doi:10.1093/brain/aws363 Brain 2013: 136; 1012–1024 | 1012

BRAIN
A JOURNAL OF NEUROLOGY

A genome-wide association study of brain lesion distribution in multiple sclerosis

Pierre-Antoine Gourraud,[1] Michael Sdika,[1] Pouya Khankhanian,[1] Roland G. Henry,[1] Azadeh Beheshtian,[1] Paul M. Matthews,[2,3] Stephen L. Hauser,[1] Jorge R. Oksenberg,[1] Daniel Pelletier[1,4] and Sergio E. Baranzini[1]

1 Department of Neurology, University of California, San Francisco, CA, USA
2 GlaxoSmithKline Clinical Imaging Centre, Hammersmith Hospital, London W12 0NN, UK
3 Department of Medicine, Centre for Neurosciences, Imperial College, London WC12 0NN, UK
4 Division of Neuro-immunology, Department of Neurology, Yale School of Medicine, New Haven, CT, USA

Correspondence to: Sergio E. Baranzini, PhD,
Department of Neurology,
School of Medicine,
University of California,
San Francisco,
675 Nelson Rising Lane, Suite 215,
San Francisco,
CA 94158, USA
E-mail: sebaran@cgl.ucsf.edu

Correspondence may also be addressed to: Daniel Pelletier, MD, Division of Neuro-immunology, Department of Neurology, Yale School of Medicine, PO Box 208018, New Haven, CT 06520-8018, USA. E-mail: daniel.pelletier@yale.edu

Brain magnetic resonance imaging is widely used as a diagnostic and monitoring tool in multiple sclerosis and provides a non-invasive, sensitive and reproducible way to track the disease. Topological characteristics relating to the distribution and shape of lesions are recognized as important neuroradiological markers in the diagnosis of multiple sclerosis, although these have been much less well characterized quantitatively than have traditional measures such as T_2 hyperintense or T_1 hypointense lesion volumes. Here, we used voxel-level 3 T magnetic resonance imaging T_1-weighted scans to reconstruct the 3D topology of lesions in 284 subjects with multiple sclerosis and tested whether this is a heritable phenotype. To this end, we extracted the genotypes from a published genome-wide association study on these same individuals and searched for genetic associations with lesion load, shape and topological distribution. Lesion probability maps were created to identify frequently affected areas and to assess the overall distribution of T_1 lesions in the subject population as a whole. We then developed an original algorithm to cluster adjacent lesional voxels (cluxels) in each subject and tested whether cluxel topology was significantly associated with any single-nucleotide polymorphism in our data set. To focus on patterns of lesion distribution, we computed the first 10 principal components. Although principal component 1 correlated with lesion load, none of the remaining orthogonal components correlated with any other known variable. We then conducted genome-wide association studies on each of these and found 31 significant associations (false discovery rate <0.01) with principal component 8, which represents a mode of variation of lesion topology in the population. The majority of the loci can be linked to genes related to immune cell function and to myelin and neural growth; some (*SYK, MYT1L, TRAPPC9, SLITKR6* and *RIC3*) have been previously associated with the distribution of white matter lesions in multiple sclerosis. Finally, we used a bioinformatics approach to identify a network of 48 interacting proteins showing genetic associations ($P < 0.01$) with cluxel topology in multiple sclerosis. This network also contains proteins expressed in immune cells and is enriched in molecules expressed in the central nervous system that contribute to neural development and regeneration. Our results show how quantitative traits derived from brain magnetic resonance

Received January 4, 2012. Revised November 13, 2012. Accepted November 15, 2012. Advance Access publication February 13, 2013

images of patients with multiple sclerosis can be used as dependent variables in a genome-wide association study. With the widespread availability of powerful computing and the availability of genotyped populations, integration of imaging and genetic data sets is likely to become a mainstream tool for understanding the complex biological processes of multiple sclerosis and other brain disorders.

Keywords: voxel-wise; GWAS; multiple sclerosis

Abbreviations: GWAS = genome-wide association study; IRSPGR = inversion recovery spoiled gradient-recalled echo; SNP = single-nucleotide polymorphism

Introduction

The success of genome-wide association studies (GWAS) in identifying common variants associated with susceptibility to complex genetic disorders has fuelled their application to assess the heritability of a variety of quantitative traits. Genetic associations with secondary or intermediate phenotypes have been reported, and GWAS of human height (Lango Allen *et al.*, 2010), eye colour (Eriksson *et al.*, 2010), hair colour, freckling (Sulem *et al.*, 2008; Eriksson *et al.*, 2010), digit length ratio (Medland *et al.*, 2010), leisure time (De Moor *et al.*, 2009) and tanning (Nan *et al.*, 2009) are examples of this approach. The development of novel methods and technologies to quantify RNA, metabolite and protein concentrations in biological fluids, and electrical properties or morphological changes in specific tissues (e.g. heart, brain) have also facilitated integration of these physiological traits with genetic variation (Cheung and Spielman, 2002; Hicks *et al.*, 2009; Newton-Cheh *et al.*, 2009; Teslovich *et al.*, 2010).

Widespread availability of brain and spinal cord MRI has revolutionized the understanding and, equally important, the diagnosis and management of multiple sclerosis, the most common cause of acquired neurological dysfunction arising during early and mid-adulthood. MRI is highly sensitive in detecting white matter hyperintense/hypointense lesions (plaques) associated with multiple sclerosis neuropathology. Furthermore, anatomical location of an injury is likely to explain, at least in part, the extent and type of neurological dysfunction experienced by a patient. In a large cross-sectional study, Charil *et al*. (2003) elegantly showed an example of this paradigm by demonstrating the close relationship between site of lesions and the type of impairment in subjects with relapsing-remitting multiple sclerosis.

In an early successful integration of MRI-derived phenotypes and genetic information, Okuda *et al*. (2009) showed that patients with multiple sclerosis carrying the susceptibility allele HLA-DRB1*15:01 display a higher volume of brain lesions than non-carriers. This observation corroborated an earlier study in optic neuritis (Hauser *et al.*, 2000) and was later confirmed in an independent population (Horakova *et al.*, 2011). These findings provided evidence that visible lesions in multiple sclerosis may be, at least in part, genetically determined. In a more recent study, we correlated genome-wide genetic variation with brain glutamate assessed *in vivo* using ^{1}H magnetic resonance spectroscopy imaging (Baranzini *et al.*, 2010). Using a protein interaction network-based pathway approach, we were able to identify associations in several genes potentially affecting the function of receptors, accessory molecules, transporters and transduction signalling of glutamate.

In a recent report, voxel-level volume differences were used as the phenotype in a GWAS of 731 elderly subjects (Hibar *et al.*, 2011). Another study from the same group successfully mapped the 3D profile of temporal lobe volume differences from 742 brain MRI scans of patients with Alzheimer's disease and healthy subjects to a single-nucleotide polymorphism (SNP) in *GRIN2B*, an ionotropic glutamate receptor. More recently, an exploratory study looking at the correlation of brain lesion distribution in 208 patients with multiple sclerosis with 69 candidate SNPs suggested genotypic influences on spatial lesion distribution (Sombekke *et al.*, 2011).

Here, we tested whether genetic variation is associated with multiple sclerosis lesion topology by a GWAS. To test this hypothesis, we analysed the distribution of multiple sclerosis lesions (captured at voxel resolution of 1 mm^3) and used that measure as a trait in a GWAS. This kind of approach represents a new trend in genomics, in which more accurate and quantitative tissue phenotypes are used to maximize the power of finding biologically meaningful genetic associations.

Materials and methods

Subjects and samples

The study included 484 subjects of northern European ancestry as described previously (Baranzini *et al.*, 2009*b*). Basic demographic and clinical characteristics of the subjects are shown in Table 1. This information was obtained by means of a longitudinal, prospective, observational on-going multiple sclerosis study at the UCSF Multiple Sclerosis Center where data were acquired according to well-established uniform protocols, stored and quality checked in an integrated computerized database. Although at the time of the study most patients had a relapsing remitting course ($n = 343$, 71%), our cohort also included individuals with clinically isolated syndrome ($n = 76$, 16%), secondary progressive ($n = 45$, 9%) and primary progressive disease ($n = 20$, 4%) (McDonald *et al.*, 2001). Clinically isolated syndrome was defined as the first well-established neurological event lasting >48 h, involving optic nerve, spinal cord, brainstem or cerebellum. In patients with clinically isolated syndrome, the presence of two or more hyperintense lesions on a T_2-weighted MRI sequence was also required for enrolment into the study. Secondary progressive multiple sclerosis was defined by 6 months of worsening neurological disability not explained by clinical relapse. Primary progressive multiple sclerosis was defined both by progressive clinical worsening for >12 months

Table 1 Cohort characteristics

Variable	Value
Cohort size (*n*)	484
Age[a] (years), p50 (p25–p75)	42 (35–50)
Disease course, *n* (%)	CIS: 76 (15.7) RRMS: 343 (70.9) SPMS: 45 (9.3) PPMS: 20 (4.1)
Gender, *n* (%)	Female: 332 (68.6) Male: 152 (31.4)
HLA-DRB1*15:01 dose, *n* (%)	0:261 (53.93) 1:188 (38.84) 2:35 (7.23)
Age of onset[a] (years), p50 (p25–p75)	33 (27.0–39.5)
MSSS[a], p50 (p25–p75)	2.44 (0.91–4.33)
EDSS[a], p50 (p25–p75)	1.5 (1.0–3.0)
Disease duration[a] (years), p50 (p25–p75)	5.85 (1.68–12.9)
Lesion volume[a] (mm^3), p50 (p25–p75)	2013 (711–4340)
Patients on disease-modifying therapy, *n* (%)	292 (60)

a p50 = median; p25 = 1st quartile; p75 = 3rd quartile.
CIS = clinically isolated syndrome; EDSS = expanded disability status scale; MSSS = multiple sclerosis severity score; PPMS = primary progressive multiple sclerosis; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis.

from symptom onset without any relapses, and abnormal CSF as defined by the presence of two or more oligoclonal bands or an elevated immunoglobulin G index. Individuals were excluded if they had experienced a clinical relapse or received treatment with glucocorticosteroids within the previous month of enrolment, as these conditions would have posed significant confounders to our study. The concomitant use of disease-modifying therapies for multiple sclerosis was permitted for inclusion in this study. For all subjects, the expanded disability status scale and multiple sclerosis severity score scores were assessed, and baseline brain MRI scans were performed within 2 weeks of entry into the study. Age of onset was defined as the first episode of focal neurological dysfunction suggestive of CNS demyelinating disease. To reduce the heterogeneity of our cohort, only patients with clinically isolated syndrome and relapsing-remitting multiple sclerosis who had a disease duration of <10 years and age of onset >20 years ($n = 284$) were studied further. The University of California, San Francisco institutional review board approved the study, and informed consent was obtained from all subjects before participation.

Imaging data

All brain MRI data were derived from high-resolution images acquired on a single 3 T GE Excite scanner (GE Healthcare Technologies) equipped with a phase-array eight-channel coil using a 3D T_1-weighted inversion recovery spoiled gradient-recalled echo (IRSPGR) sequence yielding a 1-mm^3 isometric voxel size (echo time/repetition time/inversion time = 2/7/400 ms, flip angle = 15°, number of excitations = 1, 180 slices).

Hemispheric, brainstem and cerebellar white matter lesions were segmented out directly from the high-resolution T_1-weighted 3D-IRSPGR images based on a semi-automated pixel intensity threshold with manual editing, using in-house software, and T_1 lesion masks were created as reported previously (Blum *et al.*, 2002; Mowry *et al.*, 2009).

Lesion probability maps

To create lesion probability maps, lesions were first segmented on each multiple sclerosis subject 3D-IRSPGR scan, inpainted (Sdikaand Pelletier, 2009) and non-linearly registered (Sdika, 2008) using a healthy subject (female, 42 years old) scan as reference. The lesional map of each subject was then mapped onto the reference scan using the transformation found by the previous step. The registered lesion maps were then analysed on a voxel-by-voxel basis and voxels with *P*-lesion >0 were considered lesional.

Identification of clusters of lesional voxels

With the goal of accurately representing the 3D shapes of lesions in each subject, lesional voxels that shared a common face (6), edge (12) or vertex (8) in 3D space (26 in total) were grouped into clusters. To accomplish this, we developed an algorithm that starts by considering any lesional voxel and evaluates whether each of its 26 neighbours in 3D is also a lesional voxel, and if so, it merges them. The algorithm continues to aggregate adjacent lesional voxels until the resulting cluster is only surrounded by voxels devoid of lesions. We use the term 'cluxel' to describe each such cluster of lesional voxels. For each subject, cluxel topology was defined as the 3D shape of a lesion resulting from the aggregation of neighbouring lesional voxels into cluxels. Analysis of neighbouring voxels was performed in the R statistical package using a custom algorithm (code available on request).

Genetic data

All 484 subjects with multiple sclerosis were typed for 553 139 SNPs using the Illumina 550k platform with $<2\%$ of genotyping failure rate per sample (Baranzini *et al.*, 2009*b*). SNPs (7043) were removed from the data set for missing genotypes in $>2.5\%$ of individuals and 50 340 SNPs were removed for having a minor allele frequency of $<5\%$. Finally, 443 SNPs that departed from Hardy–Weinberg equilibrium ($P < 0.001$) were removed. After these quality control steps, 495 313 (99.9%) markers remained. However, to limit the number of association tests, we also eliminated SNPs that were in moderate linkage disequilibrium ($r^2 > 0.5$), and only 208 975 were considered for further analysis. To control for multiple hypothesis testing, the false discovery rate (FDR) correction (Benjamini and Hochberg, 1995) was applied to the results of each GWAS. An analysis with all SNPs that passed quality control was also performed, and results are available in Supplementary Table 3. Clusterplots of significant SNPs are provided in Supplementary Fig. 3 and regional association plots in Supplementary Fig. 4. Distributions of the phenotype with each significant genotype are provided in Supplementary Fig. 5.

Statistical analysis

Of the 484 subjects with available genotypes, 284 fitted our stringent criteria for analysis. GWAS analyses were completed in this subset using PLINK (version 1.06) (Purcell *et al.*, 2007), and all other tests were performed in the R statistical package (version 2.9). Binary traits were analysed using logistic regression. Quantitative variables (T1LL and T2LL) that did not distribute normally were log transformed before analysis using a linear model. To reduce the heterogeneity of our cohort, we used a combination of subject stratification and covariates in the linear model. We first stratified subjects by disease type (we included only clinically isolated syndrome and relapsing-remitting

multiple sclerosis) and then adjusted by disease duration. We tested other variables (gender, disease duration, age, age of onset, expanded disability status scale, immunotherapy, etc.) and found no significant contribution in the measured phenotype (Supplementary Fig. 2). Given the negligible influence of these variables on the measured outcome, we decided not to adjust for those, to preserve power. Reported *P*-values were adjusted by the FDR method (Benjamini and Hochberg, 1995).

We performed a simulation to assess the probability of observing a *P*-value < 1 observed in the data, given an effect size equal to our estimated effect and variance equal to the sample variance of principal component 8 (PC8). In other words, this test shows our ability to find SNPs with such effect sizes given the number of samples and SNP allele frequencies found in our study. Specifically, we simulated the genotypes of 284 individuals to match the maximum autocorrelation factor of each of our hit SNPs. We then simulated a continuous phenotype based on the hit SNP's effect size and added noise from a normal distribution with mean 0 and variance equal to that of PC8. Finally, we tested for associations between the simulated genotypes and phenotypes. This process was repeated 100 000 times and the distribution of obtained *P*-values is shown in Supplementary Fig. 6.

Principal component analysis

To reduce the dimensionality of the voxel-based analysis, we used principal component analysis. This method relies on orthogonal transformation to convert a set of possibly correlated voxel-based measures into a set of values of linearly uncorrelated variables, the principal components. The first principal component explains the largest amount of variance (i.e. accounts for as much of the variability in the data as possible), and each subsequent component explains the highest remaining variance under the constraint that it is orthogonal (i.e. uncorrelated with) to the preceding components.

Module (sub-network) analysis

Systematic module searches on a highly curated protein interaction network were conducted as described (Baranzini *et al.*, 2010). Briefly, each gene product in the network was assigned a number corresponding to the *P*-value of the most strongly associated SNP for that gene with the trait (only *P*-values < 0.05 were considered). Next, the Cytoscape (www.cytoscape.org) plugin jActive modules (Ideker *et al.*, 2002) were used to identify groups of interacting gene products that were also associated with PC8 by GWAS. jActive modules convert *P*-values into *z*-scores and use a greedy algorithm to grow a sub-network (or module) from a random seed node by sequentially incorporating its neighbours in the protein interaction network. The algorithm then returns the smallest possible module that includes gene products with the most significant associations. A significance (*z*) score is assigned to each reported sub-network after evaluation of 10 000 random networks of similar size. Only modules with a score > 3 and of size > 5 were considered.

Results

Lesion probability maps

We analysed the presence and distribution of white matter lesions in 484 patients with multiple sclerosis at 1 mm^3 voxel resolution. First, the probability of a lesion to be present in each of the 67 000 voxels was calculated across all patients to build a composite lesion probability map (Fig. 1A). As expected, the lesion probability map clearly shows that the probability of a lesion is maximal near the periventricular areas and decays markedly in other brain areas. Following our earlier observation that subjects carrying the HLA-DRB1*15:01 allele consistently show higher lesion volume than those carrying other alleles (Okuda *et al.*, 2009), we next evaluated the lesion probability map stratified by the presence of this major susceptibility allele. We observed a moderate but significantly different distribution of lesional voxels in the 15:01 negative (Fig. 1B) versus the 15:01 positive group (Fig. 1C) ($P = 2.3 \times 10^{-4}$, matched Wilcoxon test). Interestingly, we also observed that more voxels with high lesion probability were found in males (Fig. 1E) compared with females (Fig. 1F) ($P = 2.2 \times 10^{-16}$, matched Wilcoxon test). Although this finding could suggest a gender-dependent effect on the total number of lesional voxels and/or their spatial distribution, it is also possible that the observed differences are caused by a larger white matter volume typical in males.

Distribution of lesional voxels

As shown in the lesion probability map, the probability for any given voxel to contain a lesion across all patients is exceedingly small (Fig. 2). Similarly, in any given subject, only a small fraction of voxels carry a lesion (lesional voxel). This limited variance in the distribution of the binary trait (lesion versus no lesion) prevented us from conducting a voxel-based genetic association because most voxels contained no lesions (uninformative). We thus considered an alternative strategy in which a combination of a threshold (T) number of lesional voxels (depicted by coloured lines in Fig. 3) and the *n*-most lesional voxels across the population (*x* axis, Fig. 3) defines two subgroups of individuals for analysis. For example, the proportion of patients displaying lesions in at least 13 of the 100 most lesional voxels (T) is ~50%, thus maximizing the power of finding a genetic association to this trait. Because multiple combinations can result in balanced proportion of observations, we tested this stratification approach over three combinations of T and Z parameters (indicated by the arrows in Fig. 3).

Genome-wide association study on lesion distribution

We performed a GWAS for each of three combinations of parameters resulting in a balanced proportion of patients (low, mid and high number of lesional voxels, denoted by arrows in Fig. 3). Although several genetic markers were nominally significant in these analyses, none surpassed the FDR correction for multiple comparisons.

As lesional voxels can adopt a variety of topological arrangements and these may be specific to each subject, we developed an algorithm that allowed merging of neighbouring lesional voxels into clusters (cluxels) to identify and characterize these spatial patterns. Using this strategy, the total number of quantitative variables for GWAS (number of cluxels per patient) is reduced by several orders of magnitude relative to a direct, voxel-based

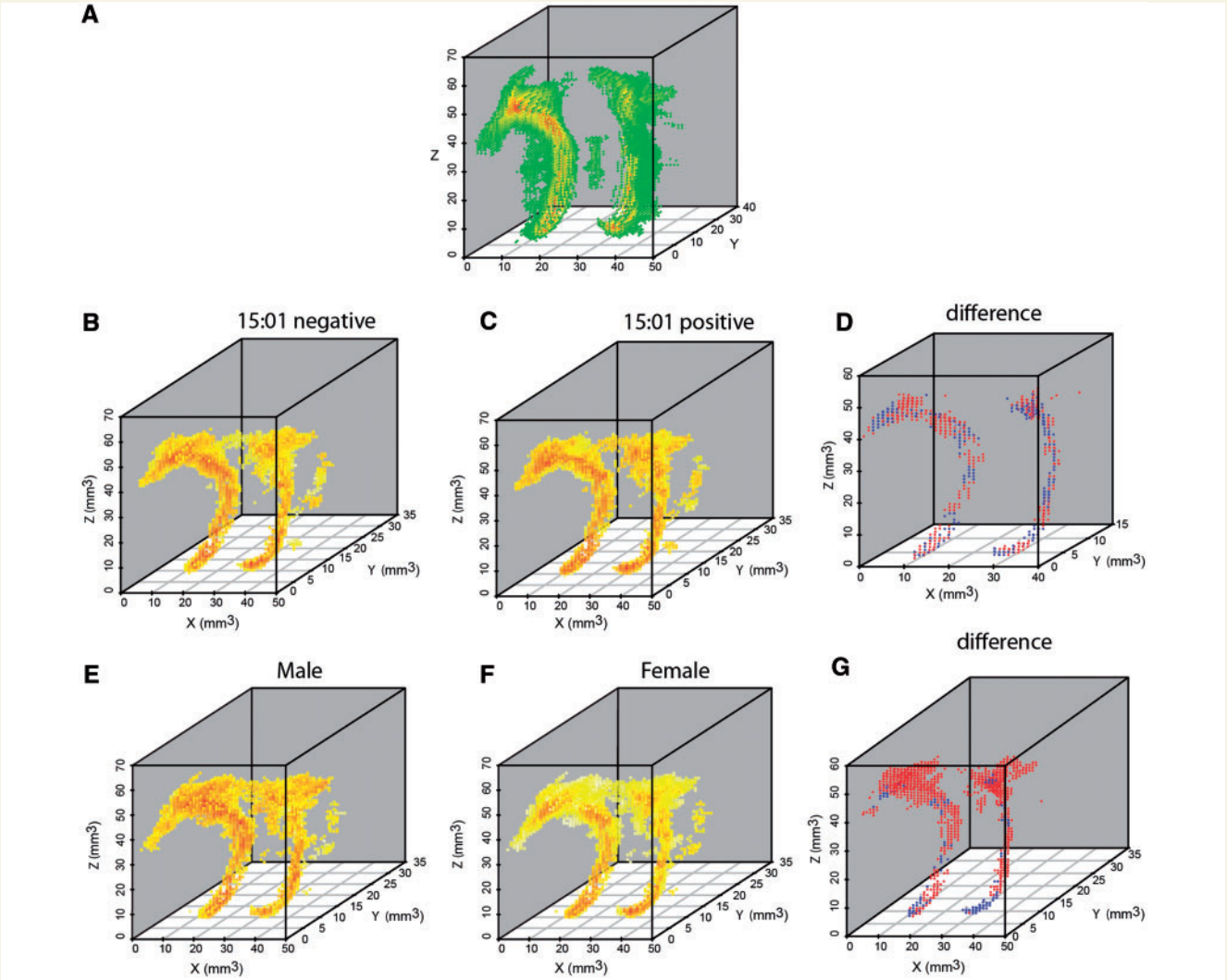

Figure 1 Lesion probability maps. (**A**) Lesion probability map over all individuals shows the highest probability of a lesion to be around the periventricular areas. Individuals carrying at least one 15:01 allele (**B**) display a significantly different lesion distribution than 15:01 negative (**C**) individuals. The colour of each voxel is proportional to the probability of there being a lesion. (red = high, yellow = medium, green = low). (**D**) The difference between panels **A** and **B**. In this lesion probability map, red voxels indicate a higher probability of a lesion for 15:01 positive subjects, whereas blue voxels indicate lower probability of a lesion for these subjects. Males (**E**) have different distribution of probabilities than females (**F**). (**G**) The difference between panels **E** and **F**. In this lesion probability map, red voxels indicate higher probability of a lesion for males, whereas blue voxels indicate lower probability of a lesion for males.

analysis and can also be used as a dichotomous trait (e.g. dense versus sparse organization) (Fig. 4). In addition, this new trait is independent of total lesion volume, as two patients with similar number of lesional voxels can have a markedly different cluxel topology, and the correlation between the number of cluxels and lesion volume is low (Fig. 4C). We computed the distribution of the number of cluxels over all subjects and chose an arbitrary cut-off ($n = 35$) to define a dichotomous phenotype (high/low number of cluxels) for a GWAS. (Fig. 4D). The pairwise minimum distance between cluxel edges and the average size of cluxels were also found to be highly variable across individuals and independent of gender and lesion volume. However, a GWAS on the number of cluxels and on the average minimal distance between cluxels did not identify markers with genome-wide (FDR-corrected) significance, and a GWAS on the average size of cluxels resulted in borderline significant results (Supplementary Tables 1 and 2).

Although the arrangement of cluxels may be markedly different across individuals, just measuring their number, size or distance from each other may not accurately capture actual cluxel topology. We then hypothesized that if the variance related to lesion volume is accounted for, the remaining variance would be enriched in cluxel topology. We thus used principal component analysis to decompose the total variance in lesional voxels and computed the correlation of each component with other relevant imaging and clinical variables (e.g. total lesion volume, expanded disability status scale, age of onset, T_1 lesion load, T_2 lesion load, total number of lesional voxels, age, disease duration, gender,

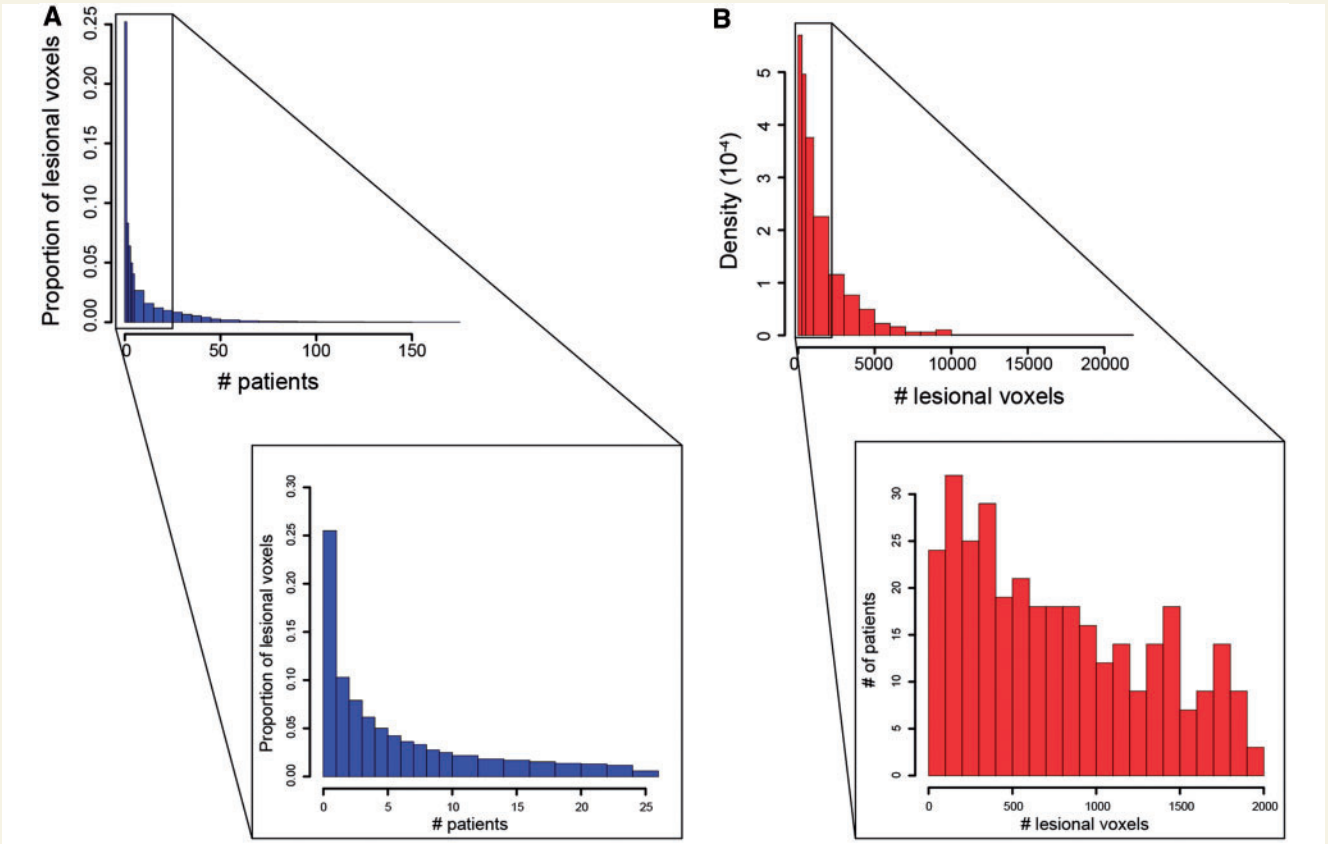

Figure 2 Lesional voxels are rare. (**A**) The proportion of voxels that contain a lesion in the given number of patients is exponentially decaying. The inset shows the distribution of voxels that contain lesions in 0–25 patients. (**B**) The proportion of patients with the given number of lesional voxels is also exponentially decaying. The inset shows the number of subjects with lesions in the top 2000 lesional voxels.

HLA-DRB1*15:01, treatments and multiple sclerosis type). Almost 10% of the total variance was explained by the first component (PC1), which in turn was highly correlated with lesion volume ($R^2 = 88\%$) (Fig. 5A and B). Given the wide distribution of lesions observed in patients with multiple sclerosis, this finding is not surprising. However, it allowed us to identify the proportion of variance explained by lesion volume itself and then focus on the remaining (orthogonal) variance, in which lesion topology (and not load) is now more represented.

We then performed GWAS with each of the remaining components (PC2–PC10) to evaluate genetic associations with orthogonal sources of variance in lesion topology within the multiple sclerosis brain. To control the rate of false discoveries, the location of lesional voxels was permuted across all voxels 100 times, thus keeping the lesional load of each individual intact while removing any spatial structure of lesions across the brain. An enrichment of significant *P*-values was observed in the original data when compared with the permuted sets (Supplementary Fig. 8). Specifically, 31 significant associations at FDR-corrected *P*-value of 0.01 were identified with PC8 (Fig. 5 and Table 2), suggesting a genetic effect in the variance explained by this component. Although 13 of these associations reach $P < 10^{-7}$, an average of only 0.64 associations exceeded this threshold in the permuted data set tested with on PC8, and an average of 2.32 associations exceeded this threshold in the permuted data sets with all 10 principal components (Supplementary Table 4). Although the voxels with the highest weight in PC8 are not arranged in a pattern that matches any particular brain structure, the associations showed to be robust to rigorous quality controls (Fig. 5D). None of the remaining principal components yielded any significant associations.

To evaluate the functional relationship among statistically significant associations, we performed a protein interaction network-based pathway analysis. This approach (as described by Baranzini *et al.*, 2009*a*) integrates information of genetic associations with protein interactions to identify groups of physically interacting proteins (sub-networks or modules) that likely work together in a biological pathway. The top-ranked network identified using this approach contained 48 genes (Fig. 6A), most of them expressed in inflammatory cells or in the CNS (Supplementary Fig. 1), which is encouraging given the suspected involvement of inflammatory and neural pathways in multiple sclerosis pathogenesis. A gene ontology analysis of the elements of this network revealed a significant enrichment in genes associated with immune response and with CNS development and function (Fig. 6B). In contrast, the top-scoring network obtained after permutation of *P*-values among all elements in the protein interaction network did not yield any significant enrichments.

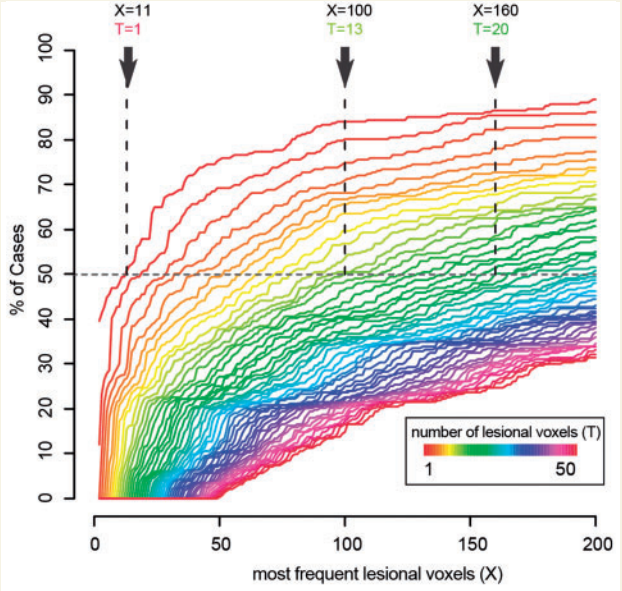

Figure 3 Proportion of subjects with lesions in top lesional voxels. Each curve shows the proportion of subjects with $\geqslant$T lesional voxels versus the number of voxels considered (where the voxels considered are the 200 voxels that contain the most lesions across patients). The colour of each curve represents 'T', the threshold minimum number of lesions among the top lesional voxels. The dotted horizontal line indicates combinations of T lesions in top $\times$ lesional voxels resulting in a more or less balanced 'cases'–'controls' design for a GWAS. The arrows depict the three balanced combinations chosen to run GWAS. For example, the first arrow indicates that 50% of cases have at least one lesion in any of the top 11 lesional voxels. The second arrow indicates that 50% of the cases have at least 13 lesions in the top 100 lesional voxels. Similarly, the third arrow indicates that 50% of the patients have at least 20 lesions in the top 160 lesional voxels. It can also be seen from this chart that ~80% of cases have at least one lesion (T = 1, red line) in the top 100 lesional voxels and ~90% in the top 200 lesional voxels.

Discussion

Although genetic association studies have scored notable achievements in the discovery of susceptibility genes in complex diseases, this search has proven more challenging than initially anticipated. The clinical and genetic heterogeneity embedded in these diseases hampers appropriate phenotyping and explains, at least in part, this incomplete success (van der Sluis *et al.*, 2010; Kutalik *et al.*, 2011; Wood *et al.*, 2011). In particular, the difficulty in defining a clear phenotype conspires against the basic assumptions of GWAS, which are aimed at identifying excess of sharing in common genetic variants among individuals with the same high-level phenotype (e.g. health or disease). In this regard, the identification of intermediate, more precise and quantitative phenotypes that are more proximal (than a physiological state of health or disease) to the genotype in the chain of events determining causality, is potentially a more informative strategy. Examples of this approach include studies that integrate data on volumetric differences from patients with Alzheimer's disease at the voxel level with genotypes from an existing GWAS to identify genetic associations involved in brain structure (Biffi *et al.*, 2010; Shen *et al.*, 2010; Stein *et al.*, 2010). Similar approaches have been used to find genetic associations with cognition and susceptibility to schizophrenia (Potkin *et al.*, 2009*a*, *b*).

In multiple sclerosis, the relationship between physical or cognitive impairment and whole-brain white matter lesion volume is generally weak. However, despite the fact that anatomical location of an injury is likely to explain (at least in part) the extent and type of neurological dysfunction, lesion volume rather than location is used as the prevalent metric. In a large cross-sectional study using lesion probability maps, Charil *et al*. (2003) showed a relationship between sites of lesions and major impairments in a group of subjects with relapsing-remitting multiple sclerosis. In this study, we have extended the concept of using lesion distribution as a phenotype to investigate the contribution of genetic variation to the topological distribution of white matter lesions in multiple sclerosis at voxel-level resolution.

To be able to perform genetic associations with white matter lesion distribution across a population, it is critical to place all subject lesion masks in a common reference space (Sdika and Pelletier, 2009). In this study, we used image registration, an automated method to perform this task. Given two brain images, registration finds the geometric transformation that maps one brain to the other (Miller *et al.*, 1993; Rueckert *et al.*, 2006; Sdika, 2008). Although non-rigid registration algorithms are able to cope with the high inter-subject variability of human brain, this method is sensitive to the presence of lesions, thus most investigators prefer to use affine registration when building multiple sclerosis lesion probability maps (Narayanan *et al.*, 1997; Charil *et al.*, 2003; Enzinger *et al.*, 2006). Indeed, affine registration is robust to the presence of lesions in the brain, but only the global shape of the brain will fit the reference and the inter-subject variability is not taken into account. To further reduce the influence of the lesion in this process, lesional voxels can be masked out of the cost function of the registration. This method improves the registration of focal lesions with respect to affine or non-rigid registration with no cost function masking (Brett *et al.*, 2001). We recently proposed a method dedicated to the problem of lesion mapping and registration of images of brains of subjects with multiple sclerosis (Sdika and Pelletier, 2009). In this method, all white matter lesions are inpainted using an original algorithm before performing the registration. The intensity of the voxels in the lesions is replaced by the intensity of surrounding white matter to visually remove them from the original image. Any registration method can then be used to register the inpainted image on another image. We have shown that the registration is improved with respect to cost function masking, especially in lesional areas.

We decided to use white matter hypointensity lesions detected on high-resolution T_1-weighted 3D-IRSPGR images as the concordance between lesions visible on T_2 spin-echo weighted and T_1-weighted 3D-IRSPGR images is ~100% (Henry *et al.*, 2009). These images allowed us to delineate sharp lesion edges, making determination of lesion topology more precise and enabling integration with readily available GWAS information from the same individuals. Results from this approach show significant

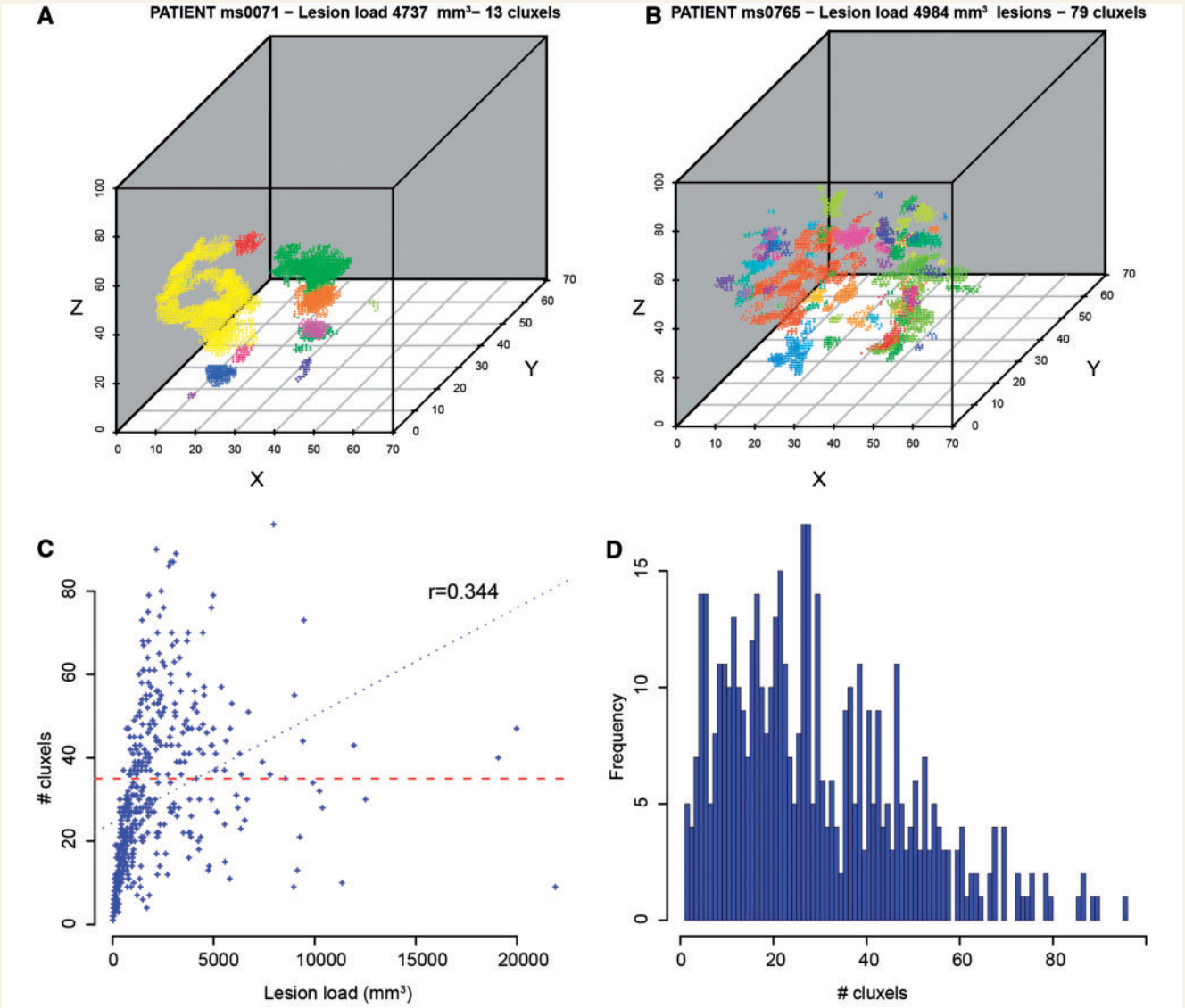

Figure 4 Cluxel topology as a trait. 3D representation of neighbouring lesional voxels (cluxels) in two patients with similar lesion volume. (**A**) The patient depicted on the left has 4737 lesional voxels arranged in 13 cluxels. (**B**) Despite having a similar lesion volume (4984 mm^3), the lesional voxels in the patient depicted on the right are arranged in 79 cluxels. These patient-specific differences form the basis for a newly defined trait in multiple sclerosis. (**C**) The black dotted line shows that the correlation between lesion volume and number of cluxels across all multiple sclerosis subjects is low (R = 0.344). (**D**) The frequency distribution of the number of cluxels per subject.

associations with lesion distribution that are independent of whole-brain lesion volume. A clear limitation of our approach is the lack of an independent replication, which typically constitutes the most reliable evidence of reported genetic associations. However, to replicate this study, a similarly sized (or larger) cohort of subjects with both genome-wide genotypes and high-resolution structural imaging are needed, a data set that currently very few groups in the world are ready to obtain. One possible source of genotyped subjects would be those participating in large GWAS. Unfortunately, these are typically multicentre efforts, and although this is acceptable for DNA collection and analysis, comparing images obtained at different sites is technically challenging and to date there is no consensus as to how to address this problem effectively. Also, although these results are suggestive of a genotypic influence on spatial distribution of lesions in multiple sclerosis, this study may be underpowered to detect unequivocal signals. With a widespread interest in quantitative imaging and broader availability of genomic data, we believe this type of study will become more common and powerful in the near future.

In our genetic study of lesion distribution (i.e. PC8), 31 associations fell below a FDR threshold of 1%. A caveat worth noticing is that most of the associated SNPs were relatively rare in our population (~5%). Although the clusterplots appeared robust (Supplementary Fig. 3), the regional association plots (Supplementary Fig. 4) uncovered several SNPs in relatively high linkage disequilibrium with the index marker that did not show evidence of association. It is possible that this observation is related to these low-frequency alleles, but further studies will be

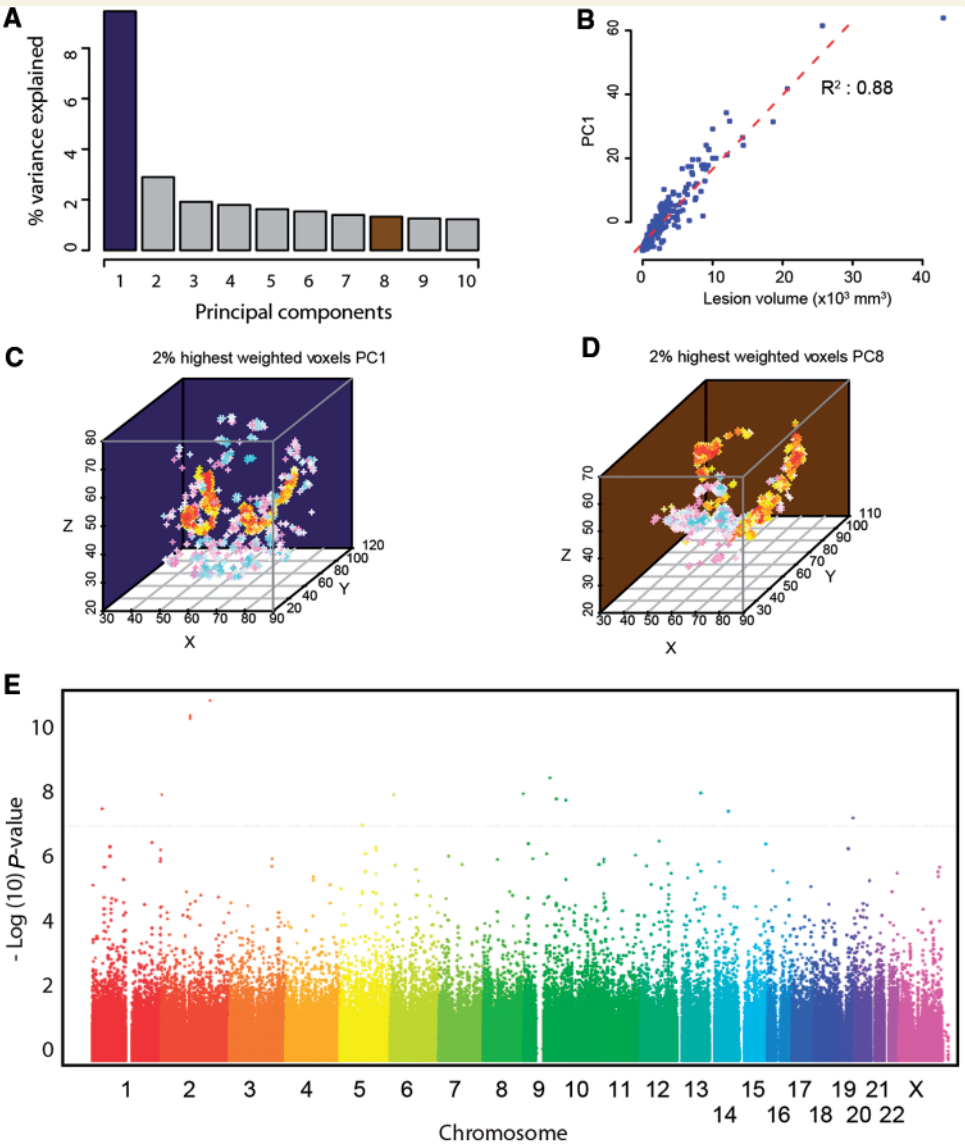

Figure 5 Principal component analysis on the distribution of lesions across all subjects. (**A**) Amount of variance explained by each of the top 10 principal components. PC1 (PC1, blue bar) explains almost 10% of the variance and is highly correlated with lesion volume (**B**). (**C**) Spatial distribution of voxels with the highest weight in PC1. Each displayed voxel is coloured according to its weight. A 'hot' colouring scale (red-orange-yellow) is used for the 2% highest-weighted voxels, and a 'cold' colouring scale (purple-blue-cyan) was used for the 2% lowest-weighted voxels. (**D**) Spatial distribution of voxels with the highest weight in PC8, which showed genome-wide (FDR-corrected) association with 31 markers. (**E**) Manhattan plot showing associations with PC8.

needed to unequivocally establish these associations. Intriguingly, when analysed altogether in a biological context, these associations are consistent with biological processes involved in brain development and lesion formation or repair (i.e. myelination), thus supporting our original hypothesis of genetic associations with lesion topology. For example, one of the most relevant associations (rs10119179, $P = 3.65 \times 10^{-9}$) maps within *SYK* (spleen tyrosine kinase) a kinase involved in tau phosphorylation (Lebouvier *et al.*, 2009) and amyloid-β oligomer-mediated microglial activation (Sondag *et al.*, 2009). The product of *SYK* has been shown to phosphorylate myelin basic protein (Shimomura *et al.*, 1993) and α-synuclein thereby preventing its aggregation, thus potentially playing an anti-neurodegenerative role (Negro *et al.*, 2002). Interestingly, DNA methylation in SYK is dynamically regulated in the human cerebral cortex throughout the lifespan, involves differentiated neurons and affects a substantial portion of genes predominantly by an age-related increase (Siegmund *et al.*, 2007). T_1 lesions non-specifically define inflammation/oedema, myelin loss and matrix destruction (a consequence both of severity of inflammation and lack of axonal regrowth). Although traditional approaches consider inflammation and tissue response as independent processes, we and others hypothesize that—in fact—they are interactive (Waxman, 2005; Hauser and Oksenberg, 2006). Our findings support this notion by showing joint

Table 2 GWAS with PC8

Chromosome	Position	SNP	Mapped gene (distance)	Beta	MAF (%)	Raw *P*-value	FDR_BH
2	174014782	rs16861690	CDCA7 (+100 Kb)	2.203	9.54	1.80E−11	3.77E-06
2	103124572	rs13410351	TMEM182 (+400 Kb)	2.008	11.3	6.12E−11	6.39E-06
9	**92672620**	**rs10119179**	**SYK (0 Kb)**	2.183	7.12	**3.65E−09**	**0.000254**
13	85343994	rs9602859	SLITRK6 (−80 Kb)	1.755	11.7	1.04E−08	0.000341
8	**141380846**	**rs6983731**	**TRAPPC9 (0 Kb)**	1.705	9.54	**1.09E−08**	**0.000341**
6	12046782	rs169715	HIVEP1 (−70 Kb)	1.891	9.54	1.14E−08	0.000341
2	2728526	rs2053906	MYT1L (−400 Kb)	1.778	9.9	1.14E−08	0.000341
9	114836373	rs10981613	ZFP37 (10 Kb)	1.712	12.01	1.56E−08	0.00039
10	8177388	rs12776126	GATA3 (20 Kb)	1.728	11.66	1.68E−08	0.00039
1	**36700319**	**rs7065**	**MRPS15[a] (0 Kb)**	1.574	10.24	**3.06E−08**	**0.00064**
14	**68189895**	**rs10483818**	**RAD51L1 (0 Kb)**	1.698	9.9	**3.62E−08**	**0.000689**
19	**61012011**	**rs16986626**	**NLRP11 (0 Kb)**	1.601	11.7	**5.72E−08**	**0.000996**
5	81207967	rs6452434	ATG10 (100 Kb)	1.62	12.01	9.41E−08	0.001512
12	**70075768**	**rs10506628**	**TSPAN8 (0 Kb)**	**1.431**	12.37	**2.82E−07**	**0.004213**
9	13581905	rs13299116	FLJ41200 (−150 Kb)	1.343	14.89	3.34E−07	0.004509
15	95430437	rs11852342	NR2F2 (900 Kb)	1.843	7.77	3.45E−07	0.004509
19	43944077	rs2368524	LGALS7 (10 Kb)	1.323	16.26	4.77E−07	0.005863
5	91806412	rs10075974	AK056485 (30 Kb)	1.486	11.31	6.67E−07	0.007428
9	80588305	rs13299822	PSAT1 (500 Kb)	1.317	14.49	6.75E−07	0.007428
11	**120288578**	**rs17124538**	**GRIK4 (0 Kb)**	**1.439**	10.95	**7.30E−07**	**0.007628**
7	36028008	rs10249851	SEPT7 (−120 Kb)	1.484	9.89	7.89E−07	0.007807
1	64565272	rs12132851	UBE2U (80 Kb)	1.558	10.95	8.23E−07	0.007807
1	246068188	rs7512555	OR11L1 (2 Kb)	1.531	10.95	9.31E−07	0.007807
9	29616136	rs1930395	LINGO2 (−800 Kb)	1.42	12.02	9.49E−07	0.007807
3	151317233	rs10513360	PFN2 (−150 Kb)	1.461	12.36	9.50E−07	0.007807
11	**8140723**	**rs16936464**	**RIC3 (0 Kb)**	**1.548**	10.95	**9.71E−07**	**0.007807**
14	**36397262**	**rs7144374**	**SLC25A21 (0 Kb)**	**1.434**	**12.72**	**1.17E−06**	**0.009081**
12	**100625756**	**rs7980436**	**CHPT1 (0 Kb)**	**1.21**	15.9	**1.32E−06**	**0.009784**
10	126063061	rs2807064	OAT (12 Kb)	1.303	12.36	1.42E−06	0.009784
13	75765667	rs12585734	AX747676 (300 Kb)	1.514	9.9	1.42E−06	0.009784
7	83673368	rs11976275	SEMA3A (−13 Kb)	1.453	12.36	1.45E−06	0.009784

a Denotes a non-synonymous SNP.
MAF = maximum autocorrelation factor; bold lettering = SNPs that map within a gene.

associations with immune and neural tissue response factors contributing to this complex phenotype. Most of the genes found to be associated with lesion topology (captured by PC8) can be traced to specific neural (e.g. axonogenesis, transmission of nerve impulse) or immune (e.g. NFkB signalling, T-cell regulation) functions.

Convincing evidence for the involvement of *SYK*, *MYT1L*, *TRAPPC9*, *SLITKR6* and *RIC3* in the development and distribution of white matter lesions in multiple sclerosis can be gathered from the literature (Shimomura *et al.*, 1993; Kim *et al.*, 1997; Wrathall *et al.*, 1998; Negro *et al.*, 2002; Aruga, 2003; Aruga and Mikoshiba, 2003; Severance and Yolken, 2007; Vana *et al.*, 2007; Llorens *et al.*, 2008; Vrijenhoek *et al.*, 2008; Baranzini *et al.*, 2009*a*; Mochida *et al.*, 2009; Philippe *et al.*, 2009; Vierbuchen *et al.*, 2010; Vilarino-Guell *et al.*, 2010) (Supplementary material). The network-based strategy followed here uses a more liberal threshold of associations to include genes that are nominally significant ($P < 0.001$) so that, by virtue of interacting among each other, define a biologically plausible module. Thus, it is possible that genes that are not significantly associated with the trait are part of any given module if they serve to functionally link two or more associated genes. However, the proportion of these non-associated genes is generally low (15/48 in the top network). Interestingly, 19 of the 30 (63%) genes with modest associations ($10^{-7} > P > 10^{-3}$) that interact physically with the top associated genes are known to be expressed in the brain. Among these, SEMA3A, RTN4R, GRM7, LRRC4C and FYN are of particular importance in the CNS owing to their role in axon guidance during development.

New multiple sclerosis lesions begin as lymphocytic infiltration around a central vein, with subsequent centrifugal spread into surrounding tissue (Barnett and Prineas, 2004). In more established lesions, and perhaps also in reactivated ones, inflammation is most prominent at the outer margins of the lesion (Gaitan *et al.*, 2011; Lassmann, 2011). The clinical significance of the heterogeneity described here is uncertain but is likely to reflect the influence of a limited number of gene variants on the propagation of lesions, influencing their initiation, evolution, termination and/or repair.

A large GWAS on multiple sclerosis susceptibility was recently reported and notably, most of the associated genes have an immunological function (Sawcer *et al.*, 2011). Although to date

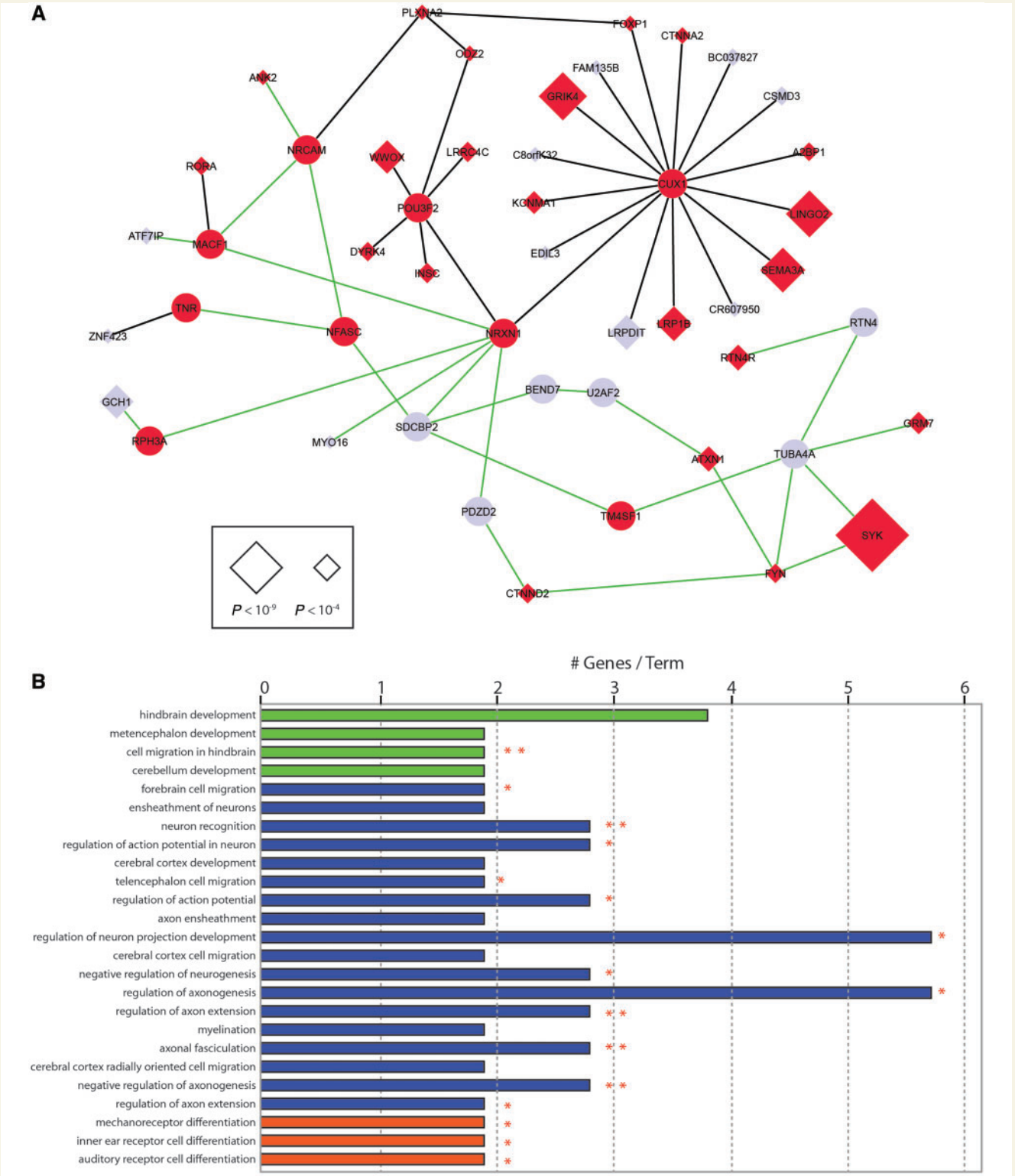

Figure 6 Top-scoring sub-network (module). (**A**) Circles and diamond represent proteins, and lines represent physical interactions (green: protein–protein, black: protein–DNA). Gene products with a significant *P*-value in the GWAS with PC8 are depicted as diamonds. Diamond size is proportional to the –log(10) of the *P*-value. Genes known to be expressed in the CNS are shown in red. (**B**) Gene ontology analysis. Bars represent gene ontology categories between GO levels 7 and 15. The length of each bar is proportional to the number of genes associated with that term. The percentage of associated genes for each term is indicated inside each bar. Asterisks denote Bonferroni-corrected (Fisher's exact test) *P*-values of enrichment for that term (*$P < 0.05$, **$P < 0.001$). Colours denote gene ontology groups with >50% similarity in their component genes.

genetic variation has been primarily associated with risk rather than expression of multiple sclerosis, earlier studies of gene-outcome associations have generally relied on bedside measures of clinical disability, which are insensitive to the number, volume or topology of brain lesions. In this study, we have identified an MRI-based, quantitative trait that is associated with common variants in patients with multiple sclerosis. The apparent strength of the associations and a compelling hypothesis for their interaction make these important candidate modifier genes for the multiple sclerosis phenotype, but independent replication of these findings is needed. The data reduction methods proposed substantially improve the potential for prospective studies to test these and related hypotheses. Joint analysis of quantitative imaging traits and genomic data should shed new light to the heterogeneity of multiple sclerosis and its responses to treatment.

Acknowledgements

The authors thank R. Lincoln, R. Guerrero, H. Mousavi, R. Gomez and A. Santaniello for sample and database management. The authors thank the patients with multiple sclerosis that participated in the study.

Funding

National Institutes of Health (RO1NS26799 to S.L.H., R01NS062885 to D.P.) and National Multiple Sclerosis Society (RG2901 to J.R.O.). S.E.B. is a Harry Weaver Neuroscience Scholar of the US National MS Society.

Supplementary material

Supplementary material is available at *Brain* online.

References

Aruga J. Slitrk6 expression profile in the mouse embryo and its relationship to that of Nlrr3. Gene Expr Patterns 2003; 3: 727–33.

Aruga J, Mikoshiba K. Identification and characterization of Slitrk, a novel neuronal transmembrane protein family controlling neurite outgrowth. Mol Cell Neurosci 2003; 24: 117–29.

Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. Hum Mol Genet 2009a; 18: 2078–90.

Baranzini SE, Srinivasan R, Khankhanian P, Okuda DT, Nelson SJ, Matthews PM, et al. Genetic variation influences glutamate concentrations in brains of patients with multiple sclerosis. Brain 2010; 133: 2603–11.

Baranzini SE, Wang J, Gibson RA, Galwey N, Naegelin Y, Barkhof F, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Hum Mol Genet 2009b; 18: 767–78.

Barnett MH, Prineas JW. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. Ann Neurol 2004; 55: 458–68.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc 1995; 57: 289–300.

Biffi A, Anderson CD, Desikan RS, Sabuncu M, Cortellini L, Schmansky N, et al. Genetic variation and neuroimaging measures in Alzheimer disease. Arch Neurol 2010; 67: 677–85.

Blum D, Yonelinas AP, Luks T, Newitt D, Oh J, Lu Y, et al. Dissociating perceptual and conceptual implicit memory in multiple sclerosis patients. Brain Cogn 2002; 50: 51–61.

Brett M, Leff AP, Rorden C, Ashburner J. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 2001; 14: 486–500.

Charil A, Zijdenbos AP, Taylor J, Boelman C, Worsley KJ, Evans AC, et al. Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets. Neuroimage 2003; 19: 532–44.

Cheung VG, Spielman RS. The genetics of variation in gene expression. Nat Genet 2002; 32 (Suppl): 522–5.

De Moor MH, Liu YJ, Boomsma DI, Li J, Hamilton JJ, Hottenga JJ, et al. Genome-wide association study of exercise behavior in Dutch and American adults. Med Sci Sports Exerc 2009; 41: 1887–95.

Enzinger C, Smith S, Fazekas F, Drevin G, Ropele S, Nichols T, et al. Lesion probability maps of white matter hyperintensities in elderly individuals: results of the Austrian stroke prevention study. J Neurol 2006; 253: 1064–70.

Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, Saxonov S, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet 2010; 6: e1000993.

Gaitan MI, Shea CD, Evangelou IE, Stone RD, Fenton KM, Bielekova B, et al. Evolution of the blood-brain barrier in newly forming multiple sclerosis lesions. Ann Neurol 2011; 70: 22–9.

Hauser SL, Oksenberg JR. The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration. Neuron 2006; 52: 61–76.

Hauser SL, Oksenberg JR, Lincoln R, Garovoy J, Beck RW, Cole SR, et al. Optic Neuritis Study Group. Interaction between HLA-DR2 and abnormal brain MRI in optic neuritis and early MS. Neurology 2000; 54: 1859–61.

Henry RG, Shieh M, Amirbekian B, Chung S, Okuda DT, Pelletier D. Connecting white matter injury and thalamic atrophy in clinically isolated syndromes. J. Neurol Sci 2009; 282: 61–6.

Hibar DP, Stein JL, Kohannim O, Jahanshad N, Saykin AJ, Shen L, et al. Voxelwise gene-wide association study (vGeneWAS): multivariate gene-based association testing in 731 elderly subjects. Neuroimage 2011; 56: 1875–91.

Hicks AA, Pramstaller PP, Johansson A, Vitart V, Rudan I, Ugocsai P, et al. Genetic determinants of circulating sphingolipid concentrations in European populations. PLoS Genet 2009; 5: e1000672.

Horakova D, Zivadinov R, Weinstock-Guttman B, Havrdova E, Tamano-Blanco M, Tyblova M, et al. HLA DRB1*1501 is only modestly associated with lesion burden at the first demyelinating event. J. Neuroimmunol 2011; 236: 76–80.

Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics 2002; 18 (Suppl 1): S233–40.

Kim JG, Armstrong RC, v Agoston D, Robinsky A, Wiese C, Nagle J, et al. Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. J Neurosci Res 1997; 50: 272–90.

Kutalik Z, Whittaker J, Waterworth D, Beckmann JS, Bergmann S. Novel method to estimate the phenotypic variation explained by genome-wide association studies reveals large fraction of the missing heritability. Genet Epidemiol 2011; 35: 341–9.

Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 2010; 467: 832–8.

Lassmann H. A dynamic view of the blood-brain barrier in active multiple sclerosis lesions. Ann Neurol 2011; 70: 1–2.

Lebouvier T, Scales TM, Williamson R, Noble W, Duyckaerts C, Hanger DP, et al. The microtubule-associated protein tau is also phosphorylated on tyrosine. J Alzheimers Dis 2009; 18: 1–9.

Llorens F, Gil V, Iraola S, Carim-Todd L, Marti E, Estivill X, et al. Developmental analysis of Lingo-1/Lern1 protein expression in the mouse brain: interaction of its intracellular domain with Myt1l. Dev Neurobiol 2008; 68: 521–41.

McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 2001; 50: 121–7.

Medland SE, Zayats T, Glaser B, Nyholt DR, Gordon SD, Wright MJ, et al. A variant in LIN28B is associated with 2D:4D finger-length ratio, a putative retrospective biomarker of prenatal testosterone exposure. Am J Hum Genet 2010; 86: 519–25.

Miller MI, Christensen GE, Amit Y, Grenander U. Mathematical textbook of deformable neuroanatomies. Proc Natl Acad Sci USA 1993; 90: 11944–8.

Mochida GH, Mahajnah M, Hill AD, Basel-Vanagaite L, Gleason D, Hill RS, et al. A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. Am J Hum Genet 2009; 85: 897–902.

Mowry EM, Beheshtian A, Waubant E, Goodin DS, Cree BA, Qualley P, et al. Quality of life in multiple sclerosis is associated with lesion burden and brain volume measures. Neurology 2009; 72: 1760–5.

Nan H, Kraft P, Qureshi AA, Guo Q, Chen C, Hankinson SE, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. J Invest Dermatol 2009; 129: 2250–7.

Narayanan S, Fu L, Pioro E, De Stefano N, Collins DL, Francis GS, et al. Imaging of axonal damage in multiple sclerosis: spatial distribution of magnetic resonance imaging lesions. Ann Neurol 1997; 41: 385–91.

Negro A, Brunati AM, Donella-Deana A, Massimino ML, Pinna LA. Multiple phosphorylation of alpha-synuclein by protein tyrosine kinase Syk prevents eosin-induced aggregation. FASEB J 2002; 16: 210–12.

Newton-Cheh C, Eijgelsheim M, Rice KM, de Bakker PI, Yin X, Estrada K, et al. Common variants at ten loci influence QT interval duration in the QTGEN Study. Nat Genet 2009; 41: 399–406.

Okuda DT, Srinivasan R, Oksenberg JR, Goodin DS, Baranzini SE, Beheshtian A, et al. Genotype-Phenotype correlations in multiple sclerosis: HLA genes influence disease severity inferred by 1HMR spectroscopy and MRI measures. Brain 2009; 132 (Pt 1): 250–9.

Philippe O, Rio M, Carioux A, Plaza JM, Guigue P, Molinari F, et al. Combination of linkage mapping and microarray-expression analysis identifies NF-kappaB signaling defect as a cause of autosomal-recessive mental retardation. Am J Hum Genet 2009; 85: 903–8.

Potkin SG, Turner JA, Fallon JA, Lakatos A, Keator DB, Guffanti G, et al. Gene discovery through imaging genetics: identification of two novel genes associated with schizophrenia. Mol Psychiatry 2009a; 14: 416–28.

Potkin SG, Turner JA, Guffanti G, Lakatos A, Torri F, Keator DB, et al. Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. Cogn Neuropsychiatry. 2009b; 14: 391–418.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–75.

Rueckert D, Aljabar P, Heckemann RA, Hajnal JV, Hammers A. Diffeomorphic registration using B-splines. Med Image Comput Comput Assist Interv 2006; 9 (Pt 2): 702–9.

Sawcer S, Hellenthal G, Pirinen M, Spencer CC, Patsopoulos NA, Moutsianas L, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011; 476: 214–9.

Sdika M. A fast nonrigid image registration with constraints on the Jacobian using large scale constrained optimization. IEEE Trans Med Imaging 2008; 27: 271–81.

Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Hum Brain Mapp 2009; 30: 1060–7.

Severance EG, Yolken RH. Lack of RIC-3 congruence with beta2 subunit-containing nicotinic acetylcholine receptors in bipolar disorder. Neuroscience 2007; 148: 454–60.

Shen L, Kim S, Risacher SL, Nho K, Swaminathan S, West JD, et al. Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. Neuroimage 2010; 53: 1051–63.

Shimomura R, Sakai K, Tanaka Y, Yonezawa K, Hashimoto E, Kasuga M, et al. Phosphorylation sites of myelin basic protein by a catalytic fragment of non-receptor type protein-tyrosine kinase p72syk and comparison with those by insulin receptor kinase. Biochem Biophys Res Commun 1993; 192: 252–60.

Siegmund KD, Connor CM, Campan M, Long TI, Weisenberger DJ, Biniszkiewicz D, et al. DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. PLoS ONE 2007; 2: e895.

Sombekke MH, Vellinga MM, Uitdehaag BM, Barkhof F, Polman CH, Arteta D, et al. Genetic correlations of brain lesion distribution in multiple sclerosis: an exploratory study. AJNR Am J Neuroradiol 2011; 32: 695–703.

Sondag CM, Dhawan G, Combs CK. Beta amyloid oligomers and fibrils stimulate differential activation of primary microglia. J Neuroinflammation 2009; 6: 1.

Stein JL, Hua X, Lee S, Ho AJ, Leow AD, Toga AW, et al. Voxelwise genome-wide association study (vGWAS). Neuroimage 2010; 53: 1160–74.

Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, et al. Two newly identified genetic determinants of pigmentation in Europeans. Nat Genet 2008; 40: 835–7.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010; 466: 707–13.

van der Sluis S, Verhage M, Posthuma D, Dolan CV. Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. PLoS ONE 2010; 5: e13929.

Vana AC, Lucchinetti CF, Le TQ, Armstrong RC. Myelin transcription factor 1 (Myt1) expression in demyelinated lesions of rodent and human CNS. Glia 2007; 55: 687–97.

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M. Direct conversion of fibroblasts to functional neurons by defined factors. Nature 2010; 463: 1035–41.

Vilarino-Guell C, Wider C, Ross OA, Jasinska-Myga B, Kachergus J, Cobb SA, et al. LINGO1 and LINGO2 variants are associated with essential tremor and Parkinson disease. Neurogenetics 2010; 11: 401–8.

Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet 2008; 83: 504–10.

Waxman SG. From neuroscience to neurology: neuroscience, molecular medicine, and the therapeutic transformation of neurology. Amsterdam; Boston: Elsevier Academic Press; 2005.

Wood AR, Hernandez DG, Nalls MA, Yaghootkar H, Gibbs JR, Harries LW, et al. Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. Hum Mol Genet 2011; 20: 4082–92.

Wrathall JR, Li W, Hudson LD. Myelin gene expression after experimental contusive spinal cord injury. J Neurosci 1998; 18: 8780–93.