# Regulatory genomic regions active in immune cell types explain a large proportion of the genetic risk of multiple sclerosis

Ramyiadarsini I Elangovan<sup>1</sup>, Giulio Disanto<sup>1,2</sup>, Antonio J Berlanga-Taylor<sup>1</sup>, Sreeram V Ramagopalan<sup>1,2</sup> and Lahiru Handunnetthi<sup>1</sup>

There is little understanding of how genetic variants discovered in recent genome-wide association studies are involved in the pathogenesis of multiple sclerosis (MS). We aimed to investigate which chromatin states and cell types explain genetic risk in MS. We used genotype data from 1854 MS patients and 5164 controls produced by the International Multiple Sclerosis Genetics Consortium and Wellcome Trust Case Control Consortium. We estimated the proportion of phenotypic variance between cases and controls explained by cell-specific chromatin state and DNase I hypersensitivity sites (DHSs) using the Genome-wide Complex Trait Analysis software. A large proportion of variance was explained by single-nucleotide polymorphisms (SNPs) in strong enhancer (SE) elements of immortalized B lymphocytes (5.39%). Three independent SNPs located within SE showed suggestive evidence of association with MS: rs12928822 (odds ratio (OR) = 0.81, 95% confidence interval (CI) = 0.73-0.89, P=2.48E-05), rs727263 (OR = 0.75, 95% CI = 0.66-0.85, P=3.26E-06) and rs4674923 (OR = 0.85, 95% CI = 0.79-0.92, P=1.63E-05). Genetic variants located within DHSs of CD19 + B cells explained the greatest proportion of variance. Genetic variants influencing the risk of MS are located within regulatory elements active in immune cells. This study also identifies a number of immune cell types likely to be involved in the causal cascade and that carry important implications for future studies of therapeutic design.

Journal of Human Genetics (2014) 59, 211-215; doi:10.1038/jhg.2014.3; published online 13 February 2014

Keywords: cellular immunity; chromatin profile; DNase hypersensitivity site; multiple sclerosis; single-nucleotide polymorphism

## INTRODUCTION

Multiple sclerosis (MS) is a complex condition characterized by demyelination in the central nervous system and progressive neurological dysfunction.<sup>1</sup> The importance of genetic factors in MS was recognized early in the study of this disease, in which a significantly elevated recurrence risk in biological relatives of affected individuals was observed.<sup>2,3</sup> The major histocompatibility complex (MHC) region makes the single, strongest genetic contribution to MS susceptibility.<sup>4</sup> In addition, recent genome-wide association studies (GWAS) have identified 110 non-MHC single-nucleotide polymorphisms (SNPs) that influence the risk of MS.<sup>5</sup> However, it is unclear how and in which cell types these risk variants exert their functional effects in the causal cascade of MS.

Chromatin is defined as the combination of DNA and nuclear proteins that regulate the expression of our genetic material. Chromatin profiles are highly cell specific and account for the large number of different cell types present in the human body. The Encyclopedia of DNA elements (ENCODE) project has recently profiled a variety of chromatin states, including regulatory regions (enhancers and promoters), repressed regions, heterochromatin (densely packed chromatin), insulator sites, transcribed regions and repetitive/copy number variation in a number of human cell types.<sup>6</sup> Similarly, DNase I hypersensitivity sites (DHSs) are highly cell specific and indicate regions of open chromatin that regulate gene expression through binding of transcription factors.<sup>7,8</sup> ENCODE researchers have also recently mapped DHSs across a variety of immune cell types, providing further insights into gene regulation in these cells.<sup>7</sup>

We have shown that GWAS data explain a considerable proportion (approximately 30%) of the phenotypic variance between MS cases and controls.<sup>9</sup> The aim of this study was to investigate which chromatin states and cell types explain most of this variance by integrating GWAS and chromatin profiling data. This analysis will advance the understanding of how genes influence disease risk and which cell types have a part in the causal pathways to MS.

E-mail: sreeram.ramagopalan@dpag.ox.ac.uk

<sup>&</sup>lt;sup>1</sup>Functional Genomics Unit, Department of Physiology, Anatomy and Genetics and Medical Research Council, University of Oxford, Oxford, UK and <sup>2</sup>Queen Mary University of London, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK

Correspondence: Dr SV Ramagopalan, Functional Genomics Unit, Department of Physiology, Anatomy and Genetics and Medical Research Council, University of Oxford, South Parks Road, Oxford OX1 3PT, UK.

Received 1 November 2013; revised 27 December 2013; accepted 28 December 2013; published online 13 February 2014

## MATERIALS AND METHODS

## Data acquisition

We used genotypic data on 475 806 SNPs from 1854 cases and 5164 controls from the United Kingdom produced by the International Multiple Sclerosis Genetics Consortium and the Wellcome Trust Case Control Consortium.<sup>5</sup> The chromatin profiles of immortalized B lymphocytes (lymphoblastoid cell lines (LCLs)), hepatocellular carcinoma cells (HepG2) and normal epidermal keratinocytes (NHEKs) were obtained from the ENCODE project.<sup>6</sup> Briefly, chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq) and expression data were used to identify different classes of chromatin states: active promoter, weak promoter, poised promoter, strong enhancer (SE), weak enhancer, polycomb repressed, heterochromatic (HC), insulator, strongly transcribed, weakly transcribed and repetitive/copy number variation.<sup>6</sup> DHSs maps for CD4 + T helper type 1 (Th1) cells, CD4 + Th17 cells, CD8 + T cells, CD19 + B cells and control HepG2 cells were also obtained from the ENCODE project.<sup>7</sup>

#### Data analysis

All genotyped SNPs were grouped based on their location within chromatin states and DHSs for each cell type. Estimates of the proportion of phenotypic variance between MS cases and controls explained by chromatin state and DHS-specific SNPs were calculated using the Genome-wide Complex Trait Analysis tool (http://gump.qimr.edu.au/gcta).<sup>10</sup> Heritability on the observed scale (proportion of phenotypic variance owing to additive genetic effects) was first estimated via residual maximum-likelihood analysis. The estimate was then transformed on the liability scale as previously described and assuming a disease prevalence of 0.001.<sup>10,11</sup> Association analysis of GWAS SNPs was conducted using PLINK (http://pngu.mgh.harvard.edu/purcell/plink/).<sup>12</sup> The potential functional effects of associated SNPs on gene expression was assessed using RegulomeDB, which is a large database of expression quantitative trait loci and predicted regulatory elements from a variety of cell types including lymphoblastoid cell lines (LCLs).<sup>13</sup>

## RESULTS

#### Calculating variance by each chromatin state for each cell type

Given the well-established immunological nature of MS, we hypothesized that a large proportion of variance would arise from regions of active chromatin in LCLs, and thus calculated the variance explained by each chromatin state in this cell type (Table 1).

We found that SNPs lying in HC explained most of the phenotypic variance between MS cases and controls (8.71%, s.e. = 2.12%). This is

Table 1 Percentage of variance explained by each chromatin state in LCLs, together with the size of each chromatin state within the genome

Chromatin state	<i>Variance explained</i> (%)	s.e. (%)	Proportion of genome covered (%)
INS	0.02	0.52	0.5
RCNV	0.05	0.13	0.2
PP	0.50	0.22	0.1
PR	1.11	0.90	3.1
ST	1.65	0.93	6.2
WE	1.95	1.05	3.1
AP	2.04	0.51	7.8
WT	2.32	1.29	10.6
WP	3.09	0.54	0.7
SE	5.39	0.83	1.7
HC	8.71	2.12	72.5

Abbreviations: AP, active promoter; HC, heterochromatic; LCL, lymphoblastoid cell line; PP, poised promoter; PR, polycomb repressed; INS, insulator; RCNV, repetitive/copy number variation; SE, strong enhancer; ST, strongly transcribed; WE, weak enhancer; WP, weak promoter; WT, weakly transcribed.

expected as HC represents a very large proportion of the genome (72.5%), and the vast majority of the genotyped SNPs are located within HC (352 578 out of 475 806). However, apart from HC, SE elements in LCLs accounted for the largest proportion of variance (5.39%, s.e. = 0.83%), despite their small representation in the genome (1.7%). Given that most of the genetic risk of MS was driven by the MHC region (9.08%, s.e. = 3.4%), we repeated the analysis after removing all MHC SNPs. In this scenario, the percentage variance explained by SNPs in SE regions was still the highest (after HC) at 4.40%, s.e. = 0.82% (Supplementary Table 1).

To show that regulatory regions specifically active in LCLs but not in other cell types were responsible for genetic risk in MS, we calculated the variance explained by each chromatin state for two additional control cell types that were unrelated to MS etiology (hepatocytes (HepG2) and keratinocytes (NHEKs)) (Figure 1). Once again, for the control cell lines, SNPs lying in HC explained most of the phenotypic variance between MS cases and controls. However, in contrast to LCLs, for both the HepG2 and NEHK cell lines, SE accounted for a minimal amount of the variance (0% and 2.18%, respectively), and actively repressed SNPs (within polycomb repressed regions) accounted for a very large contribution to the variance (8.80% and 5.87%, respectively).

The relevance of regulatory elements specifically active in LCLs appeared even clearer when we calculated the ratio of variance explained by each chromatin state between LCLs and the average of both control cell lines (Figure 2). SE SNPs in LCLs explained a proportion of variance that was 4.94 times higher than that explained by SE SNPs in HepG2 and NHEK cell lines, followed by weak promoter (4.19 times) and active promoter SNPs (4.09 times). This supports our hypothesis that active chromatin states that regulate gene expression in LCLs account for a considerable proportion of phenotypic variance between MS cases and controls.

#### Finding significant genetic associations within SE elements

We then reasoned that by testing only SNPs located within SE elements for genetic association, we could restrict the analysis to only those variants that are more likely to influence MS risk. This would reduce the significance threshold required for multiple testing correction and therefore increase our statistical power to detect associations. We performed an association test for all SNPs located within SE regions in LCLs and found four SNPs with suggestive association with MS (Table 2).

The only SNP that survived Bonferroni correction (corrected P < 0.05) was rs727263 on chromosome 13 and located within the gene UBAC2. Among the other SNPs, there were two genetic variants that have been previously associated with MS (rs12927773) and celiac disease (rs12928822). These two SNPs are in strong linkage disequilibrium ( $r^2 = 0.997$ ); therefore, they likely represent one single association signal. Interestingly, another SNP located within the same genomic region (rs7200786, within CLEC16A) is known to be associated with MS.5 However, rs12927773 and rs12928822 are not in linkage disequilibrium with rs7200786 ( $r^2 = 0.003$  for both SNPs) and the association of both rs12927773 and rs12928822 remained significant when the analysis was conditioned on the genotype at rs7200786 (rs1292773: odds ratio (OR) = 0.81, 95% confidence interval (CI) = 0.73-0.90, P = 3.8E - 05; rs12928822: OR = 0.81, 95% CI = 0.73–0.89, P = 3.6E - 05). We investigated whether any of these SNPs or SNPs in strong linkage disequilibrium  $(r^2 > 0.9)$  with them had been previously associated with gene expression using RegulomeDB. A total of 115 SNPs were tested but no expression quantitative trait loci for any gene in any cell type was identified.





Figure 1 Bar graph showing the variance explained by each chromatin state for LCLs, HepG2 and NEHK cell lines.



#### Ratio of variance for each chromatin states between the 3 cell types

Figure 2 Bar graph showing the ratio of the variance explained between LCLs and the average of NHEK and HepG2 for each chromatin state.

## Table 2 SNPs located within SE elements in LCLs with suggestive evidence of association with MS

					Bonferroni
				Uncorrected	corrected
SNP	Chr.	Position	OR (with CI)	P-value	P-value
rs12927773	16	11 403 963	0.81 (0.73–0.89)	2.78E-05	0.274
rs12928822	16	11 403 893	0.81 (0.73–0.89)	2.48E-05	0.2446
rs727263	13	100 004 108	0.75 (0.66–0.85)	3.26E - 06	0.0321
rs4674923	2	225 426 527	0.85 (0.79–0.92)	1.63E –05	0.1598

Abbreviations: Chr., chromosomes; CI, confidence interval; LCL, lymphoblastoid cell lines; OR, odds ratio; SE, strong enhancer; SNP, single-nucleotide polymorphism.

## Estimating phenotypic variance contributed by cell-specific active SNPs

The chromatin profile is highly cell specific and can considerably vary between different components of the immune system. Therefore, we used DHSs data for a variety of immune cell types and grouped all genotyped SNPs based on their location relative to these cell specific DHSs. We found that the proportions of variance explained by SNPs located within DHSs of CD19 + B cells, Th1 cells, CD8 + T cells, Th17 cells and HepG2 cells were 11.98% (s.e. = 0.82%), 10.11% (s.e. = 0.81%), 9.79% (s.e. = 0.79%), 7.53% (s.e. = 0.73%) and 4.01% (s.e. = 0.72%), respectively.

We reasoned that many of the SNPs located within immune DHSs were shared across more than one cell type. Therefore, after pooling DHSs from Th1 and Th17 into a single Th cell group, we grouped SNPs into those that were located within: (1) only CD19 + B DHSs; (2) only Th DHSs; (3) only CD8 + T DHSs; (4) Th and CD19 + B but not CD8 + T DHSs; (5) CD8 + T and CD19 + B but not Th



Figure 3 Proportion of phenotypic variance explained by SNPs within (1) only CD19+ B DHSs; (2) only Th DHSs; (3) only CD8+ T DHSs; (4) Th and CD19+ B but not CD8+ T DHSs; (5) CD8+ T and CD19+ B but not Th DHSs; (6) Th and CD8+ T but not CD19+ B DHSs; (7) Th, CD8+ T and CD19+ B DHSs.

DHSs; (6) Th and CD8 + T but not CD19 + B DHSs; (7) Th, CD8 + T and CD19 + B DHSs. The proportion of phenotypic variance explained by SNPs located in only CD19 + B DHSs (3.75%, s.e. = 0.58%) was higher than that explained by SNPs in only CD8 + T DHSs (0.92%, s.e. = 0.39%) and Th DHSs (2.53%, s.e. = 0.59%). However, the highest proportion of variance was explained by SNPs located in DHSs that were shared across all cell types (3.83%, s.e. = 0.61%) (Figure 3).

## DISCUSSION

MS is a complex disorder of unknown etiology. We have shown that SNPs accounting for most of the genetic risk associated with MS are located within regulatory regions that are specifically active in immune cells. In particular, the proportion of variance explained by active enhancer and promoter elements in LCLs was almost five times higher than was observed in the two control cell types. In contrast, a large proportion of variance was explained by SNPs located within repressed genomic regions of the two control cell types, suggesting that MS risk variants are likely to be functionally inactive in non-immune cells. Even after removing variants falling within the MHC region, SE in LCLs still accounted for a significant percentage of the genetic risk of MS. This indicates an important role for genetic variation within SEs and thus gene expression in influencing the risk of MS. It is interesting to find that this effect is not confined to the MHC but is homogeneously distributed across the genome.

By testing for association of only SNPs located within SE elements in LCLs, we were able to reduce the number of tests performed and identified four SNPs that were suggestive of association with MS. One of them (rs727263) was located within *UBAC2*, a gene that has been previously associated with the risk of Behcet's disease.<sup>14,15</sup> Two additional SNPs located on chromosome 16 showed suggestive association with MS (rs12928822 and rs12927773). These two SNPs are in strong linkage disequilibrium with each other and their association is independent of another confirmed MS susceptibility locus in the same genomic region (rs7200786 within *CLEC16A*).<sup>5</sup> Notably, rs12928822 and

rs12927773 have been previously associated with the risk of MS and coeliac disease.<sup>16,17</sup> Several genes are located near rs12928822 and rs12927773 including *PRM1*, *PRM2*, *SOCS1* and *TNP2*. Interestingly, *SOCS1* is involved in the suppression of cytokine signaling required for downregulation of immune cell function and therefore represents a plausible candidate.<sup>18</sup> However, these associations would need to be replicated in an independent cohort of individuals before they can be considered established MS-associated loci.

SNPs located within immune-specific DHSs explained a larger proportion of variance than SNPs located in DHSs of a non-immune cell type such as HepG2. Interestingly, we found that among different immune cell types, the proportion of variance explained by CD19+ B-cell-specific SNPs was greater than that explained by Th1, Th17 and CD8 + cytotoxic T-cell-specific SNPs. These results emphasize the role played by B cells in the pathogenesis of MS. The most common immunological finding in MS patients is the presence of IgG oligoclonal bands in their cerebrospinal fluid and this finding lends support to the presence of an abnormal B-cell activation within their central nervous system.<sup>19</sup> Furthermore, B-cell abnormalities influence both conversion to clinically definite MS, MRI activity, onset of relapses and disease progression.<sup>20-25</sup> Finally, clinical trials have shown that antibody-mediated depletion of B cells is highly effective in diminishing MRI activity and onset of clinical relapses.<sup>26,27</sup> However, the largest proportion of variance was explained by those SNPs located within DHSs shared across all these cell types. This finding further highlights the complexity of this disease and suggests that the etiology of MS is unlikely to be driven by a single cell type.

To conclude, we have used a novel approach to integrate functional genomics and GWAS data, and it is shown that SNPs located within regulatory elements active in immune cells (particularly in B and T cells) explain a large proportion of the phenotypic variance between MS cases and healthy controls. Genetic variants that influence the risk of MS are therefore likely to act by changing the chromatin landscape and influencing the expression of neighboring genes. Similar analyses in other immunological cell types relevant to MS and functional studies are required to further elucidate how MS-associated genetic variants exert their effects in the causal cascade. This approach is likely to yield more specific and effective treatments in the future than what is currently available.

## CONFLICT OF INTEREST

RIE, GD, AJB, LH and SVR report no competing interest.

#### ACKNOWLEDGEMENTS

This work was funded by the Medical Research Council (Grant No. G0801976) and a research fellowship FISM-Fondazione Italiana Sclerosi Multipla-Cod. (2010/B/5 to GD). The study sponsors had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; and preparation, review or approval of the manuscript. All authors state that this research was carried out independently of the influence of funding bodies. RIE and SVR had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Author contributions: Study concept and design: Disanto, Handunnetthi and Ramagopalan. Acquisition of data: Elangovan, Disanto and Ramagopalan. Analysis and interpretation of data: Elangovan and Disanto. Drafting of the manuscript: Elangovan. Critical revision of the manuscript for important intellectual content: Disanto, Berlanga-Taylor, Handunnetthi and Ramagopalan. Study supervision: Ramagopalan.

- 1 Compston, A. & Coles, A. Multiple sclerosis. Lancet 372, 1502–1517 (2008).
- 2 Ebers, G. C., Sadovnick, A. D. & Risch, N. J. Canadian Collaborative Study Group. A genetic basis for familial aggregation in multiple sclerosis. *Nature* **377**, 150–151 (1995).
- 3 Willer, C. J., Dyment, D. A., Risch, N. J., Sadovnick, A. D. & Ebers, G. C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl Acad. Sci.* USA 100, 12877–12882 (2003).
- 4 Ramagopalan, S. V. & Ebers, G. C. Multiple sclerosis: major histocompatibility complexity and antigen presentation. *Genome Med.* 1, 105 (2009).
- 5 International multiple sclerosis genetics consortium (IMSGC), Beecham, A. H., Patsopoulos, N. A., Xifara, D. K., Davis, M. F., Kemppinen, A. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013).
- 6 Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- 7 Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
- 8 Sheffield, N. C., Thurman, R. E., Song, L., Safi, A., Stamatoyannopuolos, J. A., Lenhard, B. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res.* 23, 777–788 (2013).
- 9 Watson, C. T., Disanto, G., Breden, F., Giovannoni, G. & Ramagopalan, S. V. Estimating the proportion of variation in susceptibility to multiple sclerosis captured by common SNPs. *Sci. Rep.* 2, 770 (2012).
- 10 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. 88, 76–82 (2011).
- 11 Lee, S. H., Wray, N. R., Goddard, M. E. & Visscher, P. M. Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.* 88, 294–305 (2011).
- 12 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
- 13 Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797 (2007).
- 14 Sawalha, A. H., Hughes, T., Nadig, A., Vuslat, Y., Kenan, A., Gokhan, K. et al. A putative functional variant within the UBAC2 gene is associated with increased risk of Behcet's disease. Arthritis. Rheum. 63, 3607–3612 (2011).

- 15 Hou, S., Shu, Q., Jiang, Z., Chen, Y., Li, F., Chen, F. et al. Replication study confirms the association between UBAC2 and Behcet's disease in two independent Chinese sets of patients and controls. Arthritis Res. Ther. 14, R70 (2012).
- 16 Dubois, P. C., Trynka, G., Franke, L., Hunt, K. A., Romanos, J., Curtotti, A. *et al.* Multiple common variants for celiac disease influencing immune gene expression. *Nat. Genet.* **42**, 295–302 (2010).
- 17 International Multiple Sclerosis Genetics Consortium. Comprehensive follow-up of the first genome-wide association study of multiple sclerosis identifies KIF21B and TMEM39A as susceptibility loci. *Hum. Mol. Genet.* **19**, 953–962 (2010).
- 18 Leikfoss, I. S., Mero, I. L., Dahle, M. K., Lie, B. A., Harbo, H. F., Spurkland, A. *et al.* Multiple sclerosis-associated single-nucleotide polymorphisms in CLEC16A correlate with reduced SOCS1 and DEXI expression in the thymus. *Genes Immun.* **14**, 62–66 (2013).
- 19 Freedman, M. S., Thompson, E. J., Deisenhammer, F., Giovanonni, G., Grimsley, G., Keir, G. *et al.* Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement. *Arch. Neurol.* **62**, 865–870 (2005).
- 20 Disanto, G., Morahan, J. M., Barnett, M. H., Giovannoni, G. & Ramagopalan, S. V. The evidence for a role of B cells in multiple sclerosis. *Neurology* 78, 823–832 (2012).
- 21 Howell, O. W., Reeves, C. A., Nicholas, R., Carrassiti, D., Roncaroli, F., Magliozzi, R. et al. Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain* **134**, 2755–2771 (2011).
- 22 Brettschneider, J., Czerwoniak, A., Senel, M., Fang, L., Kassubeck, J., Pinkhardt, E. et al. The chemokine CXCL13 is a prognostic marker in clinically isolated syndrome (CIS). PLoS One 5, e11986 (2010).
- 23 Brettschneider, J., Tumani, H., Kiechle, U., Muche, R., Richards, G., Lehmensiek, V. et al. IgG antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome. *PLoS One* **4**, e7638 (2009).
- 24 Joseph, F. G., Hirst, C. L., Pickersgill, T. P., Ben-Shlomo, Y., Robertson, N. P. & Scolding, N. J. CSF oligoclonal band status informs prognosis in multiple sclerosis: a case control study of 100 patients. *J Neurol. Neurosurg. Psychiatry* **80**, 292–296 (2009).
- 25 Khademi, M., Kockum, I., Andersson, M. L., Iaocobaeus, E., Brundin, L., Sellebjerg, F. et al. Cerebrospinal fluid CXCL13 in multiple sclerosis: a suggestive prognostic marker for the disease course. *Mult. Scler.* **17**, 335–343 (2011).
- 26 Hauser, S. L., Waubant, E., Arnold, D. L., Vollmer, T., Antel, J., Fox, R. J. et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N. Engl. J. Med.* 358, 676–688 (2008).
- 27 Kappos, L., Li, D., Calabresi, P. A., O'Connor, P., Bar-Or, A., Barkhof, F. et al. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebocontrolled, multicentre trial. *Lancet* **378**, 1779–1787 (2011).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)