Neuroscience Letters 450 (2009) 281–286

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Correlation-based multivariate analysis of genetic influence on brain volume

David R. Hardoon[a,*], Ulrich Ettinger[b], Janaina Mourão-Miranda[b], Elena Antonova[b], David Collier[c], Veena Kumari[d], Steven C.R. Williams[b], Michael Brammer[b]

[a] Computational Statistics & Machine Learning Centre, Dept. of Computer Science, University College London, London WC1E 6BT, United Kingdom
[b] Centre for Neuroimaging Sciences, Institute of Psychiatry, De Crespigny Park, King's College London, United Kingdom
[c] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, De Crespigny Park, King's College London, United Kingdom
[d] Department of Psychology, Institute of Psychiatry, De Crespigny Park, King's College London, United Kingdom

ARTICLE INFO

Article history:
Received 11 August 2008
Received in revised form 11 November 2008
Accepted 14 November 2008

Keywords:
Canonical correlation analysis
Kernel methods
Multivariate genetic analysis
Multivariate sMRI analysis

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 with brain structure. The implementation of the method is demonstrated on genetic and structural magnetic resonance imaging (MRI) data acquired from a group of 16 healthy subjects by showing the multivariate genetic influence on grey and white matter.

Macroscopic features of brain morphology and volume are known to be highly heritable [14]. Given the observation of volumetric and structural brain changes in a number of neuropsychiatric conditions [8,10], 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 [2,3].

Recently, MRI studies have begun to identify specific molecular genetic candidates for these brain changes. In a typical genetic MRI paradigm, participants are classified on the basis of DNA analysis according to a certain genotype (e.g. genotype 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 [7,11].

However, the analysis of single-gene effects is limited as complex phenotypes are likely influenced by multiple genetic polymorphisms. Also, there are likely to be effects of epistasis, i.e. gene–gene interactions, on brain 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 et al. [12,13]. In some situations, the use of categorical labels may not be optimal and in an earlier study [6], 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. We show that by

* Correspondence author. Tel: +44 20 7679 0425; fax: +44 20 7387 1397.
E-mail address: D.Hardoon@cs.ucl.ac.uk (D.R. Hardoon).

0304-3940/$ – see front matter © 2008 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2008.11.035

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.

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, S.D. = 4.33) and participants had spent an average (S.D.) 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.

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 [5]. DNA was extracted from whole blood by isolation of the white cells followed by a modified phenlol-chloroform procedure. After extraction the DNA was re-suspended in TRIS-EDTA (10 mM tris pH 7.4, 0.1 mM 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 317 K 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).

Genotyping was performed using an Illumina Infinium HumanHap 300 k genotyping bead array according to the manufacturers protocols. All samples were genotyped individually. For quality control, arrays yielding a call rate below 99.9% were excluded from the study. The SNPs on the illumina array have been selected to avoid having many SNPs in strong linkage disequilibrium (LD), as this would yield redundant information.

Participants underwent MRI scanning using a General Electric Signa Advantage scanner at 1.5 T. A three-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 5.1 ms, a repetition time of 18 ms, a flip angle of 20° and a field of view of 240 mm × 240 mm × 192 mm for a resulting voxel dimension of 0.9375 mm × 0.9375 mm × 1.5 mm. Grey/white matter discrimination was achieved by means of an inversion time of 450 ms.

First, each structural image was reoriented to the anteroposterior 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 software.[1] SPM5 implements a unified segmentation/normalisation framework, which is a single probabilistic model combining tissue classification approach in native space, non-uniformity correction, and nonlinear registration to the standard (MNI) space in one procedure [1]. Third, segmented and normalised images were modulated with Jacobian determinates, which involves scaling each image by the amount of contraction incurred during non-linear 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 software.[2] 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.

CCA is a technique, proposed by Hotelling [9] 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 refered 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. In this paper we use the kernel variant of CCA. Due to space limitation we refer the reader to [6] for exact details and derivations.

In the following section we elaborate on our experimental setup, analyses and results. Our data set consisted of 16 subjects. Genotyping of the sample yielded 94 SNPs in each subject. Each SNP is coded as 3 genotypes, i.e. a subject is for any given SNP examined here an A-allele homozgote (AA), a heterozygote (AB), or a B-allele homozygote (BB). 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 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 (or AA), 2 (or AB) and 3 (or BB) 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 for each

[1] http://www.fil.ion.ucl.ac.uk/spm/.
[2] http://dbm.neuro.uni-jena.de/vbm/segmentation/modulation/.

SNP, although the direction for each SNP across the subjects is the same. We overcome this issue by representing the SNP genotypes in a three-dimensional orthogonalised label such that $1 \perp 2 \bigwedge 1 \perp 3 \bigwedge 2 \perp 3$. In other words we use the following representation of the genotypes:

$1 \rightarrow [0, 0, 1]$
$2 \rightarrow [0, 1, 0]$
$3 \rightarrow [1, 0, 0]$

as our corresponding label representation of the 3 genotypes. This method therefore also allows for the identification of heterozygosity effects (e.g. AB > AA = BB). The analysis thus is of additive genetic effects.

Following the KCCA procedure we wish to analyse the effects of all SNPs on 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 address the question of methods reliability by randomly splitting the samples into two equal groups and comparing the resulting analysis of the groups, i.e. we compute a correlation value between the two resulting normalised genetic and MRI weight maps. We repeat this procedure 100 times and give the average correlation values and standard deviation in Table 1. The high correlation value and low standard deviation represent an agreement for the different analyses which in turn constitute an argument for the reliability of the proposed method.

Table 1
Average correlation values for the genetic and MRI analysis on a group comparison study.

	Genetics	MRI
Grey	0.8707 ± 0.0289	0.8891 ± 0.0810
White	0.8111 ± 0.0455	0.8920 ± 0.1015

The subjects were split into two random groups of eight subjects and the correlation of the resulting analysis between the two groups was computed. This process was repeated 100 times and is reported with the standard deviation.

To avoid overinterpreting our results due to the small sample size, we confine ourselves to a plot in Fig. 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 we are able to compare between the grey and white resulting weight vectors. The individual allele contribution for each of the SNP weightings is given in Fig. 2. While there is a difference between the genetic influence on white and grey matter it is not apparent. We separate the individual genotypes weights and plot a direct comparison in Fig. 3.

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.

Fig. 1 shows the SNP "loading" profiles with voxelwise volume variation of grey and white matter, e.g. the relative correlations of different SNPs with voxelwise volumetric variation within grey and

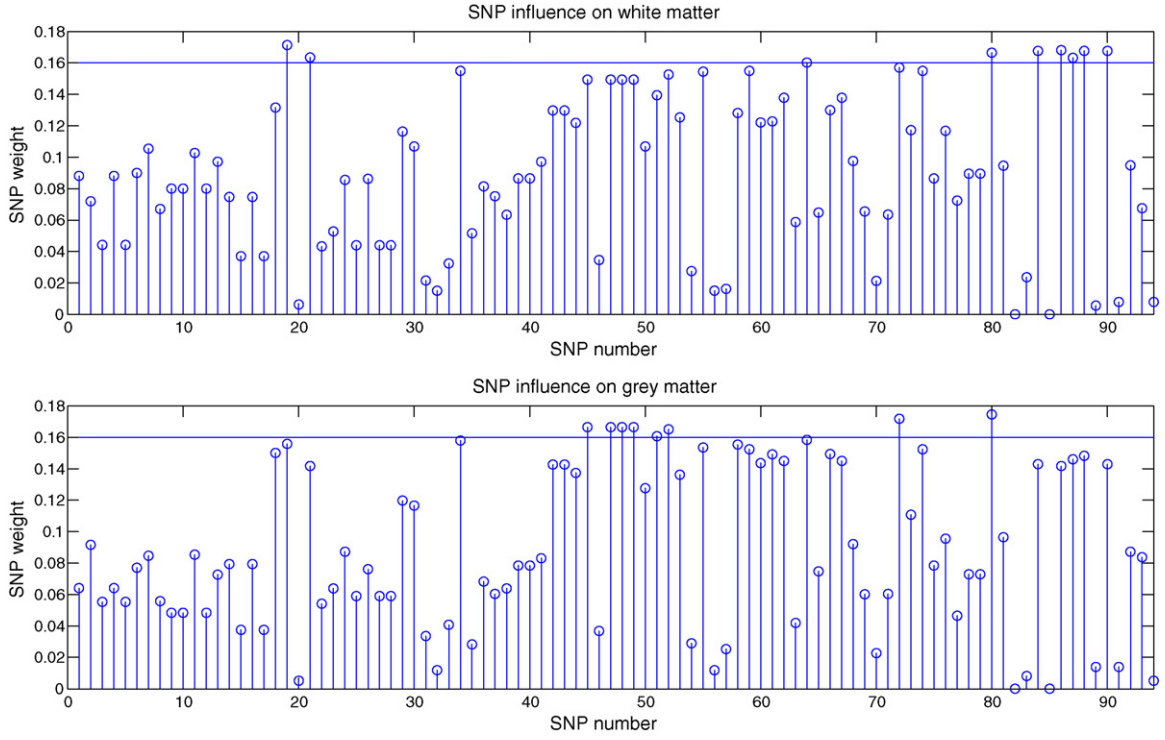

Fig. 1. The contributing largest SNP (absolute) genotype value. For each SNP the allele with the largest, absolute, value is given.

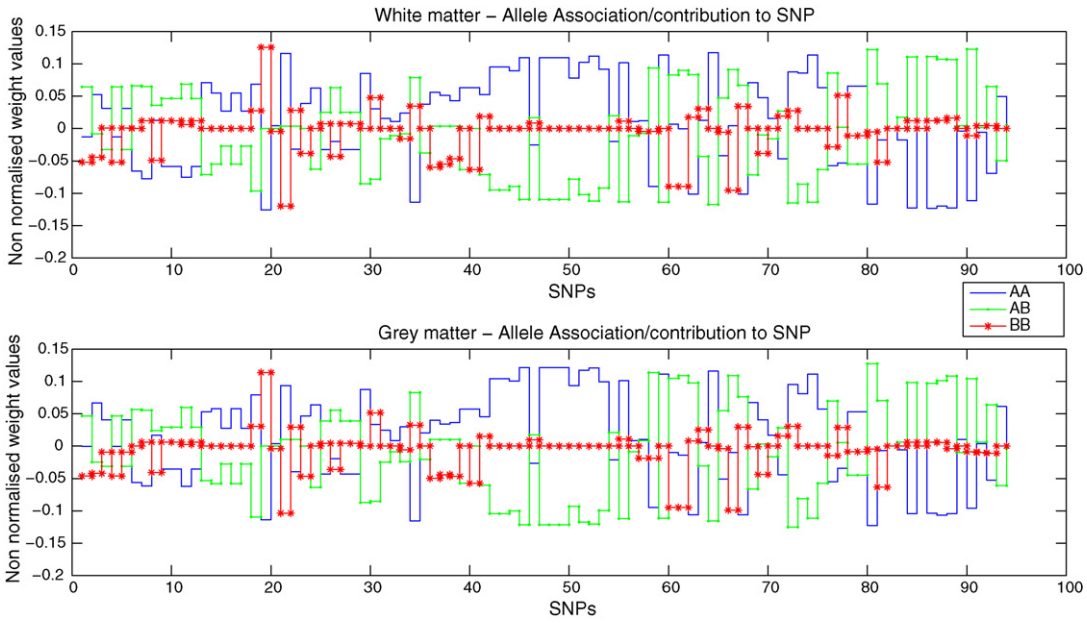

Fig. 2. Genotype (AA, AB, BB) contributions to changes in grey and white matter volume.

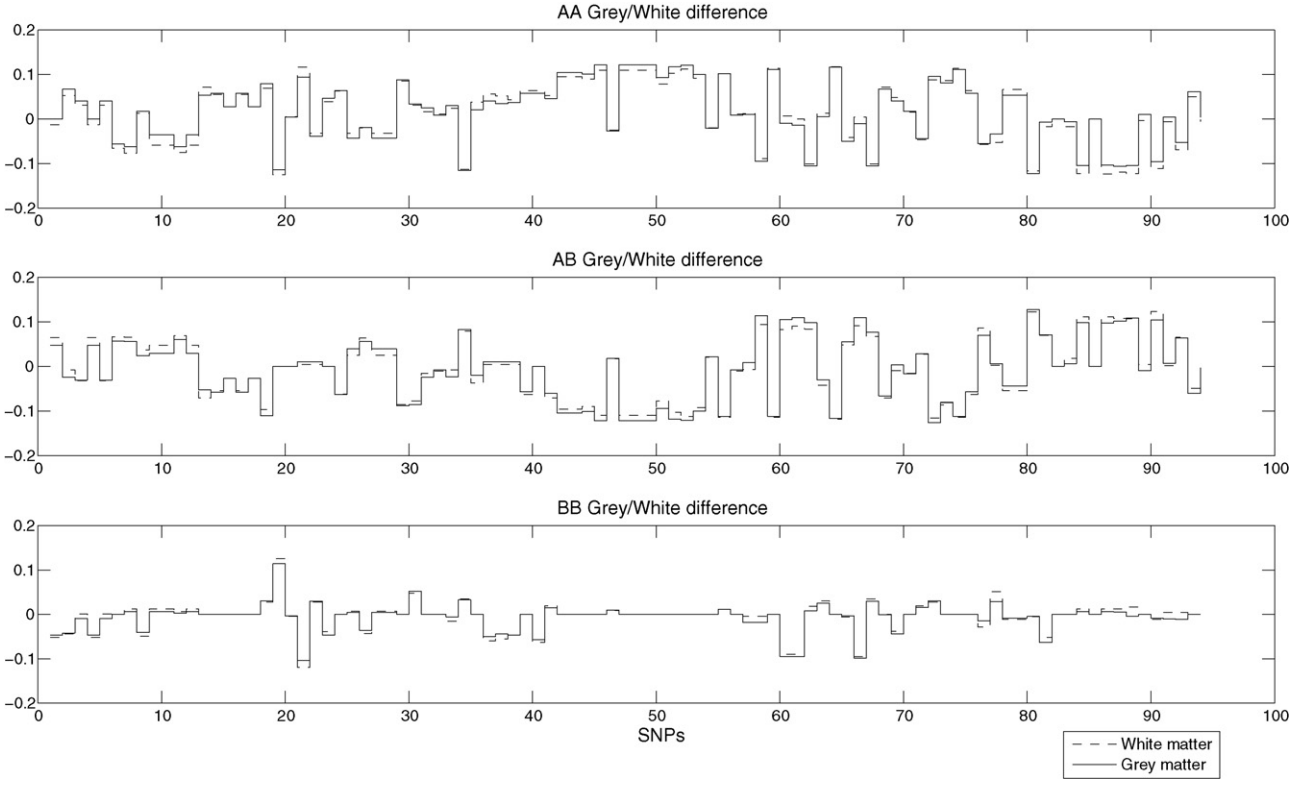

Fig. 3. The differences between influences on grey and white matter volume by genotype (AA, AB, BB).

white matter areas of the brain. We heuristically threshold at 0.16 so to give the top ≈ 10% of SNP with the largest influence. The SNP's with the largest apparent correlations with volumetric variation are shown in Table 2. Figs. 4 and 5 shows 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. The maps were displayed using AFNI.[3] The colour bar indicates mapping from the KCCA weights to colours. To generate a colour bar that is symmetric around zero the values in the weight maps are rescaled in such a way that the absolute maximum is assign the value of $+1$ and the colour scale runs from -1 to $+1$. The absolute maximum was 12.74 for the gray matter weight and 25.73 for the white matter weight. Both weights were normalised (normalised $\mathbf{w} = (\mathbf{w}/(\|\mathbf{w}\|_2))$) and multiplied by 1000. For each map the threshold corresponded a 20% of their respective maximum.

We should be very conservative in our interpretations of the results from a small, preliminary study. However, it appears (a) the

[3] http://afni.nimh.nih.gov/.

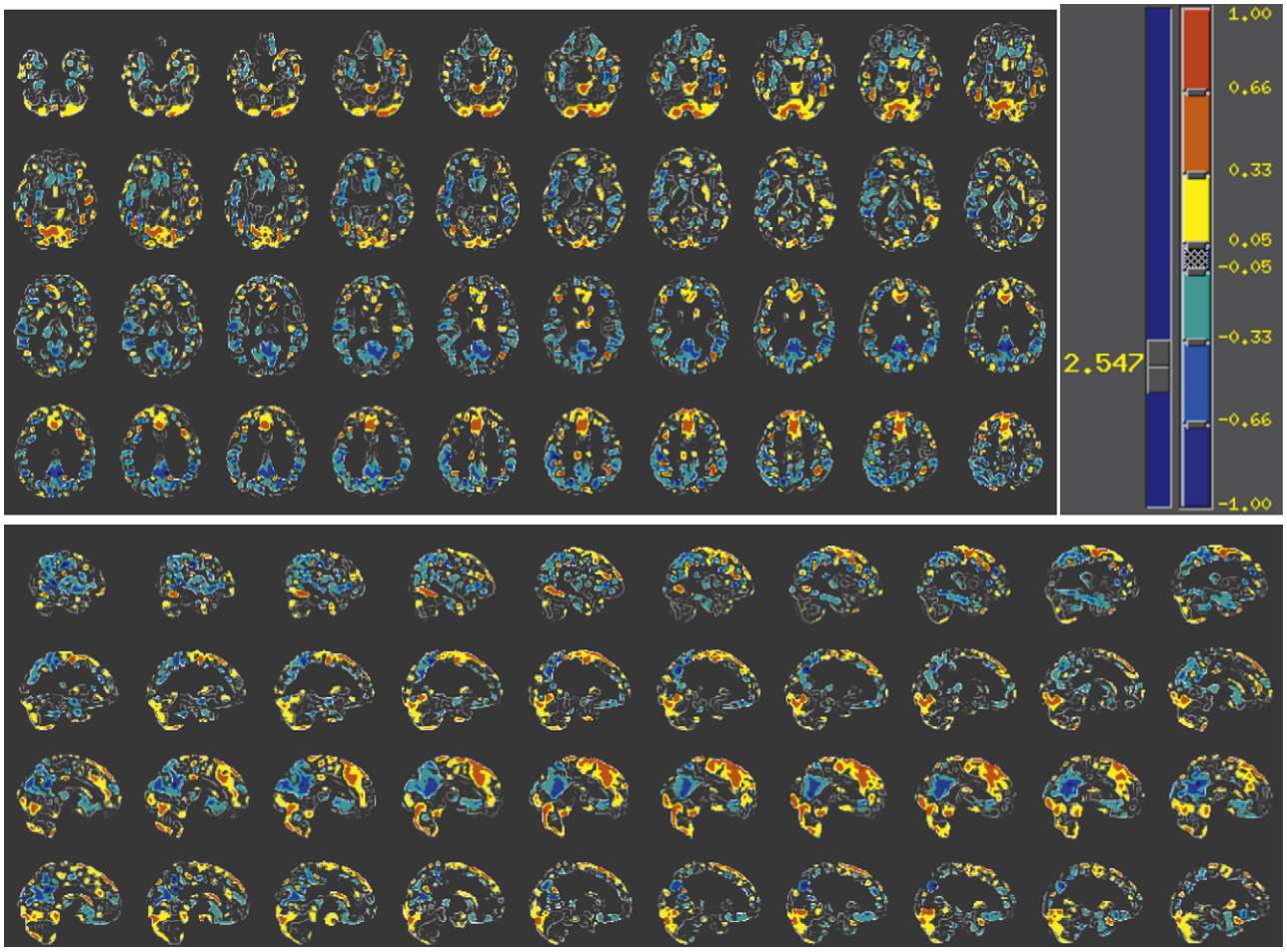

Fig. 4. Weighted MRI of genetic influence on grey matter volume. The colour bar indicates mapping from the KCCA weights to colours. The blue and red represent opposite change influence where blue is the negative values and red is the positive values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

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. Considering 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. 6 mm) should be applied.

Table 2
SNPs with a weight influence greater than 0.16 on grey or white matter.

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

The SNP number in parentheses corresponds to the numbering of the SNPs in the order as they are presented.

Broadly speaking, the red areas include cerebellum, occipital cortex, anterior cingulate and 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. With this caveat in mind, however, it is of interest to note that all except two (rs909525 and rs6609257) SNPs with a weight influence of more than 0.16 (see Table 2) are from the microcephalin (MCPH1) gene. MCPH1 is thought to be involved in the regulation of human brain size [4] and has been shown to be associated with cranial volume in a recent study [15]. However, it should be borne in mind that linkage disequilibrium amongst SNPs in the same gene is common; this potential problem should be addressed in future studies.

We conclude by arguing that KCCA with a linear kernel has potential to investigate 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.

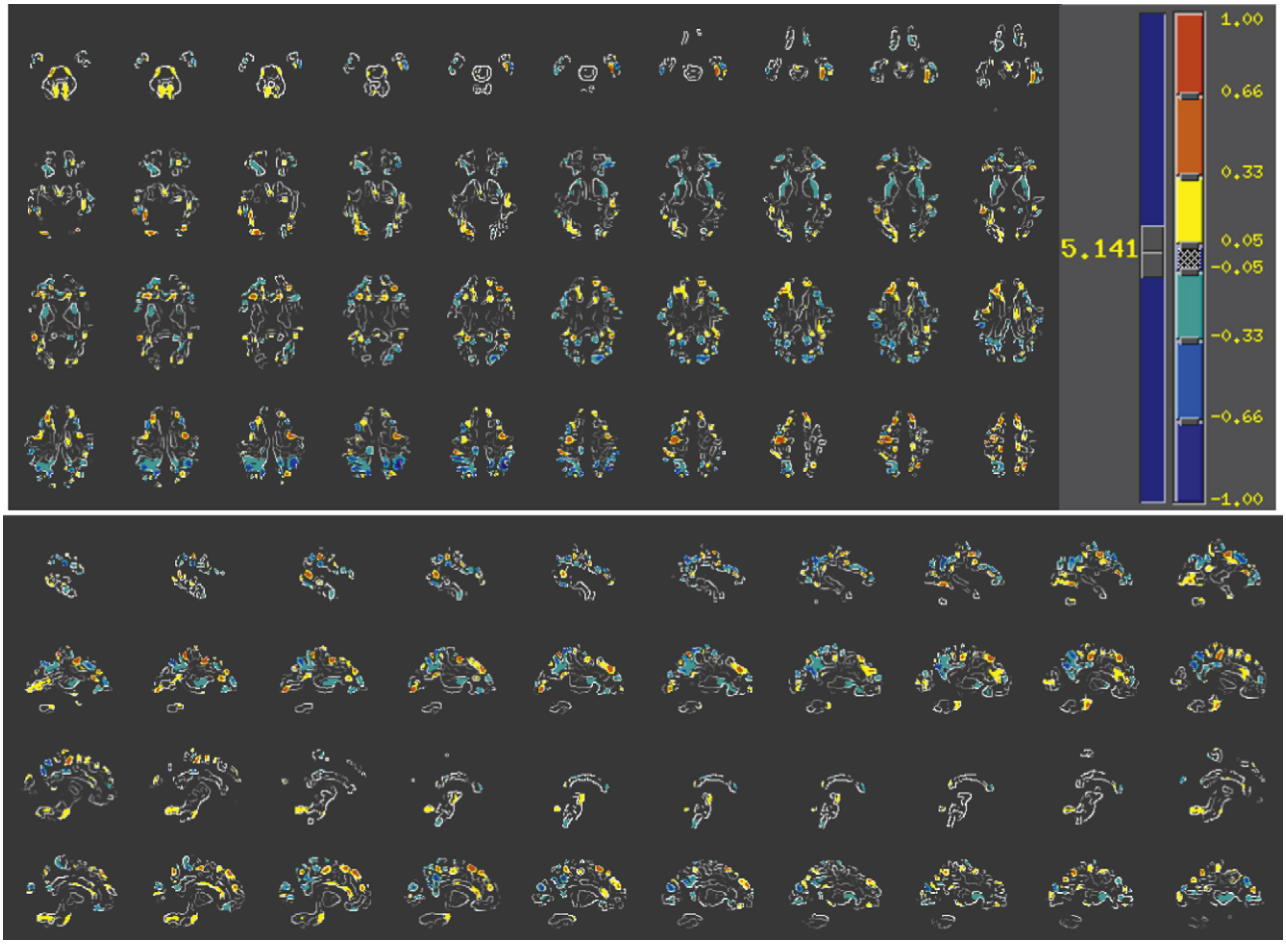

Fig. 5. Weighted MRI of genetic influence on white matter volume. The colour bar indicates mapping from the KCCA weights to colours. The blue and red represent opposite change influence where blue is the negative values and red is the positive values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

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. Ulrich Ettinger is funded by a NIHR (National Institute for Health Research) Personal Award. The views expressed in this publication are those of the authors and not necessarily those of the NHS, NIHR or Department of Health.

References

[1] J. Ashburner, K. Friston, Unified segmentation, Neuroimage 26 (3) (2005) 839–851.
[2] C.E. Bearden, T.G. van Erp, P.M. Thompson, A.W. Toga, T.D. Cannon, Cortical mapping of genotype-phenotype relationships in schizophrenia, Hum. Brain Mapp. 28 (2007) 519–532.
[3] U. Ettinger, M. Picchioni, S. Landau, K. Matsumoto, N.E. van Haren, N. Marshall, N.H. Hall, K. Schulze, T. Toulopoulou, N. Davies, T. Ribchester, P. McGuire, R.M. Murray, Magnetic resonance imaging of the thalamus and adhesio interthalamica in twins with schizophrenia, Arch. Gen. Psychiatry 64 (2007) 401–409.
[4] P.D. Evans, S.L. Gilbert, N. Mekel-Bobrov, E.J. Vallender, J.R. Anderson, L.M. Vaez-Azizi, S.A. Tishkoff, R.R. Hudson, B.T. Lahn, Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans, Science 209 (5741) (2005) 1717–1720.
[5] B. Freeman, N. Smith, C. Curtis, L. Huckett, J. Mill, I.W. Craig, 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 (2003) 67–72.
[6] D.R. Hardoon, J. Mourao-Miranda, M. Brammer, J. Shawe-Taylor, Unsupervised analysis of FMRI data using kernel canonical correlation, NeuroImage 37 (4) (2007) 1250–1259.
[7] B. Ho, T.H. Wassink, D.S. O'Leary, V.C. Sheffield, N.C. Andreasen, Catechol-o-methyl transferase val158met gene polymorphism in schizophrenia: working memory, frontal lobe MRI morphology and frontal cerebral blood flow, Mol. Psychiatry 10 (2005) 229–287.
[8] R. Honea, T.J. Crow, D. Passingham, C.E. Mackay, Regional deficits in brain volume in schizophrenia: a meta-analysis of voxel-based morphometry studies, Am. J. Psychiatry 162 (2005) 2233–2245.
[9] H. Hotelling, Relations between two sets of variates, Biometrika 28 (1936) 312–377.
[10] M.J. Kempton, J.R. Geddes, U. Ettinger, S.C. Williams, P.M. Grasby, Meta-analysis, database and meta-regression of 98 structural imaging studies in bipolar disorder, Arch. Gen. Psychiatry 65 (9) (2008) 1017–1032.
[11] A.M. McIntosh, B.J. Baig, J. Hall, D. Job, H.C. Whalley, G.K. Lymer, T.W. Moorhead, D.G. Owens, P. Miller, D. Porteous, S.M. Lawrie, E.C. Johnstone, Relationship of catechol-o-methyltransferase variants to brain structure and function in a population at high risk of psychosis, Biol. Psychiatry 61 (2007) 1127–1134.
[12] J. Mourao-Miranda, A.L.W. Bokde, C. Born, H. Hampel, S. Stetter, Classifying brain states and determining the discriminating activation patterns: support vector machine on functional MRI data, NeuroImage 28 (2005) 980–995.
[13] J. Mourao-Miranda, E. Reynaud, F. McGlone, G. Calvert, M. Brammer, The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject FMRI data, NeuroImage 33 (4) (2006) 1055–1065.
[14] J.S. Peper, R.M. Brouwer, D.I. Boomsma, R.S. Kahn, E. Hulshoff Pol, Genetic influences on human brain structure: a review of brain imaging studies in twins, Hum. Brain Mapp. 28 (2007) 464–473.
[15] J.K. Wang, Y. Li, B. Su, A common SNP of MCPH1 is associated with cranial volume variation in Chinese population, Hum. Mol. Genet. 17 (9) (2008) 1329–1335.