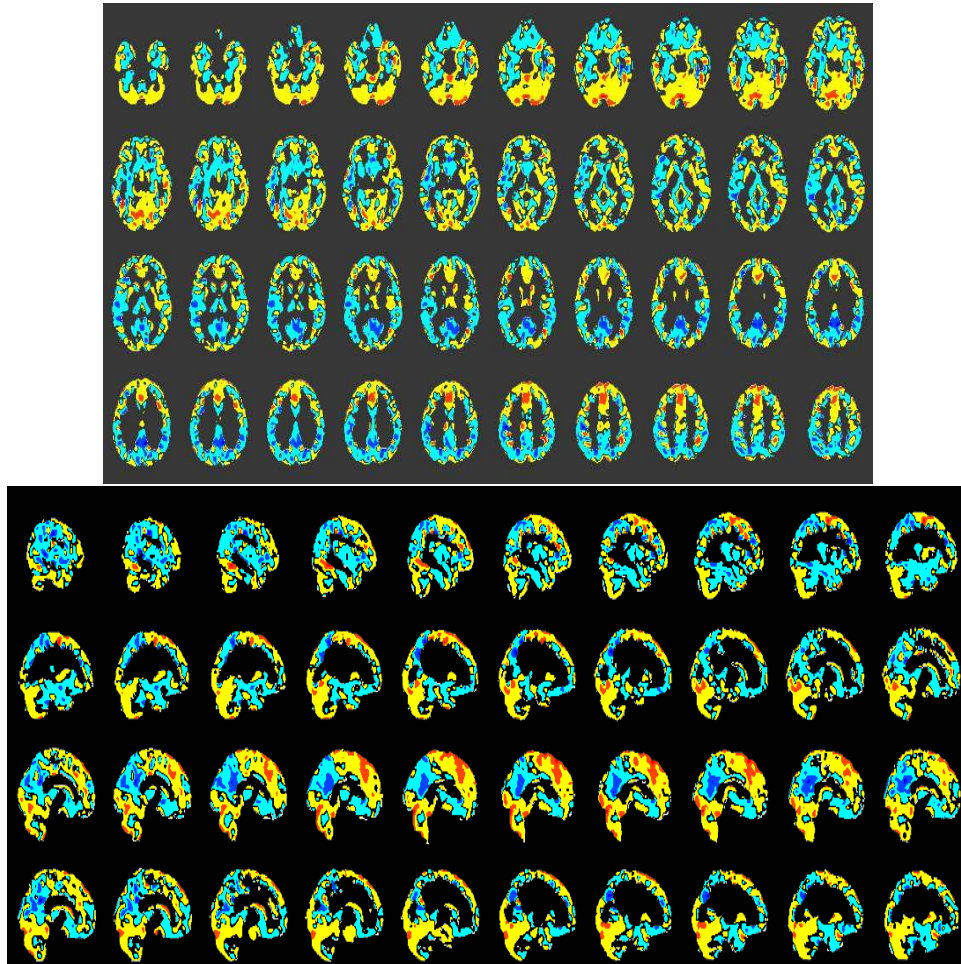


Figure 2: Weighted MRI of grey matter volume influence. The blue and red represent opposite change influence where blue is the negative values and red is the positive values.

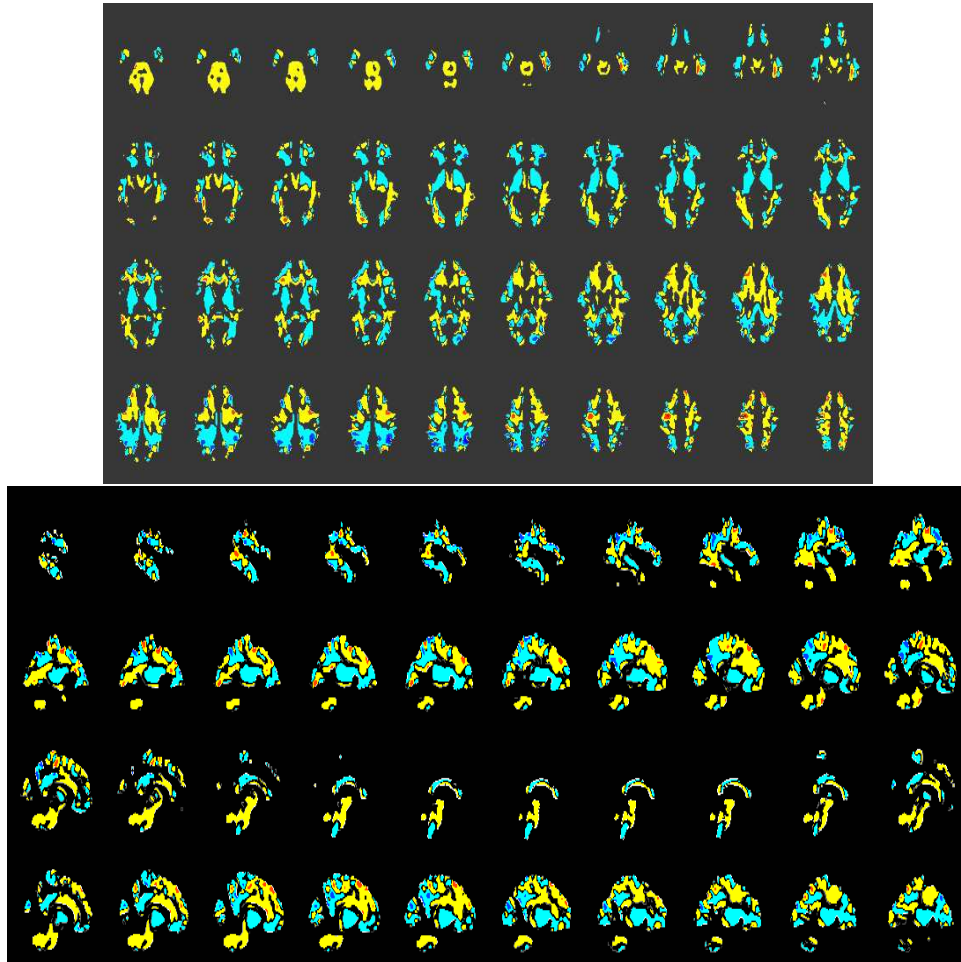


Figure 3: Weighted MRI of white matter volume influence. The blue and red represent opposite change influence where blue is the negative values and red is the positive values.

different SNPs with voxelwise volumetric variation within grey and white matter areas of the brain. We heuristically threshold at 0.16 so to give the top $\approx 10\%$ of SNP with the largest influence. The SNP's with the largest apparent correlations with volumetric variation are shown in Table 1. Figures 2 and 3 show the voxel-wise loadings in grey and white matter volume of SNP correlations, e.g how much these voxels are under the combined genetic influence of all the SNPs studied.

We should be very conservative in our interpretations of the results from a small, preliminary study. However, it appears (a) the method enables us to place some ranking on the possible importance of SNPs in driving structural brain change and (b) that the brain areas under apparent SNP-related influence fall into a number of broad regions. The regions shown in red and blue are affected in opposite directions but we do not at present read these as increases and decreases in grey or white matter volume. Considered our voxel size and a relatively coarse smoothing kernel in the present analysis, the results should be applied as representing associations between genes and large-scale volumetric variations. For fine-grained volumetric variations a higher image resolution (i.e. $1 \times 1 \times 1$) with smaller smoothing kernel (e.g. 6mm) should be applied.

Broadly speaking, the red areas include cerebellum, occipital cortex, anterior cingulate and some lateral frontal regions. The blue regions include posterior cingulate, parietal cortex and some temporal regions. We refrain at present from placing firm interpretations on these results pending replication and extension to larger groups.

We conclude by saying that KCCA with a linear kernel is capable of investigating the interactional effects of multiple genetic influences on brain structure. We hope that the possibilities raised by this method will be usefully exploited to clarify the nature of multiple gene influences on brain structure and function in large samples of healthy and diseased subjects.

Acknowledgements

We are grateful for partial support from an EC FP6 grant, Project 037761, "SGENE". David R. Hardoon is supported by the EPSRC project Le Strum, EP-D063612-1.

References

- Ashburner, J. & Friston, K. (2005). Unified segmentation. *Neuroimage*, *26*(3), 839–851.
- Bach, F. & Jordan, M. (2002). Kernel independent component analysis. *Journal of Machine Learning Research*, *3*, 1–48.
- Bearden, C. E., van Erp, T. G., Thompson, P. M., Toga, A. W. & Cannon, T. D. (2007). Cortical mapping of genotype-phenotype relationships in schizophrenia. *Human Brain Mapping*, *28*, 519–532.
- Ettinger, U., Picchioni, M., Landau, S., Matsumoto, K., van Haren, N. E., Marshall, N., Hall, N. H., Schulze, K., Touloupoulou, T., Davies, N., Ribchester, T., McGuire, P. & Murray, R. M. (2007). Magnetic resonance imaging of the thalamus and adhesio interthalamica in twins with schizophrenia. *Arch Gen Psychiatry*, *64*, 401–409.
- Freeman, B., Smith, N., Curtis, C., Hockett, L., Mill, J. & Craig, I. W. (2003). Dna from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behav Genet*, *33*, 67–72.
- Fukumizu, K., Bach, F. R. & Gretton, A. (2006). Consistency of kernel canonical correlation analysis. *Journal of Machine Learning Research*, *8*, 361–383.
- Fyfe, C. & Lai, P. L. (2001). Kernel and nonlinear canonical correlation analysis. *International Journal of Neural Systems*, *10*, 365–377.
- Hardoon, D. R., Mourao-Miranda, J., Brammer, M. & Shawe-Taylor, J. (2007). Unsupervised analysis of fmri data using kernel canonical correlation. *NeuroImage*, *37* (4), 1250–1259.
- Hardoon, D. R., Szedmak, S. & Shawe-Taylor, J. (2004). Canonical correlation analysis: an overview with application to learning methods. *Neural Computation*, *16*, 2639–2664.
- Ho, B., Wassink, T. H., O’Leary, D. S., Sheffield, V. C. & Andreasen, N. C. (2005). Catechol-o-methyl transferase val158met gene polymorphism in schizophrenia: working memory, frontal lobe mri morphology and frontal cerebral

- blood flow. *Mol Psychiatry*, 10, 287–229.
- Honea, R., Crow, T. J., Passingham, D. & Mackay, C. E. (2005). Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies. *Am J Psychiatry*, 162, 2233–2245.
- Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28, 312–377.
- Kempton, M. J., Geddes, J. R., Ettinger, U., Williams, S. C. & Grasby, P. M. (2008). Meta-analysis, database and meta-regression of 98 structural imaging studies in bipolar disorder. *ArchGenPsychiatry*, *In Press*, –.
- Kuss, M. & Graepel, T. (2002). The geometry of kernel canonical correlation analysis. Technical report, Max Planck Institute for Biological Cybernetics.
- McIntosh, A. M., Baig, B. J., Hall, J., Job, D., Whalley, H. C., Lymer, G. K., Moorhead, T. W., Owens, D. G., Miller, P., Porteous, D., , S. M. L. & Johnstone, E. C. (2007). Relationship of catechol-o-methyltransferase variants to brain structure and function in a population at high risk of psychosis. *BiolPsychiatry*, 61, 1127–1134.
- Mourao-Miranda, J., Bokde, A. L. W., Born, C., Hampel, H. & Stetter, S. (2005). Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. *NeuroImage*, 28, 980–995.
- Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G. & Brammer, M. (2006). The impact of temporal compression and space selection on svm analysis of single-subject and multi-subject fmri data. *NeuroImage*, 33:4, 1055–1065.
- Peper, J. S., Brouwer, R. M., Boomsma, D. I., Kahn, R. S. & Pol, H. E. H. (2007). Genetic influences on human brain structure: a review of brain imaging studies in twins. *HumBrain Mapp*, 28, 464–473.
- Shawe-Taylor, J. & Cristianini, N. (2004). *Kernel Methods for Pattern Analysis*. Cambridge University Press.