

# On the Brain Structure Heterogeneity of Autism: Parsing out Acquisition Site Effects With Significance-Weighted Principal Component Analysis

Francisco Jesús Martínez-Murcia,<sup>1</sup> Meng-Chuan Lai,<sup>2,3,4,5</sup>  
Juan Manuel Górriz,<sup>1\*</sup> Javier Ramírez,<sup>1</sup> Adam M. H. Young,<sup>4</sup>  
Sean C. L. Deoni,<sup>6</sup> Christine Ecker,<sup>7,8</sup> Michael V. Lombardo,<sup>4,9</sup>  
MRC AIMS Consortium,<sup>†</sup> Simon Baron-Cohen,<sup>4,10</sup> Declan G. M. Murphy,<sup>7,8</sup>  
Edward T. Bullmore,<sup>10,11</sup> and John Suckling<sup>10,11</sup>

<sup>1</sup>Department of Signal Theory Networking and Communications, C/Periodista Daniel Saucedo Aranda S/N, E-18071, University of Granada, Granada, Spain

<sup>2</sup>Child and Youth Mental Health Collaborative at The Centre for Addiction and Mental Health and The Hospital for Sick Children, Toronto, Ontario, Canada

<sup>3</sup>Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

<sup>4</sup>Department of Psychiatry, Autism Research Centre, University of Cambridge, Cambridge, United Kingdom

<sup>5</sup>Department of Psychiatry, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan

<sup>6</sup>Advanced Baby Imaging Laboratory, School of Engineering, Brown University, Providence, Rhode Island

<sup>7</sup>Department of Forensic and Neurodevelopmental Sciences, Sackler Institute for Translational Neurodevelopment, London, United Kingdom

<sup>8</sup>Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

<sup>9</sup>Department of Psychology and Center for Applied Neuroscience, University of Cyprus, Nicosia, Cyprus

The Medical Research Council Autism Imaging Multicentre Study Consortium (MRC AIMS Consortium) is a UK collaboration between the Institute of Psychiatry (IoP) at King's College, London, the Autism Research Centre, University of Cambridge, and the Autism Research Group, University of Oxford. The Consortium members are in alphabetical order: Anthony J. Bailey (Oxford), Simon Baron-Cohen (Cambridge), Patrick F. Bolton (IoP), Edward T. Bullmore (Cambridge), Sarah Carrington (Oxford), Marco Catani (IoP), Bhisudev Chakrabarti (Cambridge), Michael C. Craig (IoP), Eileen M. Daly (IoP), Sean C. L. Deoni (IoP), Christine Ecker (IoP), Francesca Happé (IoP), Julian Henty (Cambridge), Peter Jezzard (Oxford), Patrick Johnston (IoP), Derek K. Jones (IoP), Meng-Chuan Lai (Cambridge), Michael V. Lombardo (Cambridge), Anya Madden (IoP), Diane Mullins (IoP), Clodagh M. Murphy (IoP), Declan G. M. Murphy (IoP), Greg Pasco (Cambridge), Amber N. V. Ruigrok (Cambridge), Susan A. Sadek (Cambridge), Debbie Spain (IoP), Rose Stewart (Oxford), John Suckling (Cambridge), Sally J. Wheelwright (Cambridge), and Steven C. Williams (IoP).

Contract grant sponsor: UK Medical Research Council AIMS network; Contract grant number: G0400061; Contract grant sponsors:

“Vicerrectorado de Relaciones Internacionales de la Universidad de Granada” and CEI BioTic Granada; Contract grant: “Convocatoria de Movilidad Internacional de Estudiantes de Doctorado Curso 2013/2014”; Contract grant sponsor: MINECO/FEDER; Contract grant numbers: TEC2012-34306 and TEC2015-64718-R; Contract grant sponsor: Consejería de Economía, Innovación, Ciencia y Empleo (Junta de Andalucía, Spain); Contract grant numbers: P09-TIC-4530 and P11-TIC-7103

Download the code used in this article at <https://github.com/SiPBA/swpca>

\*Correspondence to: Juan Manuel Górriz, Department of Signal Theory Networking and Communications, C/ Periodista Daniel Saucedo Aranda S/N, E-18071, University of Granada, Granada, Spain. E-mail: [gorriz@ugr.es](mailto:gorriz@ugr.es)

Received for publication 20 January 2016; Revised 7 October 2016; Accepted 13 October 2016.

DOI: 10.1002/hbm.23449

Published online 24 October 2016 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)).

<sup>10</sup>Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, United Kingdom

<sup>11</sup>Department of Psychiatry, Brain Mapping Unit, University of Cambridge, Cambridge, United Kingdom

**Abstract:** Neuroimaging studies have reported structural and physiological differences that could help understand the causes and development of Autism Spectrum Disorder (ASD). Many of them rely on multisite designs, with the recruitment of larger samples increasing statistical power. However, recent large-scale studies have put some findings into question, considering the results to be strongly dependent on the database used, and demonstrating the substantial heterogeneity within this clinically defined category. One major source of variance may be the acquisition of the data in multiple centres. In this work we analysed the differences found in the multisite, multi-modal neuroimaging database from the UK Medical Research Council Autism Imaging Multicentre Study (MRC AIMS) in terms of both diagnosis and acquisition sites. Since the dissimilarities between sites were higher than between diagnostic groups, we developed a technique called Significance Weighted Principal Component Analysis (SWPCA) to reduce the undesired intensity variance due to acquisition site and to increase the statistical power in detecting group differences. After eliminating site-related variance, statistically significant group differences were found, including Broca's area and the temporo-parietal junction. However, discriminative power was not sufficient to classify diagnostic groups, yielding accuracies results close to random. Our work supports recent claims that ASD is a highly heterogeneous condition that is difficult to globally characterize by neuroimaging, and therefore different (and more homogenous) subgroups should be defined to obtain a deeper understanding of ASD. *Hum Brain Mapp* 38:1208–1223, 2017. © 2016 The Authors Human Brain Mapping Published by Wiley Periodicals, Inc.

**Key words:** autism spectrum disorder; structural magnetic resonance imaging; structural heterogeneity; voxel based morphometry

## INTRODUCTION

Autism Spectrum Disorder (ASD) is a neurodevelopmental syndrome characterized by social and communication impairment as well as restricted, repetitive patterns of behaviour, interests or activities. The delimitation of both functionally and structurally affected areas in the brain in such an etiologically and neurobiologically heterogeneous condition has long been a major concern [Ecker and Murphy, 2014; Lai et al., 2013; Lenroot and Yeung, 2013]. With this context, the use of large samples is of fundamental importance, and has been addressed by establishing multicentre collaborations such as the UK Medical Research Council Autism Imaging Multi-centre Study (MRC AIMS) [Ecker et al., 2012, 2013, 2015] and the Autism Brain Imaging Data Exchange (ABIDE) [Di Martino et al., 2014]

Multicentre studies with structural (sMRI) and functional magnetic resonance imaging (fMRI) are increasingly common, allowing for recruitment of larger samples in shorter periods of time. However, the use of images acquired at different sites still poses a major challenge. In addition to logistical difficulties, such as regulatory approvals and data protection, a number of technical and methodological issues can potentially affect the resulting maps, introducing undesired intensity and geometric variance.

### Abbreviations

AAL	Advanced Automated Labelling
AAM	Active Appearance Models
ABIDE	Autism Brain Imaging Data Exchange
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADOS	Autism Diagnostic Observation Schedule
ANOVA	Analysis of variance
ASD	Autism Spectrum Disorder
CBM	Component-based morphometry
CSF	Cerebrospinal fluid
GM	Grey matter
IR	Inversion Recovery
MNI	Montreal Neurological Institute
MRI	Magnetic resonance imaging
PCA	Principal component analysis
SBM	Source based morphometry
SPGR	Spoilt gradient recall
SSPF	Steady state free procession
SVC	Support vector classifier
SVD	Singular value decomposition
SWPCA	Significance-weighted principal component analysis
VBM	Voxel-based-morphometry
WASI	Wechsler abbreviated scale of intelligence
WM	White matter

This issue has been addressed in other neurological conditions, such as Alzheimer's Disease [Jovicich et al., 2006; Stonnington et al., 2008], where group differences are well known, and demonstrating that the impact of a correction for site on the resulting neurobiological differences is relatively small. However, these effects have a stronger impact in psychiatric conditions where the atypical radiological signs on MRI are often subtle and require large samples of patients to observe on-average differences relative to control samples. Recent meta-analyses point to differences being inconsistently reported in schizophrenia [Friedman and Glover, 2006; Turner et al., 2013], psychosis [Clementz et al., 2016; Wang et al., 2015], and ASD (using the multi-centre ABIDE database) [Haar et al., 2014].

These inconsistencies can arise from a variety of variance sources, ranging from the multi-level (phenotypic, neurobiological, and etiological) heterogeneities of the conditions to technical issues that include differences in scanner make, model, manufacturer, static field strength, field inhomogeneities, slew rates and image reconstruction [Van Horn and Toga, 2009], as well as acquisition problems such as within-acquisition participant head motion. Field inhomogeneities are a source of misinterpretation of the data even when the same MRI system manufacturer and model are used [Van Horn and Toga, 2009]. Furthermore, results in [Pearlson, 2009] demonstrate that a single scanner can change with time, which makes some widely used strategies, for example collecting controls first and patients later, a flawed approach. Recent neuroimaging research on ASD [Haar et al., 2014] has shown that, while analyses performed on a particular database (acquired on a single platform) could yield coherent regions, the atypical structures are often inconsistent across the wider literature using different databases. Therefore, new methodologies focused on reducing multi-site variance may be potentially helpful in increasing the power to identify the characteristic neurobiological signature of autism, should there be one.

A number of approaches have been proposed to reduce between-site variance. Geometric distortions caused by magnetic fields inhomogeneities have been widely studied [Jovicich et al., 2006; Stonnington et al., 2008]. Furthermore, there exist a number of diffeomorphic registration algorithms, such as DARTEL [Ashburner, 2007] or ANTS [Avants et al., 2010], intended to reduce inter-subject—and by extension, inter-site—variations. In the case of intensity variance, the issue has already been addressed in the MRC AIMS database by leveraging sMRI sequences that yield quantitative estimates of relaxation times [Deoni et al., 2008] that have been demonstrated to reduce single-site effects compared to weighted sequences. However, images acquired using this technique still yield between-site differences [Suckling et al., 2014].

To address the problem of intensity variance and improve the homogeneity of the images across different sites, we have developed a new post-acquisition

methodology that enhances derived maps (e.g. grey or white matter volumes) by means of ameliorating site effects. This method, called significance-weighted principal component analysis (SWPCA), can be applied as part of pre-processing before computing whole-brain or regional statistical analysis. The algorithm proceeds by performing a principal component analysis (PCA) over the whole database of images and later computing the statistical significance of each component in relation to a categorical variable, in this case the acquisition site. This information is used to reconstruct the datasets using a weighted strategy that effectively reduces intensity inhomogeneities due to site effects.

This article is organised as follows: First, in “Material and Methods” section, the methodology is presented. It comprises the presentation of the MRC AIMS image database and its pre-processing, PCA, one-way analysis of variance (ANOVA), and how these two statistical procedures are combined in a weighted approach to create the corrected maps. In “Results” section, the algorithm is applied to maps of grey and white matters and estimates of relaxation times, with qualitative and quantitative results presented before and after applying SWPCA to the images. “Discussion” section is used to discuss the relevance of the SWPCA algorithm and the results of the case-control comparison, and finally in “Conclusions” section, we draw conclusions concerning multi-site studies and the neurobiology of autism.

## MATERIAL AND METHODS

### Image Database

Structural MRI were analysed from 136 adult, right-handed males (68 with ASD and 68 matched controls) with no significant mean differences in age and full-scale IQ, acquired from the centres contributing to the UK Medical Research Council Autism Imaging Multi-centre Study (MRC AIMS) [Ecker et al., 2012, 2013, 2015] and recruited by advertisement. In this work, only participants recruited at the Institute of Psychiatry, King's College London (LON) and the Autism Research Centre, University of Cambridge (CAM) were included where an equivalent set of images were acquired from each participant.

Participants were excluded from the study if they had a history of major psychiatric disorder or medical illness affecting brain function (e.g. psychosis or epilepsy), or current drug misuse (including alcohol), or were taking anti-psychotic medication, mood stabilizers or benzodiazepines.

All participants with ASD were diagnosed according to International Classification of Diseases, 10th Revision (ICD-10) research criteria, and confirmed using the Autism Diagnostic Interview-Revised (ADI-R) [Lord et al., 1994]. Autism Diagnostic Observation Schedule (ADOS) [Lord et al., 2000] was performed, but the score was not considered as an inclusion criteria. ASD participants, to be

**TABLE I. Demographics of the participants included in the analysis**

Database	Group	<i>N</i>	Age ( $\mu \pm \sigma$ years)	IQ ( $\mu \pm \sigma$ )
LON	ASD	39	28.74 $\pm$ 6.52	111.28 $\pm$ 13.13
	CTL	40	25.30 $\pm$ 6.62	104.67 $\pm$ 11.16
CAM	ASD	29	26.83 $\pm$ 4.64	115.83 $\pm$ 11.88
	CTL	28	26.75 $\pm$ 7.32	115.25 $\pm$ 13.67
ALL	ASD	68	25.90 $\pm$ 6.95	109.03 $\pm$ 13.31
	CTL	68	27.93 $\pm$ 5.87	113.22 $\pm$ 12.81

included, must have scored above the ADI-R cut-off in the three domains of impaired reciprocal social interaction, communication and repetitive behaviours and stereotyped patterns, although failure to reach cut-off in one of the domains by one point was permitted. Intellectual ability was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI) [Wechsler, 1999], ensuring the participants fell within the high-functioning range on the spectrum defined by a full-scale IQ > 70. The demographics of the participants are shown in detail in Table 1.

Structural MRI were obtained using Driven Equilibrium Single Pulse Observation of T1 and T2 (DESPOT1, DESPOT2) [Deoni et al., 2008] at King’s College London and University of Cambridge, both with 3T GE Medical Systems HDx scanners. Using multiple spoiled gradient recall (SPGR) acquisitions in the DESPOT1 sequence and steady state free precession (SSPF) acquisitions in the DESPOT2 sequence, with different flip angles and repetition times, quantitative T1 and T2 maps (qT1 and qT2) were calculated with a custom ImageJ plug-in package. Correction of main and transmit magnetic field (B0 and B1) inhomogeneity effects was performed during the estimation of T1 and T2.

For accurate registration to the standard stereotaxic space of the Montreal Neurological Institute (MNI), a simulated T1-weighted Inversion Recovery (IR) images (synT1) were created based on the qT1 maps (Ecker et al., 2012, 2013; Lai et al., 2012). The synT1 images were then segmented using New Segment into grey (GM) and white matter (WM) maps, and normalized to the MNI space using DARTEL in SPM8 [Friston et al., 2007], with modulation (preserve volume) to retain information of regional/local GM and WM volumes, and smoothed with a 3mm FWHM Gaussian Kernel to account for inter-subject mis-registration. The synT1, qT1 and qT2 maps were also registered to the standard MNI space using the same DARTEL flow fields, but without modulation (preserve concentration) to retain information of regional/local T1 contrast, T1 relaxation time, and T2 relaxation times, and smoothed with a 3mm FWHM Gaussian kernel. Therefore, there were five different modalities: qT1, qT2, synT1 map, GM and WM maps, for each every participant, which allows us to observe the impact of our SWPCA correction of site-related undesired variance on quantitative (qT1 and

qT2), simulated (synT1) images and probability maps (GM and WM).

During the pre-processing of the images, several procedures targeted the reduction of inter-subject and inter-site geometric distortion, amongst them the correction of B0 and B1 field inhomogeneity effects and the registration to MNI space. Many other algorithms have been proposed to help in this task. However, the study of their relative performance lies beyond the scope of this article. Following image registration, it was assumed that only the intensity of the maps was affected between sites.

### Intracerebral Mask

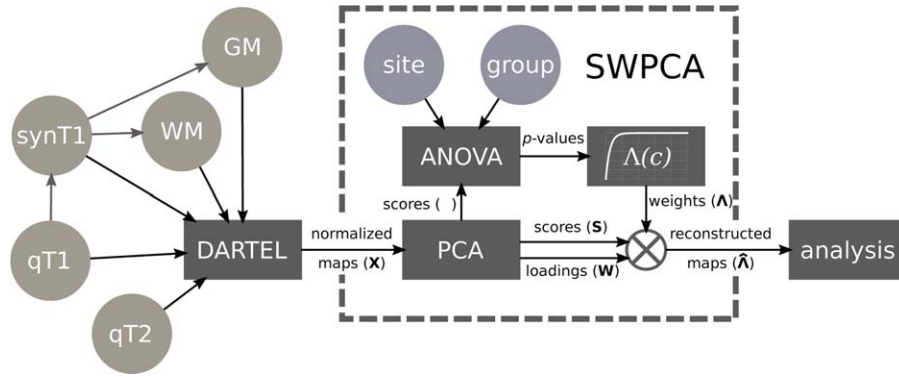
Prior to any processing of the brain images, masks were applied that restricted the analysis to the brain parenchyma. These binary masks comprised voxels in the DARTEL study-specific template with GM (or WM) probabilities > 0.3, whilst excluding voxels with a cerebrospinal fluid (CSF) tissue probability > 0.3, to avoid including CSF regions where T1 and T2 values are extremely high. Throughout this work three masks were used: selectively GM (excluding WM and CSF); selectively WM (excluding GM and CSF); both GM and WM (excluding CSF).

### Significance Weighted Principal Component Analysis

The SWPCA is an algorithm to reduce, in this case, undesired intensity variance introduced by multi-site image acquisition. SWPCA takes any dataset of pre-processed images, spatially normalized, and decomposes them into their variance components to then provide a corrected dataset where these undesired variance components have been reduced. To do so, PCA was applied to each modality in turn to obtain the component scores and component loadings. Since PCA is a data-driven approach, it was only used to decompose the source images, and after this procedure, a one-way ANOVA estimated the relation between each variance component and a given categorical variable, in our case, the acquisition site. The between-site variability in the variance component was then identified by its corresponding *P*-value. Finally, these *P*-values were transformed into a weighting matrix  $\Lambda$  that weighted the influence of each variance component in a final PCA reconstruction of the corrected maps. The procedure is summarized in Figure 1.

### Principal component analysis

The first step in the SWPCA algorithm was to perform a PCA decomposition of the dataset into a set of orthogonal components that model the variance present in the images, in an analogous way to the paradigm described in Active Appearance Models (AAM) [Cootes et al., 2001].



**Figure 1.**

Summary of the SWPCA algorithm, along with its context in the pipeline used in this article. Circles represent the input data, both images (green shading) and class (group and acquisition site, purple shading). Rectangles represent the different procedures applied, comprising the DARTEL normalization and registration, the different steps contained in SWPCA -PCA, ANOVA and obtaining the weighting function  $\Lambda(c)$ - and the subsequent analysis.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations  $X$  of possibly correlated variables, where  $X$  is a  $K \times N$  matrix, with  $K$  participants (in this case, with one image per participant) and  $N$  the number of voxels, into a set of  $N$  linearly uncorrelated variables called Principal Components (PC, also known as component loadings or the mixing matrix)  $W$  of size  $N \times N$  whose linear combination using a vector of component scores  $s_K$  can perfectly recompose each image. The set of these component scores  $S$  (size  $K \times N$ ) was estimated as:

$$S = XW \quad (1)$$

This transformation computes a sequence of PCs, maximally explaining the variability of the data while maintaining orthogonality between components. PCA was computed using singular value decomposition (SVD):

$$X = U\Sigma V^* \quad (2)$$

where  $U$  is an  $K \times K$  orthogonal matrix,  $\Sigma$  is an  $K \times N$  diagonal matrix with non-negative real numbers on the diagonal, and the  $N \times N$  unitary matrix  $V^*$  denotes the conjugate transpose of the  $N \times N$  unitary matrix  $V$ . With this decomposition both the component scores and estimates of the set of components loadings  $W$  were obtained. In this work the truncated form of SVD was used such that only the first  $C$  components were considered, where most of the variability of the data was concentrated:

$$S_C = U_C \Sigma_C = XW_C \quad (3)$$

where  $S_C$  is the set of component scores using the first  $C$  components (size  $K \times C$ ). To achieve reasonable performance with minimal information loss, it was assumed that

the number of components was the same as the number of images,  $C = K$ . Thus, a partial reconstruction of the original signal could be undertaken:

$$\hat{X} = S_C A_C \quad (4)$$

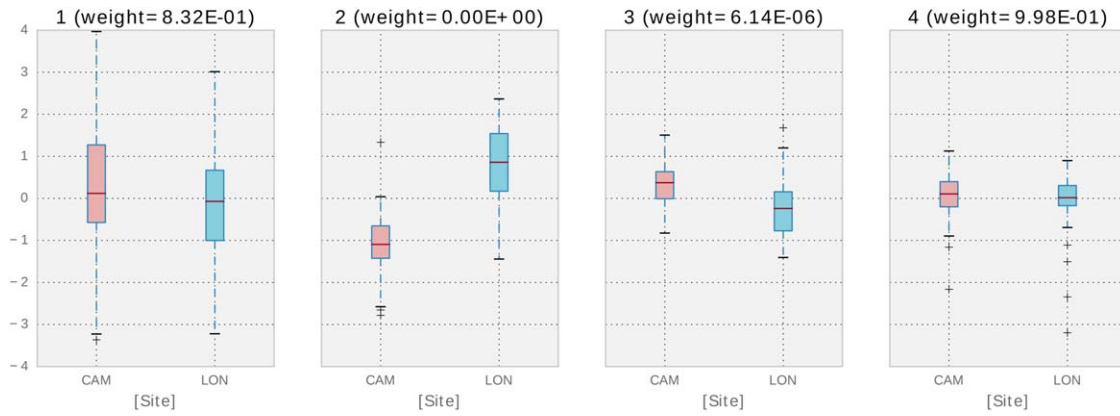
where  $A_C$  is the pseudoinverse of the truncated matrix of component loadings  $W_C$ , and  $\hat{X}$  is the reconstructed set of images.

### One-way analysis of variance

The estimated PCs effectively model the variability of the image dataset. The next step was to assess each PC as a source of inter-site variance with ANOVA. ANOVA estimates the  $F$ -statistic, defined as the ratio between the estimated variance within groups and the variance between groups:

$$F = \frac{MS_{within}}{MS_{between}} = \frac{SS_{within}/(G-1)}{SS_{between}/(K-G)} = \frac{\sum_i n_i (\bar{Y}_i - \bar{Y})^2 / (G-1)}{\sum_{ij} (Y_{ij} - \bar{Y}_i)^2 / (K-G)} \quad (5)$$

Where  $MS_{within}$  and  $MS_{between}$  are the mean squares within- and between-groups respectively,  $G$  is the number of separate groups (in our case, two),  $\bar{Y}$  is the sample mean of a certain feature (in our case, the sample mean of all  $K$  values of a given component score),  $\bar{Y}_i$  is the sample mean of the features belonging to group  $i = 1 \dots G$ ,  $Y_{ij}$  is the  $j$ th observation of a feature belonging to group  $i$  and  $n_i$  is the number of participants in the  $i$ th group. The  $F$ -distribution allows an easy computation of  $P$ -values, given the number of groups and degrees of freedom. The  $F$ -statistic and  $P$ -values were computed independently for each component



**Figure 2.**

Box-plot of the distribution of the component scores at each site in the four first components. We assume that bigger differences between distributions imply a bigger influence of the acquisition site on the portion of variance modelled by that component and therefore, to parse out those differences, the resulting weight will be smaller. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

score and acquisition site, and then used in the SWPCA algorithm.

### Weighting function

To obtain a set of corrected maps, a new signal matrix of all maps of the same modality,  $\hat{X}$ , was estimated with the influence of the PCs with variance related to acquisition site, assessed via the  $P$ -values, reduced. To do so, Eq. (4) was modified to include a square matrix  $\Lambda$  (dimension  $C \times C$ ) whose diagonal contains a weight  $\lambda_c$  for each component that depends on its  $P$ -value; that is,

$$\hat{X} = S\Lambda A \quad (6)$$

The computation of each  $\lambda_c$ , for each component, was performed using the Laplace distribution, modified so that the weights were on the interval  $[0, 1]$ :

$$\lambda_c(p_c, p_{th}) = 1 - e^{-\frac{p_c}{p_{th}}} \quad \forall p_c \in [0, 1] \quad (7)$$

where  $p_c$  is the statistical significance of the  $c$ -th component with respect to the acquisition site and  $p_{th}$  is the statistical threshold for significance; that is,  $p_{th} = 0.05$ . This weighting ensured that most of the components of variance that are not related to the acquisition site are kept unchanged, while at the same time it strongly reduces the influence of components with  $P$ -values less than the threshold.

This procedure is illustrated in Figure 2, where a box-plot of the distribution of the first four principal component scores is shown. Since we have assumed that substantial differences imply a bigger influence of the acquisition site on the portion of variance modelled by that component, the resulting weight is reduced, and the contribution of that component to the reconstructed signal will be smaller. After computing all weights, most of the

sources that are related to the acquisition site (for example, the second and third components) have been parsed out while keeping all other sources of variance.

### Experimental Settings and Validation

To validate the effects of the SWPCA algorithm on the inter-site variance, experiments were undertaken to assess the reduction of the undesired site variance in the original datasets, and its impact on the between-group signal. Two kind of analysis were performed: a characterization of voxel-wise differences, and a classification analysis.

Voxel-wise differences between groups were characterized using voxel-based-morphometry (VBM) [Ashburner and Friston, 2000], comprising preprocessing (registration, smoothing) and mass-univariate  $t$ -test on the smoothed maps from each modality. SWPCA is included (when needed) in this pipeline as a plug in, after the smoothing and before the computation of the test. Permutation testing assessed the significance of the relationship between the tested and target variables. A max-type procedure was used to obtain family-wise, whole-brain corrected  $P$ -values [Freedman and Lane, 1983]. Additionally, a component-based morphometry (CBM), based on source based morphometry (SBM) [Xu et al., 2009] was used. This procedure provided  $Z$ -maps for visual inspection comparable to those obtained in VBM, by selecting component loadings  $W$ , scaling them to unit standard deviation and weighting their contribution to the final map with their statistical significance, computed using the same permutation inference as in VBM.

A classification analysis was undertaken using a common classification pipeline [Khedher et al., 2015; López et al., 2009] consisting of preprocessing, feature extraction and classification. SWPCA is used as a plug-in here as well, after the preprocessing and before the feature

extraction step. We used PCA on the images for feature reduction and a Support Vector Classifier (SVC) with linear kernel, as implemented in LIBSVM [Chang and Lin, 2011], to classify the component scores in both corrected and uncorrected datasets (i.e. with and without SWPCA).

The classification was validated using stratified 10-fold cross-validation [Kohavi, 1995]. In brief, 9 subsets of the dataset were used for extraction of the PCs and training of the classifier with the remaining subset used for testing. This procedure was repeated for each subset, repeated 10 times to avoid possible bias and random effects of the partitions. The average and standard deviation of the accuracy (acc), sensitivity (sens) and specificity (spec) values for each repetition were recorded.

For each modality independently, the following experiments were performed:

**Experiment 1:** To demonstrate the ability of the SWPCA algorithm to reduce undesired effects due to acquisition site, the PCA + SVC pipeline was applied to the datasets labelled by acquisition site. Classification accuracy was compared to datasets with and without SWPCA. VBM was then applied to identify the spatial location of the between-site differences. This was undertaken on the whole database (ALL), and subgroups containing only ASD or CTL participants.

**Experiment 2:** The discrimination ability of each modality, acquired at different sites was assessed by classification performance of individuals from London (LON) and Cambridge (CAM) was separately assessed, using group (ASD and CTL) as the labels.

**Experiment 3:** To assess the impact of SWPCA on the datasets when characterizing the differences between ASD and CTL groups, the classification pipeline comprising PCA + SVC, as well as VBM and CBM, were applied to all participants with group as the labels.

## RESULTS

### Experiment 1: Effect of Acquisition Site

The first experiment was to demonstrate the ability of SWPCA to reduce the intensity variance related to acquisition site. To do so, we first performed a VBM analysis in all five modalities (qT1, qT2, synT1, GM and WM) separately, with the uncorrected (without applying SWPCA) and the corrected (after applying SWPCA) maps, using the acquisition site as labels.

To illustrate where the sources of variance of the acquisition sites are located, Figure 3 shows a brain  $t$ -map of significant ( $P < 0.01$ ,  $|t| > 2.57$ ) GM and WM between-site differences. The biggest reductions in variance were found in qT1 and synT1 maps, where high variability between acquisition sites, especially in the right hemisphere, was

substantially reduced after the application of SWPCA. The reduction in the qT2, GM and WM maps was smaller, although noticeable.

To quantify the impact of this variance reduction on the between-groups effects, the classification analysis was undertaken. Higher accuracy values imply that the maps contain site-related patterns that were significant, whereas accuracy close to 0.5 indicates that the site-related variance was low. The test was applied to ALL, and also to the ASD and CTL subgroups. The classification results are presented in Table 2.

Performance results indicate clear advantages of using SWPCA, in particular in the case of qT1 and synT1 which were associated with strong site-dependent variance. These results are also consistent with the reduction of significant between-group areas observed in Figure 3.

The between-site differences were smaller for GM and WM maps, possibly due their reduced sensitivity. Since fractional occupancy values are abstract, unitless values derived from each image they are less influenced by the acquisition site effects. For qT2 maps, the site-related differences were greater for the CTL participants than ASD where, according to the classification accuracy, they were nearly indistinguishable. Acquisition site differences were therefore noticeably reduced in the CTL and ALL databases, but not in the ASD.

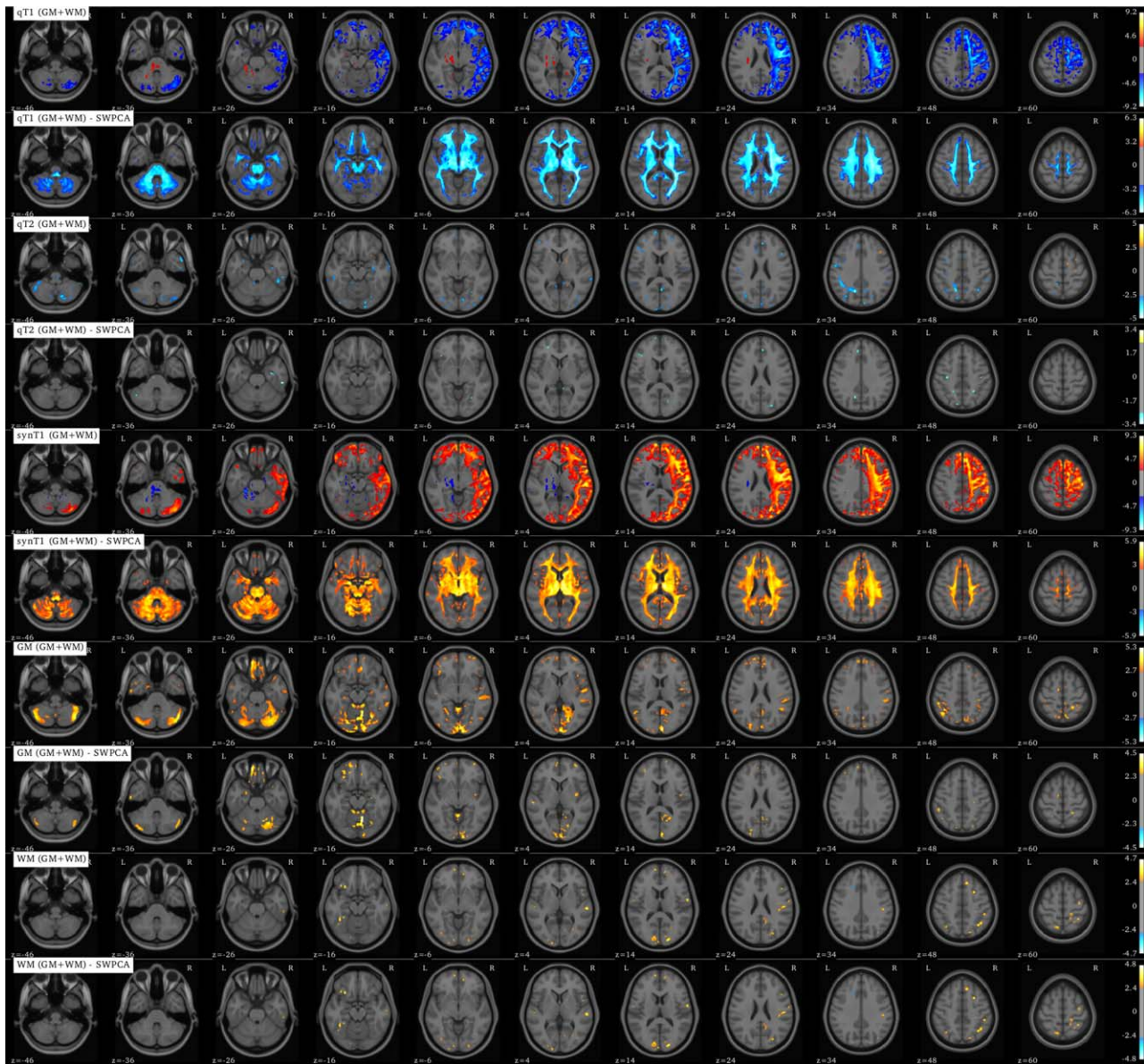
### Experiment 2: Within-Site between-Group Differences

In this second experiment, accuracy, sensitivity and specificity in the between-group comparison were recorded for images acquired from each site. This is an estimation of the discrimination ability of the different modalities without the influence of the site effects; Table 3. For all modalities, most of the values are close to a random classifier (~50%), indicative of having either no significant differences between groups, or having spatially heterogeneous patterns of sMRI measures across individuals where mass-univariate approaches are sub-optimal in detecting group differences. It is interesting to note that the London sample contained more between-group differences than those acquired in Cambridge.

### Experiment 3: Effect of SWPCA on Group Differences

Finally, group differences were characterised with and without applying site-effects reduction via SWPCA to the five modalities.

Whole-brain VBM analysis was performed on the corrected and uncorrected maps from each modality. Figure 4 depicts the brain  $t$ -maps of significant ( $P < 0.01$ ,  $|t| > 2.57$ ) qT1, qT2, synT1, GM and WM between-group differences, using ALL, with the GM + WM mask, before and after applying SWPCA, so that the reduction of site-related



**Figure 3.**

Brain t-map (voxel-based morphometry) of significant ( $P < 0.01$ ,  $|t| > 2.57$ ) GM and WM between-group differences using qT1, qT2, synT1, GM and WM modalities after applying SWPCA to remove site effects. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

variability can be observed. Some of the highlighted areas after applying SWPCA are inconsistent across modalities, with spurious peaks and noise, including a large area around the ventricles in the qT1 and synT1 modalities related to some abnormal participants that will be discussed later. However, there were some areas that were consistent across modalities. Significant areas found across at least 4 of the 5 modalities correspond to the Advanced Automated Labelling (AAL) [Tzourio-Mazoyer et al., 2002] areas of: (A) right superior frontal gyrus, Brodmann areas

6 ( $z = 60$ ); (B) the pars opercularis of the left inferior frontal gyrus, Brodmann areas 44; (C) the pars triangularis of the left inferior frontal gyrus, Brodmann areas 45; (D) the posterior part of the left middle temporal gyrus ( $z = 24$ ); CSF filled spaces on the margins of the ventricles ( $z = -6, 4, 14, 24$ ); and the left crus I of cerebellar hemisphere ( $z = -26$ ).

The complementary CBM (Section 2.4) analysis was performed on the most significant components. The resulting regions, statistically thresholded with  $|Z| > 2.57$



**TABLE 2. Between-site classification accuracy ( $\pm$  standard deviation) for different modalities and masks without and with SWPCA correction**

Modality	Mask	ALL		CTL		ASD	
		No SWPCA	SWPCA	No SWPCA	SWPCA	No SWPCA	SWPCA
qT1	GM+WM	0.875 $\pm$ 0.083	0.530 $\pm$ 0.130	0.847 $\pm$ 0.141	0.543 $\pm$ 0.115	0.769 $\pm$ 0.145	0.553 $\pm$ 0.093
	GM	0.849 $\pm$ 0.085	0.535 $\pm$ 0.107	0.835 $\pm$ 0.154	0.501 $\pm$ 0.090	0.712 $\pm$ 0.161	0.575 $\pm$ 0.084
	WM	0.865 $\pm$ 0.082	0.447 $\pm$ 0.071	0.876 $\pm$ 0.128	0.441 $\pm$ 0.058	0.813 $\pm$ 0.127	0.575 $\pm$ 0.153
qT2	GM+WM	0.596 $\pm$ 0.128	0.503 $\pm$ 0.093	0.615 $\pm$ 0.196	0.454 $\pm$ 0.075	0.506 $\pm$ 0.192	0.476 $\pm$ 0.103
	GM	0.596 $\pm$ 0.126	0.493 $\pm$ 0.097	0.549 $\pm$ 0.187	0.478 $\pm$ 0.108	0.497 $\pm$ 0.197	0.425 $\pm$ 0.091
	WM	0.612 $\pm$ 0.131	0.560 $\pm$ 0.128	0.576 $\pm$ 0.195	0.550 $\pm$ 0.146	0.541 $\pm$ 0.185	0.575 $\pm$ 0.172
synT1	GM+WM	0.904 $\pm$ 0.073	0.563 $\pm$ 0.060	0.919 $\pm$ 0.100	0.440 $\pm$ 0.057	0.807 $\pm$ 0.151	0.631 $\pm$ 0.098
	GM	0.879 $\pm$ 0.090	0.576 $\pm$ 0.035	0.899 $\pm$ 0.108	0.526 $\pm$ 0.079	0.800 $\pm$ 0.145	0.587 $\pm$ 0.042
	WM	0.904 $\pm$ 0.076	0.582 $\pm$ 0.047	0.894 $\pm$ 0.111	0.574 $\pm$ 0.038	0.859 $\pm$ 0.112	0.468 $\pm$ 0.101
GM	GM+WM	0.595 $\pm$ 0.133	0.586 $\pm$ 0.141	0.582 $\pm$ 0.192	0.566 $\pm$ 0.093	0.481 $\pm$ 0.169	0.468 $\pm$ 0.152
	GM	0.620 $\pm$ 0.141	0.585 $\pm$ 0.078	0.604 $\pm$ 0.227	0.574 $\pm$ 0.038	0.499 $\pm$ 0.188	0.525 $\pm$ 0.114
WM	GM+WM	0.659 $\pm$ 0.139	0.448 $\pm$ 0.066	0.635 $\pm$ 0.180	0.507 $\pm$ 0.144	0.522 $\pm$ 0.206	0.525 $\pm$ 0.198
	WM	0.639 $\pm$ 0.124	0.549 $\pm$ 0.072	0.578 $\pm$ 0.194	0.516 $\pm$ 0.126	0.549 $\pm$ 0.160	0.526 $\pm$ 0.136

(corresponding to  $P < 0.01$ ), were superimposed on the MNI template, and are depicted in Figure 5. A reduction of significant between-group areas after applying SWPCA is evident in most modalities, but particularly noticeable in the qT1 and qT2. In WM no significant regions were observed, neither before nor after SWPCA. The significant regions identified in any modality corresponded to the AAL areas of the CSF filled areas around the ventricles (planes  $z = -6, 4, 14, 24$ ), the right middle temporal gyrus (plane  $z = 14$ ) and the left crus I of cerebellar hemisphere (plane  $z = -26$ ). However, none of these regions were repeated over more than two of the modalities, except for the large areas around ventricles that were caused by abnormalities in three participants, which will be discussed later.

Performance results for the classification analysis applied to ALL are shown in Table 4. Between-group results were quite similar before or after applying SWPCA,

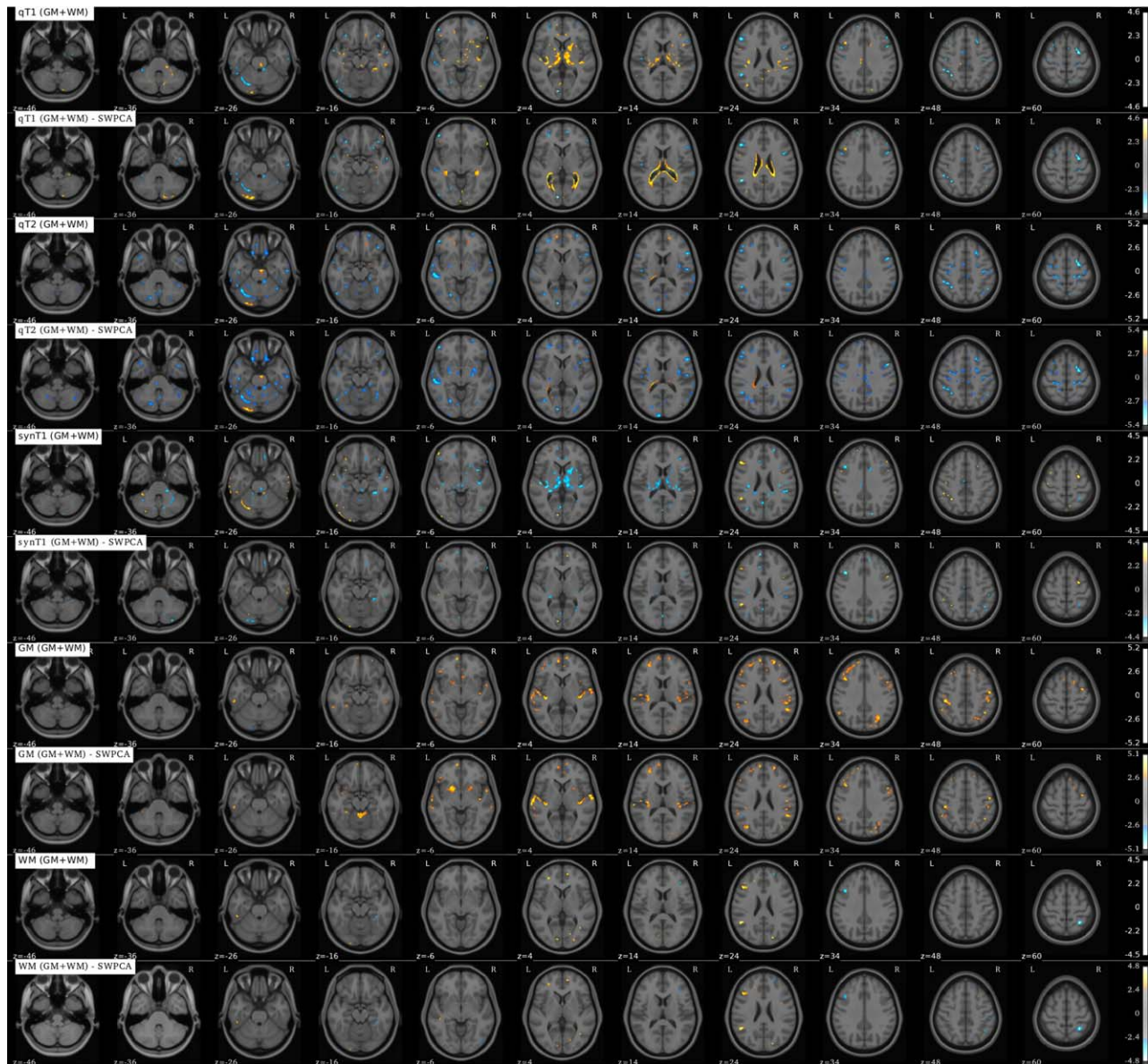
although reducing between-site variance generally reduced the performance towards a random classifier. The results in this table match the overall effects that were found in Figure 4, where most spurious significance peaks disappeared after applying SWPCA, but some regions were highlighted. These regions, where SWPCA did not seem to eliminate the significant areas but enhanced them, could be responsible for the accuracy increment in the analysis of the qT2 modality, and the GM with GM mask.

## DISCUSSION

Brain anatomical and functional differences between ASD participants and controls have been explored by a number of previous studies [Di Martino et al., 2014; Ecker et al., 2015; Hernandez et al., 2015; Lenroot and Yeung, 2013; Zürcher et al., 2015]. Many affected structures have

**TABLE 3. Classification accuracy (Acc), sensitivity (Sen) and specificity (Spec)  $\pm$  standard deviation for each modality and mask using the participants acquired at the LON and CAM sites**

Modality	Mask	LONDON			CAMBRIDGE		
		Acc	Sens	Spec	Acc	Sens	Spec
qT1	GM+WM	0.603 $\pm$ 0.175	0.512 $\pm$ 0.260	0.692 $\pm$ 0.237	0.504 $\pm$ 0.193	0.492 $\pm$ 0.276	0.515 $\pm$ 0.307
	GM	0.501 $\pm$ 0.157	0.440 $\pm$ 0.244	0.565 $\pm$ 0.245	0.484 $\pm$ 0.201	0.488 $\pm$ 0.300	0.480 $\pm$ 0.327
	WM	0.505 $\pm$ 0.174	0.485 $\pm$ 0.248	0.526 $\pm$ 0.242	0.451 $\pm$ 0.197	0.465 $\pm$ 0.297	0.435 $\pm$ 0.296
qT2	GM+WM	0.628 $\pm$ 0.168	0.535 $\pm$ 0.246	0.719 $\pm$ 0.237	0.467 $\pm$ 0.181	0.527 $\pm$ 0.307	0.417 $\pm$ 0.314
	GM	0.539 $\pm$ 0.149	0.425 $\pm$ 0.220	0.654 $\pm$ 0.222	0.491 $\pm$ 0.196	0.548 $\pm$ 0.316	0.430 $\pm$ 0.298
	WM	0.619 $\pm$ 0.194	0.585 $\pm$ 0.262	0.655 $\pm$ 0.250	0.472 $\pm$ 0.195	0.448 $\pm$ 0.283	0.492 $\pm$ 0.290
synT1	GM+WM	0.665 $\pm$ 0.158	0.578 $\pm$ 0.224	0.755 $\pm$ 0.238	0.479 $\pm$ 0.201	0.478 $\pm$ 0.318	0.475 $\pm$ 0.316
	GM	0.547 $\pm$ 0.159	0.475 $\pm$ 0.237	0.622 $\pm$ 0.252	0.514 $\pm$ 0.218	0.477 $\pm$ 0.322	0.555 $\pm$ 0.342
	WM	0.515 $\pm$ 0.185	0.520 $\pm$ 0.288	0.506 $\pm$ 0.254	0.509 $\pm$ 0.209	0.472 $\pm$ 0.317	0.542 $\pm$ 0.316
GM	GM+WM	0.513 $\pm$ 0.171	0.507 $\pm$ 0.252	0.518 $\pm$ 0.245	0.488 $\pm$ 0.202	0.445 $\pm$ 0.318	0.528 $\pm$ 0.285
	GM	0.586 $\pm$ 0.174	0.610 $\pm$ 0.247	0.564 $\pm$ 0.270	0.521 $\pm$ 0.187	0.522 $\pm$ 0.303	0.535 $\pm$ 0.289
WM	GM+WM	0.471 $\pm$ 0.181	0.455 $\pm$ 0.245	0.488 $\pm$ 0.278	0.489 $\pm$ 0.206	0.502 $\pm$ 0.319	0.483 $\pm$ 0.314
	WM	0.465 $\pm$ 0.174	0.445 $\pm$ 0.243	0.484 $\pm$ 0.268	0.468 $\pm$ 0.210	0.488 $\pm$ 0.292	0.448 $\pm$ 0.305



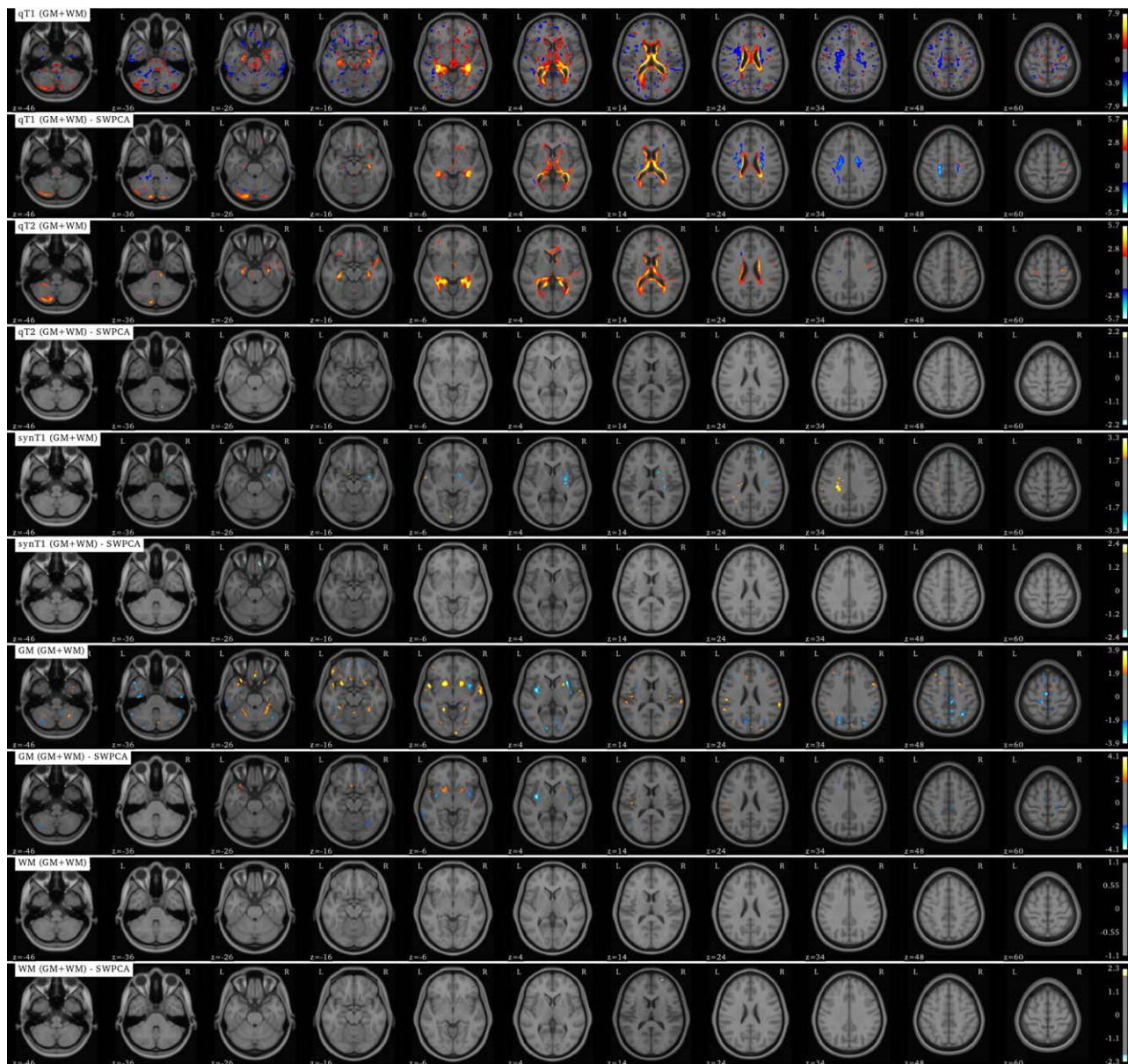
**Figure 4.**

Brain t-map (voxel-based morphometry) of significant ( $P < 0.01$ ,  $|t| > 2.57$ ) grey and white matter differences in ASD using qT1, qT2, synT1, GM and WM images before and after applying SWPCA to remove site effects. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

been proposed in each of these studies, however as a recent large-scale study points out [Haar et al., 2014], these are frequently inconsistent throughout the literature. Researchers argue that most of these structures are database-dependent, and since many studies use multi-site acquisition procedures, the variance introduced by each acquisition site is a probable source of Type I errors.

The technical and logistical drawbacks of multicentre studies are widely documented, including participant

recruitment procedures [Pearlson, 2009] and technical effects that range from the usage of different equipment or acquisition parameters [Van Horn and Toga, 2009] to physical changes that affect the performance of MRI scanners across time [Pearlson, 2009]. There is general recognition that standardization is needed to ensure the uniformity of the acquired maps. Different approaches have been used in large-scale studies, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) where



**Figure 5.**

Brain Z-map (component-based morphometry) of significant ( $P < 0.01$ ,  $|Z| > 2.57$ ) grey and white matter differences in ASD using qT1, qT2, synT1, GM and WM images before and after applying SWPCA to remove site effects. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

human “phantoms” were used to perform a preparatory optimisation of MRI scanning platforms [Friedman and Glover, 2006].

There are two major types of site effects, regardless of their source: geometric distortions and intensity inhomogeneities. In this work, we focused on the latter, since much of the geometric distortion has been eliminated during acquisition (see section “Image Database”), and the DARTEL normalization and registration acts as a homogenizing

step, reducing both between-site and between-subject geometric differences, substantially reducing the impact of the site-related differences.

Regarding intensity correction, in the MRC AIMS database used in this study [Ecker et al., 2012, 2013, 2015], a standardization procedure based on quantitative imaging [Deoni et al., 2008] was used to minimize inter-site variance and improve the signal-to-noise contrast. However, as the between-site analysis in “Experiment 1: Effect of

**TABLE 4. Classification accuracy (Acc), sensitivity (Sen), and specificity (Spec) ± standard deviation for the different modalities and masks using ALL, before and after applying SWPCA**

Modality	Mask	No SWPCA			SWPCA		
		Acc	Sens	Spec	Acc	Sens	Spec
qT1	GM+WM	0.564 ± 0.123	0.503 ± 0.179	0.625 ± 0.177	0.435 ± 0.123	0.499 ± 0.181	0.371 ± 0.178
	GM	0.523 ± 0.112	0.468 ± 0.162	0.580 ± 0.192	0.458 ± 0.120	0.477 ± 0.187	0.441 ± 0.210
	WM	0.504 ± 0.131	0.475 ± 0.191	0.533 ± 0.194	0.484 ± 0.123	0.511 ± 0.179	0.456 ± 0.194
qT2	GM+WM	0.578 ± 0.115	0.487 ± 0.208	0.669 ± 0.178	0.593 ± 0.136	0.546 ± 0.206	0.640 ± 0.194
	GM	0.554 ± 0.135	0.492 ± 0.194	0.614 ± 0.181	0.526 ± 0.144	0.512 ± 0.209	0.543 ± 0.222
	WM	0.516 ± 0.138	0.508 ± 0.198	0.522 ± 0.216	0.499 ± 0.137	0.477 ± 0.209	0.521 ± 0.196
synT1	GM+WM	0.596 ± 0.132	0.509 ± 0.194	0.680 ± 0.172	0.577 ± 0.130	0.479 ± 0.208	0.676 ± 0.183
	GM	0.587 ± 0.139	0.509 ± 0.210	0.665 ± 0.169	0.483 ± 0.136	0.489 ± 0.218	0.480 ± 0.200
	WM	0.496 ± 0.139	0.500 ± 0.189	0.492 ± 0.194	0.487 ± 0.134	0.513 ± 0.189	0.461 ± 0.211
GM	GM+WM	0.498 ± 0.120	0.486 ± 0.197	0.507 ± 0.203	0.490 ± 0.123	0.514 ± 0.197	0.465 ± 0.182
	GM	0.574 ± 0.121	0.571 ± 0.189	0.579 ± 0.163	0.593 ± 0.127	0.602 ± 0.172	0.587 ± 0.190
WM	GM+WM	0.499 ± 0.132	0.506 ± 0.189	0.487 ± 0.181	0.521 ± 0.129	0.510 ± 0.209	0.532 ± 0.180
	WM	0.506 ± 0.143	0.488 ± 0.219	0.526 ± 0.197	0.507 ± 0.122	0.521 ± 0.165	0.492 ± 0.193

Acquisition Site” section suggests, this strategy still results in variance that makes it easier to distinguish scanning sites than diagnostic groups. For example, when using qT1 the accuracy for LON vs. CAM classification was >80%, whilst when classifying ASD vs. CTL it was 52%. This marks the substantial effect of site variance on the maps’ intensity distribution, even when the multi-site study employs quantitative imaging protocol on the same model of scanner platform across sites. However, with the inclusion of GM and WM maps, we can observe that the inhomogeneities found on qT1 or synT1 barely affected the segmentation procedure.

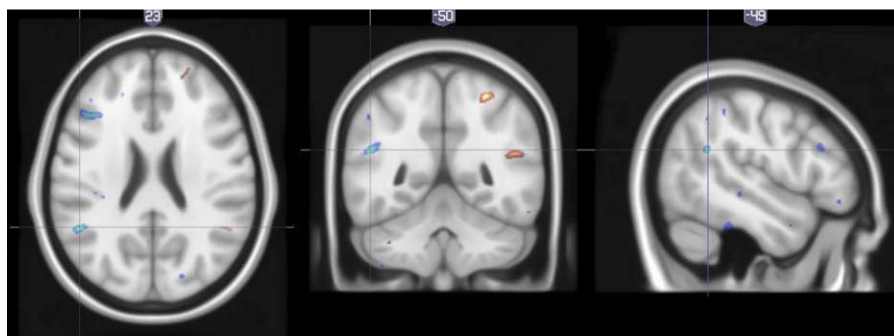
In this work, the approach we have taken is to perform a multivariate decomposition of each dataset into a number of components that explain different portions of variance. The following step was to identify the components of variance that are due to multi-site acquisition and reduce them. Decomposition was completed using PCA and then, to identify which of the components were linked to acquisition site, we performed an ANOVA on the component scores. Finally, using the weighting function defined in Sec. 2.3.3, we reconstructed the original signal reducing the undesired variance, in what we called Significance Weighted PCA (SWPCA). The method has proven its ability in reducing undesired variance, quantifiable by means of the accuracy obtained in a site vs. site classification. In this case, SWPCA reduced the accuracy from >0.8 to approximately ~0.5, a random classifier, suggesting that most site-related variance was eliminated.

A simpler approach such as applying a voxel-by-voxel ANOVA would also be useful to reduce the acquisition site effects [Suckling et al., 2012]. However, SWPCA is a multivariate approach that still offers major advantages over this voxel-wise algorithm, and similar algorithms have found utility in text document searches [Kriegel et al., 2008; Tavoli et al., 2013; Zhang and Nguyen, 2005].

First, PCA models the different sources of variance of the dataset, whereas a simple voxel-wise ANOVA only removes mean site differences, which might result in less statistical power. Secondly, SWPCA is multivariate in nature, where each component contains information that potentially affects all voxels. Together, these two features allow SWPCA to identify the components linked to the undesired effects, and reduce their impact with a weighted reconstruction approach, reducing the general variance related to the acquisition site. However, this increased power reveals a major drawback: SWPCA needs at least a moderate number of participants to work properly. That is the reason why we cannot apply SWPCA to databases such as ADNI [Friedman and Glover, 2006] or ABIDE [Di Martino et al., 2014], where the number of participants acquired at each site is small, or to the six travelling phantoms used in the calibration of the MRC AIMS study.

There exist a number of similar multivariate methods that model the influence of categorical variables, such as the well-known Partial Least Squares (PLS) algorithm [Vinzi et al., 2010] or Surrogate Variable Analysis (SVA) [Leek and Storey, 2007]. In the first case, both PLS and SWPCA take categorical variables  $Y$  along with the data  $X$  as inputs to partition the influence of these into components. However, the most significant difference is the underlying model. Whilst SWPCA estimates the principal components blindly using their variance, which is what we aim to reduce, and performs an ANOVA afterwards, PLS uses the categorical variable in the computation of the covariance matrix and then estimates the components.

On the other hand, SVA, used for gene expression studies [Leek and Storey, 2007], is more comparable to SWPCA. The SVA algorithm uses a number of decomposition and significance estimation steps to construct a set of surrogate variables; that is, variables that account for the unmodeled variance and expression heterogeneity. While



**Figure 6.**

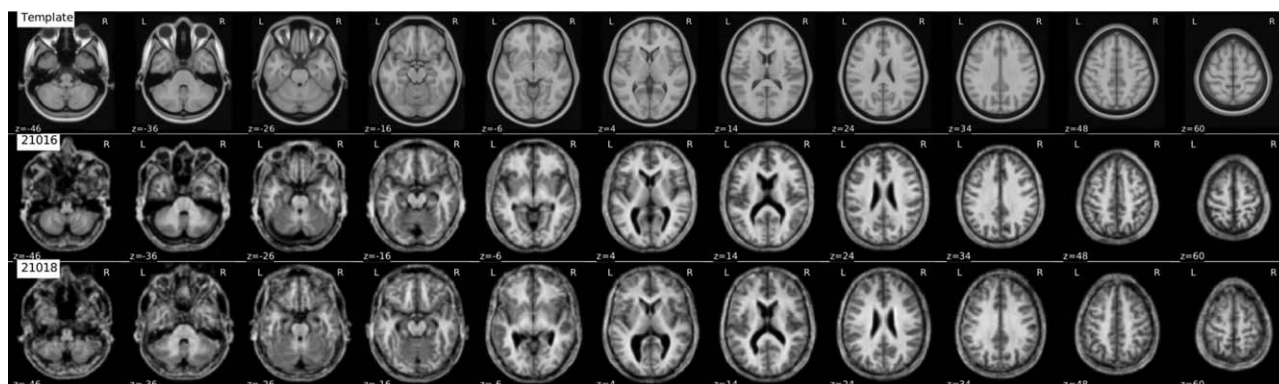
Location of the significant region that we have labelled D (posterior part of the superior temporal gyrus) within the MNI template.

similar to SWPCA in the steps used (i.e. SVD decomposition and significance estimation), their approaches are fundamentally different. SVA constructs a higher complexity model that starts by eliminating the contribution of primary variables to produce a number of unknown hidden (surrogate) variables, whereas SWPCA is intended to reduce complexity by producing variance-reduced maps to reduce the influence of previously known, but unconsidered, variables and facilitate a subsequent analysis focused only on the relevant variables.

Focusing on the VBM results, after performing the site-effects removal by SWPCA significant between-group differences were noted in five areas: (A) the right superior frontal gyrus; (B) the pars opercularis of the left inferior frontal gyrus; (C) the pars triangularis of the left inferior frontal gyrus; (D) the posterior part of the left middle temporal gyrus; and (E) the left crus I of cerebellar hemisphere. The first three regions are within Brodmann areas 6, 44 and 45. However, when examining the projection of the region D onto the MNI template (see Fig. 6), it is also located in the posterior part of the left superior temporal

gyrus. Therefore, D corresponds closely with the region between Brodmann areas 22 and 39, the Temporo-Parietal Junction (TPJ), with negative  $t$ -value at the left side (containing Wernicke's area) and positive  $t$ -value at the right side.

The role of these regions in autism has received much attention. Brodmann areas 44 and 45, that together make the Broca's Area (of importance in speech production and a proposed part of the human mirror neuron system [Nishitani et al., 2005]), is a region where mirror neuron dysfunction has been consistently reported in ASD-affected children [Dapretto et al., 2006] and adults [Hadjikhani et al., 2006; Lopez-Hurtado and Prieto, 2008; Verly et al., 2014]. Wernicke's area, contained in the left TPJ, is also linked to language, and has been associated with ASD in several works [Hadjikhani et al., 2006; Kriegel et al., 2008; Verly et al., 2014]. Additionally, the right TPJ has been proposed as related to mentalizing and has been repeatedly implicated in autism [Barnea-Goraly et al., 2004], including a fMRI study of a subsample of this same AIMS dataset [Lombardo et al., 2011]. The right superior



**Figure 7.**

The template used in this work compared to two of the participants with abnormal ventricle size (21016 and 21018). Atrophy of the cerebellum in participant 21016 can also be appreciated, responsible for some of the 'highlighted' areas in qT1, qT2 and synT1  $t$ -maps (see Fig. 4).

frontal gyrus (region A) is more equivocal, with some studies [Ecker et al., 2010, 2012] reporting abnormalities in this area, while others [Hadjikhani et al., 2006; Segovia et al., 2014] report no significant differences. Our analyses reveal no differences in the insula and amygdala, brain structures frequently linked to autism.

Some regions, particularly in qT2, synT1 and segmented GM maps show potentially spurious significance peaks around the ventricles and especially in the left crus I of cerebellar hemisphere (region E). After examining the database, two individuals had appreciable structural abnormalities in the form of abnormal ventricle size and cerebellar atrophy, as can be seen in Figure 7. It is possible that these participants influenced the computation of the *t*-maps, and therefore are responsible for the significance in region E and areas surrounding the ventricles and, since they are part of the LON subdataset, could also be responsible for the increased classification accuracy of the quantitative T1 and T2, and the synthetic T1 maps in this subdataset.

After observing the influence of these participants on the computation of the *t*-maps, we can assume that most of the structural differences in ASD are so subtle that the influence of just one or two images can impact on the final results. This, along with the poor performance of the classification pipeline presented in Section 3, dramatically reduces the significance of the aforementioned *t*-maps. Therefore, the existing evidence leads to the conclusion that ASD presents as either undetectable structural differences or, more likely, with such heterogeneous differences that are difficult to establish a common pattern even after reducing the variance introduced by acquisition site.

It may be the case that cohorts of individuals examined at different sites are somehow systematically biased towards a specific type of patient (in ways that we cannot see simply based on phenotypic information), then site-related intensity variability is also enriched with important variability about nested autism subgroups. So with any technique trying to remove the site-related inhomogeneity, the subgroup information could also be removed. Together, the evidence supports the claim that defining meaningful subgroups based on different measures, such as genetic profiling, clinical co-morbidities or sensory sensitivities, is the most urgent next step for ASD research [Haar et al., 2014].

## CONCLUSIONS

In this work, a novel method called Significance Weighted PCA (SWPCA) is proposed. The method comprises a principal component extraction, the characterization of their statistical significance according to a categorical variable (in this case, acquisition site) and the reconstruction of the images using a weighted approach. This approach was tested with an ASD database and demonstrated its ability for reducing inter-site variance, which we have

characterized with a site vs. site classification of the images. A priori, the images yielded  $>0.8$  of accuracy, in contrast to a  $\sim 0.5$  accuracy between groups, suggesting that the images contained many differences depending on acquisition site. After minimizing site-related variance, statistically significant group differences were found for example in Broca's area and the temporoparietal junction. However, their discriminative power was not sufficient to classify diagnostic groups, yielding accuracy results close to random. Our work supports recent claims that ASD is a highly heterogeneous syndrome/diagnostic category that is difficult to characterize globally using neuroimaging and therefore different (and more homogeneous) subgroups should be defined using imaging and perhaps other biological measures to obtain a deeper understanding of ASD.

## REFERENCES

- Ashburner J (2007): A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113.
- Ashburner J, Friston KJ (2000): Voxel-based morphometry—The methods. *Neuroimage* 11:805–821.
- Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, Gee JC (2010): The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 49: 2457–2466.
- Barnea-Goraly N, Kwon H, Menon V, Eliez S, Lotspeich L, Reiss AL (2004): White matter structure in autism: Preliminary evidence from diffusion tensor imaging. *Biol Psychiatry* 55: 323–326.
- Chang CC, Lin CJ (2011): LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27.
- Clementz BA, Sweeney JA, Hamm JP, Ivleva EI, Ethridge LE, Pearlson GD, Keshavan MS, Tamminga CA (2016): Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *Am J Psychiatry* 173:373–384.
- Cootes TF, Edwards GJ, Taylor CJ (2001): Active appearance models. *IEEE Trans Pattern Anal Mach Intell* 23:681–685.
- Dapretto M, Davies MS, Pfeifer JH, Scott AA, Sigman M, Bookheimer SY, Iacoboni M (2006): Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nat Neurosci* 9:28–30.
- Deoni SC, Williams SC, Jezzard P, Suckling J, Murphy DG, Jones DK (2008): Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. *Neuroimage* 40:662–671.
- Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M, Deen B, Delmonte S, Dinstein I, Ertl-Wagner B, Fair DA, Gallagher L, Kennedy DP, Keown CL, Keyser C, Lainhart JE, Lord C, Luna B, Menon V, Minshew NJ, Monk CS, Mueller S, Müller RA, Nebel MB, Nigg JT, O'Hearn K, Pelphrey KA, Peltier SJ, Rudie JD, Sunaert S, Thioux M, Tyszka JM, Uddin LQ, Verhoeven JS, Wenderoth N, Wiggins JL, Mostofsky SH, Milham MP (2014): The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol Psychiatry* 19:659–667.
- Ecker C, Murphy D (2014): Neuroimaging in autism—from basic science to translational research. *Nat Rev Neurol* 10:82–91.

- Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG, MRCAIMSC (2010): Investigating the predictive value of whole-brain structural MR scans in autism: A pattern classification approach. *Neuroimage* 49:44–56.
- Ecker C, Suckling J, Deoni SC, Lombardo MV, Bullmore ET, Baron-Cohen S, Catani M, Jezzard P, Barnes A, Bailey AJ, et al. (2012): Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: A multicenter magnetic resonance imaging study. *Arch Gen Psychiatry* 69:195–209.
- Ecker C, Ginestet C, Feng Y, Johnston P, Lombardo MV, Lai MC, Suckling J, Palaniyappan L, Daly E, Murphy CM, Williams SC, Bullmore ET, Baron-Cohen S, Brammer M, Murphy DGM, MRC AIMS Consortium (2013): Brain surface anatomy in adults with autism: The relationship between surface area, cortical thickness, and autistic symptoms. *JAMA Psychiatry* 70:59.
- Ecker C, Bookheimer SY, Murphy DGM (2015): Neuroimaging in autism spectrum disorder: Brain structure and function across the lifespan. *Lancet Neurol* 14:1121–1134.
- Freedman D, Lane D (1983): A nonstochastic interpretation of reported significance levels. *J Business Economic Stat* 1: 292–298.
- Friedman L, Glover GH (2006): Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging* 23:827–839.
- Friston, KJ, Ashburner, J, Kiebel, SJ, Nichols, TE, Penny, WD (2007): *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. London, UK: Academic Press.
- Haar S, Berman S, Behrmann M, Dinstein I (2014): Anatomical abnormalities in autism? *Cereb Cortex*:bhu242
- Hadjikhani N, Joseph RM, Snyder J, Tager-Flusberg H (2006): Anatomical differences in the mirror neuron system and social cognition network in autism. *Cereb Cortex* 16:1276–1282.
- Hernandez LM, Rudie JD, Green SA, Bookheimer S, Dapretto M (2015): Neural signatures of autism spectrum disorders: Insights into brain network dynamics. *Neuropsychopharmacology* 40:171–189.
- Jovicich J, Czanner S, Greve D, Haley E, van der Kouwe A, Gollub R, Kennedy D, Schmitt F, Brown G, Macfall J, Fischl B, Dale A (2006): Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* 30:436–443.
- Khedher L, Ramirez J, Górriz JM, Brahim A, Segovia F (2015): Early diagnosis of Alzheimer's disease based on partial least squares, principal component analysis and support vector machine using segmented MRI images. *Neurocomputing* 151:139–150.
- Kohavi R (1995): A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 14:1137–1145.
- Kriegel H-P, Kröger P, Schubert E, Zimek A. (2008): A general framework for increasing the robustness of PCA-based correlation clustering algorithms. In: *International Conference on Scientific and Statistical Database Management*. Berlin Heidelberg: Springer, pp 418–435.
- Lai MC, Lombardo MV, Chakrabarti B, Ecker C, Sadek SA, Wheelwright SJ, Murphy DGM, Suckling J, Bullmore ET, MRC AIMS Consortium, Baron-Cohen S (2012): Individual differences in brain structure underpin empathizing-systemizing cognitive styles in male adults. *Neuroimage* 61:1347–1354.
- Lai MC, Lombardo MV, Chakrabarti B, Baron-Cohen S (2013): Subgrouping the autism “spectrum”: reflections on DSM-5. *PLoS Biol* 11:e1001544.
- Leek JT, Storey JD (2007): Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genetics* 3: e161.
- Lenroot RK, Yeung PK (2013): Heterogeneity within autism spectrum disorders: What have we learned from neuroimaging studies? *Front Hum Neurosci* 7:733.
- Lombardo M, Chakrabarti B, Bullmore E, Consortium MA, Baron-Cohen S (2011): Specialization of right temporo-parietal junction for mentalizing and its relation to social impairments in autism. *Neuroimage* 56:1832–1838.
- Lopez-Hurtado E, Prieto JJ (2008): A microscopic study of language-related cortex in autism. *Am J Biochem Biotechnol* 4: 130–145.
- López M, Ramírez J, Górriz JM, Salas-Gonzalez D, Álvarez I, Segovia F, Puntonet CG (2009): Automatic tool for the Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *IET Electron Lett* 45:389–391.
- Lord C, Rutter M, Le Couteur A (1994): Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 24:659–685.
- Lord C, Risi S, Lambrecht L, Cook J, EH, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000): The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord* 30:205–223.
- Nishitani N, Schürmann M, Amunts K, Hari R (2005): Broca's region: From action to language. *Physiology* 20:60–69.
- Pearlson G (2009): Multisite collaborations and large databases in psychiatric neuroimaging: Advantages, problems, and challenges. *Schizophr Bull* 35:1–2.
- Segovia F, Holt R, Spencer M, Górriz JM, Ramírez J, Puntonet CG, Phillips C, Chura L, Baron-Cohen S, Suckling J (2014): Identifying endophenotypes of autism: A multivariate approach. *Front Comput Neurosci* 8:
- Stonnington CM, Tan G, Kloppel S, Chu C, Draganski B, Jack CR Jr, Chen K, Ashburner J, Frackowiak RS (2008): Interpreting scan data acquired from multiple scanners: A study with Alzheimer's disease. *Neuroimage* 39:1180–1185.
- Suckling J, Barnes A, Job D, Brennan D, Lymer K, Dazzan P, Marques TR, MacKay C, McKie S, Williams SR, Williams SCR, Deakin B, Lawrie S (2012): The Neuro/PsyGRID calibration experiment: Identifying sources of variance and bias in multicenter MRI studies. *Hum Brain Mapp* 33:373–386.
- Suckling J, Henty J, Ecker C, Deoni SC, Lombardo MV, Baron-Cohen S, Jezzard P, Barnes A, Chakrabarti B, Ooi C, Lai M-C, Williams SC, Murphy DGM, Bullmore E, for the MRC AIMS Consortium (2014): Are power calculations useful? A multicentre neuroimaging study. *Hum. Brain Mapp* 35: 3569–3577.
- Tavoli, R, Kozegar, E, Shojafar, M, Soleimani, H, Pooranian, Z (2013): Weighted PCA for improving Document Image Retrieval System based on keyword spotting accuracy. In: *Telecommunications and Signal Processing (TSP)*, 2013 36th International Conference on IEEE. pp. 773–777.
- Turner JA, Damaraju E, van Erp TG, Mathalon DH, Ford JM, Voyvodic J, Mueller BA, Belger A, Bustillo J, McEwen S, Potkin SG, Fbirn, Calhoun VD (2013): A multi-site resting state fMRI study on the amplitude of low frequency fluctuations in schizophrenia. *Front Neurosci* 7:137.
- Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002): Automated

- anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15:273–289.
- Van Horn JD, Toga AW (2009): Multi-site neuroimaging trials. *Curr Opin Neurol* 22:370.
- Verly M, Verhoeven J, Zink I, Mantini D, Peeters R, Deprez S, Emsell L, Boets B, Noens I, Steyaert J, Lagae L, De Cock P, Rommel N, Sunaert S (2014): Altered functional connectivity of the language network in ASD: Role of classical language areas and cerebellum. *Neuroimage Clin* 4:374–382.
- Vinzi V, Chin WW, Henseler J, Wang H, et al. (2010) *Handbook of Partial Least Squares*. Berlin Heidelberg: Springer-Verlag.
- Wang Z, Meda SA, Keshavan MS, Tamminga CA, Sweeney JA, Clementz BA, Schretlen DJ, Calhoun VD, Lui S, Pearlson GD (2015): Large-scale fusion of gray matter and resting-state functional MRI reveals common and distinct biological markers across the psychosis spectrum in the B-SNIP cohort. *Front Psychiatry* 6:174.
- Wechsler D (1999): *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.
- Xu L, Groth KM, Pearlson G, Schretlen DJ, Calhoun VD (2009): Source-based morphometry: The use of independent component analysis to identify gray matter differences with application to schizophrenia. *Hum Brain Mapp* 30:711–724.
- Zhang J, Nguyen TN (2005): A new term significance weighting approach. *J Intell Inform Syst* 24:61–85.
- Zürcher NR, Bhanot A, McDougle CJ, Hooker JM (2015): A systematic review of molecular imaging (PET and SPECT) in autism spectrum disorder: Current state and future research opportunities. *Neurosci Biobehav Rev* 52:56–73.