DNase hypersensitive sites and association with multiple sclerosis

Giulio Disanto[1,2,†], Geir Kjetil Sandve[3,†], Vito A.G. Ricigliano[1], Julia Pakpoor[1], Antonio J. Berlanga-Taylor[1], Adam E. Handel[1], Jens Kuhle[2], Lars Holden[4], Corey T. Watson[5], Gavin Giovannoni[2], Lahiru Handunnetthi[1] and Sreeram V. Ramagopalan[1,*]

[1]Department of Physiology, Anatomy and Genetics and Medical Research Council Functional Genomics Unit, University of Oxford, South Parks Road, Oxford OX1 3PT, UK, [2]Queen Mary University of London, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, UK, [3]Department of Informatics and Centre for Cancer Biomedicine, University of Oslo, Blindern, Oslo, Norway, [4]Statistics for Innovation, Norwegian Computing Center, 0314 Oslo, Norway and [5]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia

Genome-wide association studies (GWASs) have shown that approximately 60 genetic variants influence the risk of developing multiple sclerosis (MS). Our aim was to identify the cell types in which these variants are active. We used available data on MS-associated single nucleotide polymorphisms (SNPs) and deoxyribonuclease I hypersensitive sites (DHSs) from 112 different cell types. Genomic intervals were tested for overlap using the Genomic Hyperbrowser. The expression profile of the genes located nearby MS-associated SNPs was assessed using the software GRAIL (Gene Relationships Across Implicated Loci). Genomic regions associated with MS were significantly enriched for a number of immune DHSs and in particular T helper (Th) 1, Th17, CD8+ cytotoxic T cells, CD19+ B cells and CD56+ natural killer (NK) cells (enrichment = 2.34, 2.19, 2.27, 2.05 and 1.95, respectively; $P < 0.0001$ for all of them). Similar results were obtained when genomic regions with suggestive association with MS and additional immune-mediated traits were investigated. Several new candidate MS-associated genes located within regions of suggestive association were identified by GRAIL (*CARD11, FCRL2, CHST12, SYK, TCF7, SOCS1, NFKBIZ* and *NPAS1*). Genetic data indicate that Th1, Th17, cytotoxic T, B and NK cells play a prominent role in the etiology of MS. Regions with confirmed and suggestive association have a similar immunological profile, indicating that many SNPs truly influencing the risk of MS actually fail to reach genome-wide significance. Finally, similar cell types are involved in the etiology of other immune-mediated diseases.

INTRODUCTION

Multiple sclerosis (MS) is a highly debilitating disorder of the central nervous system characterized by immune-mediated demyelination and progressive neurological dysfunction (1). Both genetic and environmental factors have been shown to influence the risk of MS and a number of immune modulating treatments are either available to MS patients or currently being tested (2–5). Despite these recent advances in MS research, the exact mechanisms driving myelin and neuronal loss are largely unknown. As a consequence, most MS treatments are unspecific and target a large number of cell types rather than being tailored therapies. Understanding the agents driving MS onset and their cell type specificity is therefore necessary to improve our strategies of disease prevention and treatment.

It is plausible that genetic variants associated with a certain disease would act through influencing the cell type(s) triggering its onset. Deoxyribonuclease I (DNase I) hypersensitive sites (DHSs) represent precise and reliable indicators of open and active chromatin across the genome. Recent studies have shown

that these marks are highly cell specific and can be used to identify pathogenic cell types in complex traits (6). The Encyclopedia of DNA Elements (ENCODE) Project has investigated the overlap between MS-associated genetic variants and DHSs across a number of cell types (6). In particular, ENCODE researchers used data from a recently published meta-analysis of genome-wide association studies (GWASs) in MS (7) and compared the proportion of SNPs meeting increasingly significant association *P*-values thresholds that were located inside DHSs of each cell type to the proportion of all SNPs inside DHSs of the same cell type (6).

However, this strategy is biased by the considerable variability of linkage disequilibrium (LD) across the genome. Indeed, associated SNPs may be in strong LD with either a very large number (even hundreds) or a small number (only a few) SNPs. The inevitable consequence of this disparity is that some associated SNPs will count more than others in the overall estimate. This represents a serious issue in the case of MS where the main association is located in the major histocompatibility complex (MHC), in which LD is much stronger than in the rest of the genome. Indeed, ENCODE concluded that MS-associated variants were particularly active in CD3+ cord blood T cells, but this association was driven by SNPs with *P*-values $<10^{-10}$ (6). Notably, in MS only a large number of SNPs that are located within the MHC region reach such small association *P*-values ($<10^{-10}$) (7).

Although most genetic susceptibility to MS resides within the MHC, more than 50 genetic loci located outside this region have now been associated with MS risk and many others show suggestive association (7, 8). We therefore aimed to provide a more realistic estimate of the overlap between MS-associated variants and DHSs by assigning an equal weight to each independent association. Finally, we investigated additional diseases and assessed to what extent different immune-mediated traits share common pathogenic cell types.

RESULTS

We collected data on MS-associated SNPs from the 'Catalog of Published Genome-Wide Association Studies' (http://www.genome.gov/gwastudies/). We estimated the tendency of these associations to be located within regions of cell-specific open chromatin using a set of 147 DHSs tracks from 112 different cell types produced by the ENCODE project (Supplementary Material, Table S1) (see Methods) (6). The overlap between MS-associated regions and DHSs was higher than that expected only for a number of immune cell types. The highest enrichment values were observed for various T-cell types and in particular Th1, Th17 and CD8+ cytotoxic T cells (enrichment = 2.34, 2.19 and 2.27, respectively; *P* < 0.0001 for all of them). DHSs in CD19+ B cells and CD56+ natural killer (NK) also showed a striking tendency of overlap with MS-associated regions (enrichment = 2.05 and 1.95, respectively; *P* < 0.0001) (Fig. 1 and Supplementary Material, Table S2). The overlap between immune DHSs and MS-associated regions remained similarly significant when correcting for the distance of MS-associated SNPs from coding regions (*P* < 0.001 for all cell types), and there was a very strong correlation between the original and the distance-corrected *P*-values (Pearson's *r* = 0.967, *P* < 2.2e − 16) (Supplementary Material, Fig. S1).

We then calculated enrichment considering SNPs with confirmed and suggestive evidence of association as defined in methods. There was a clear overrepresentation of immune DHSs in both groups, and there was a considerable correlation between confirmed and suggestive enrichment values (Pearson's rho = 0.86; *P* < 2.2e − 16) (Fig. 2 and Supplementary Material, Table S3). We further confirmed the immunological nature of suggestive associations by using the software GRAIL (Gene Relationships Across Implicated Loci) (9). This tool is able to examine relationships between genes in disease-associated
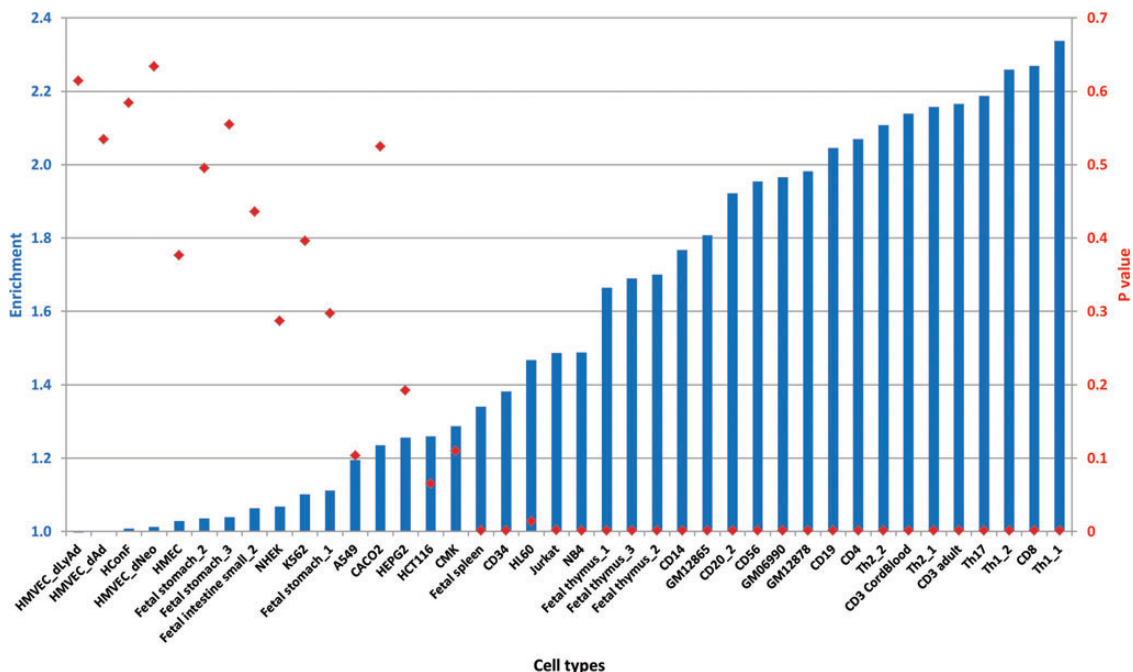

Figure 1. Enrichment and significance of overlap between MS-associated genomic regions and cell-specific DHSs. Only the top 40 cell types are shown in the figure. For cell types with more than one replicate available, the number after the underscore indicates the number of replicates (e.g. Th1_2 = Th1 second replicate).
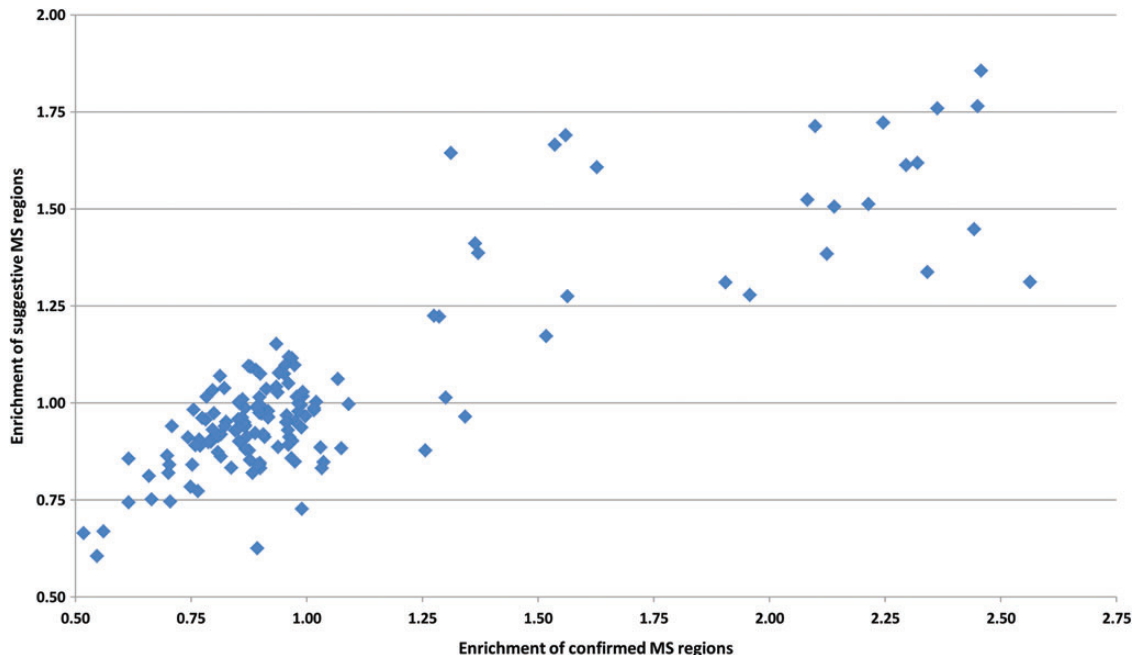
Figure 2. Correlation between enrichment of regions with confirmed versus suggestive association with MS. Each dot represents a different cell type.

loci based on the Novartis Gene Expression Atlas (see Methods) (9, 10). We observed that both confirmed and suggestive MS candidate genes were expressed in the several immune cell types that were also suggested by the DHSs analysis (CD4+ and CD8+ T, B and NK cells (Supplementary Material, Table S4)). We found a large number of connections between genes with confirmed and suggestive association with MS and eight genes located within suggestive regions were connected within the network more than expected by chance only (*CARD11, FCRL2, CHST12, SYK, TCF7, SOCS1, NFKBIZ* and *NPAS1*) (Fig. 3 and Supplementary Material, Table S4).

Finally, we investigated the DHSs profile of 11 additional complex traits. Among immune-mediated diseases, the top enriched cell types were CD8+ T cells in psoriasis, Th17 in type 1 diabetes and rheumatoid arthritis, Th1 in celiac disease and Crohn's disease, cord blood CD3+ T and CD19+ B cells in systemic sclerosis, B cells in systemic lupus erythematosus, CD14+ monocytes and colorectal carcinoma cells in ulcerative colitis. Among non-immune-mediated diseases, the top cell types were hepatocellular carcinoma cells and fetal intestine in colorectal cancer, CD14+ monocytes in leprosy and fetal brain in schizophrenia (Table 1 and Supplementary Material, Table S2).

DISCUSSION

We observed that genomic regions with confirmed association with MS overlap with immune-specific DHSs more than expected by chance only. Notably, we did not find any evidence for MS-associated genetic variants being particularly enriched for neuronal active chromatin. These findings add further support to the hypothesis that the pathogenesis of MS is primarily immune-mediated (8, 11). By assigning an equal weight to each MS association (rather than inflating the final estimate

with MHC variants), we were able to show that DHSs from a variety of T cell subtypes (CD8+, Th1, Th17), CD19+/ CD20+ B and CD56+ NK cells have a strong tendency to overlap with MS-associated regions. This is substantially different from the previous observation that pointed at cord blood CD3+ T cells as the major determinant of MS pathogenesis and considerably underestimated the role played by additional cell types (in particular Th1, Th17 and CD8+T) (6).

Interestingly, the same overrepresentation of DHSs from immune cells was found when using SNP with both suggestive and confirmed association with MS. This observation was further confirmed by the fact that genes located near confirmed and suggestive genetic associations are particularly expressed in the same cell types highlighted by the DHSs analysis. Taken together these findings strongly suggest that many genetic variants that influence the risk of MS actually fail to reach genome-wide significance and additional common variants associated with MS remain to be uncovered.

That an important role in MS is played by T cells is perhaps unremarkable, although we here show the importance of both CD4+ helper T and CD8+ T subsets. MS-associated genomic regions are enriched for genes involved in T-cell proliferation and activation (8). Furthermore, the main MS genetic risk factor resides in the MHC class II region, which plays a central role in the development of CD4+ T-cell central tolerance (12, 13). The importance of CD8+ T cells is increasingly being recognized, especially given the fact that clonally expanded CD8+ T cells typically outnumber CD4+ T cells in active MS lesions and that HLA class I genes (necessary for CD8+ T central tolerance development) also influence MS risk (8, 14–16). Accumulating evidence also supports a role for NK cells and their receptors in regulating tissue damage and autoimmune responses in both EAE mice and human MS patients (17, 18).

The role of B cells in the pathogenesis of MS has been more debated. We have recently shown that MS-associated regions
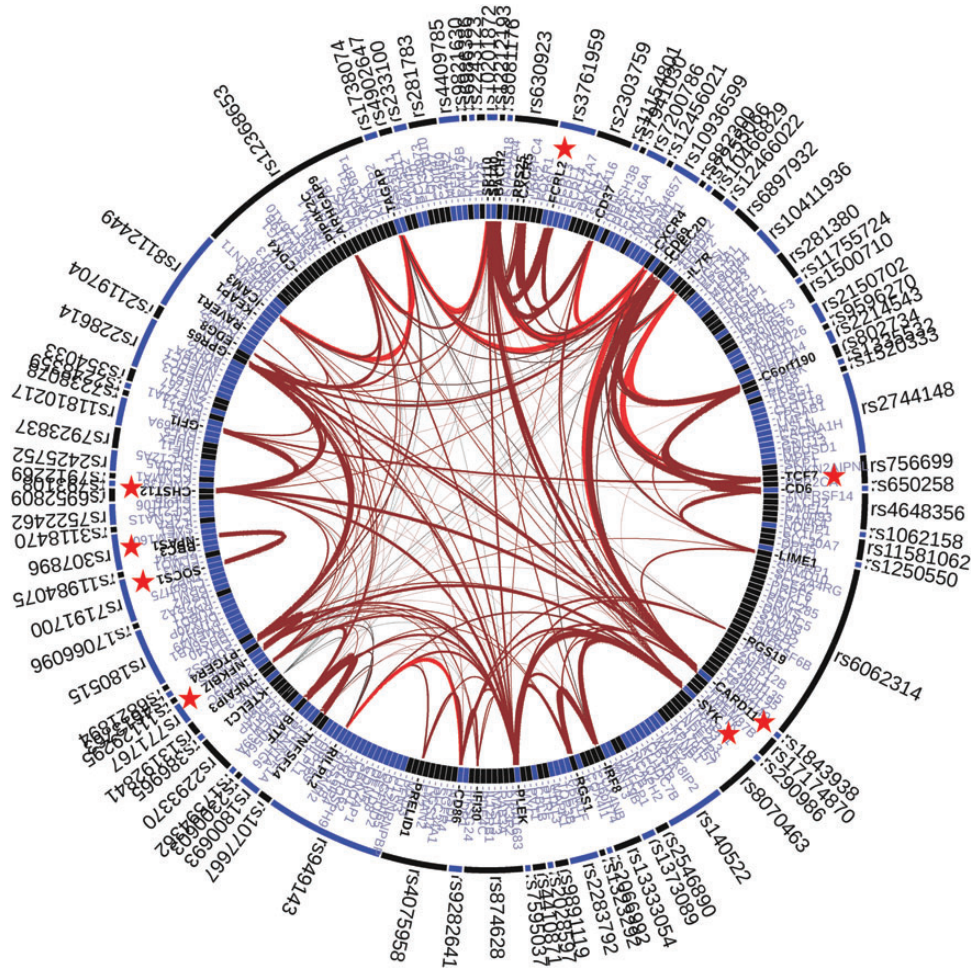
Figure 3. Gene–gene connections found by GRAIL based on the Novartis Gene Expression Atlas. Those genes that participate in the network are indicated in bold, while red stars indicate genes located within the regions of suggestive association that are connected within the network more than expected by chance.

Table 1. Top enriched cell types and tissues in the remaining 11 conditions tested (GM12878 = B lymphoblastoid cell line; HEPG2 = hepatocellular carcinoma cells)

Disease	Top cell type/tissue	Enrichment	*P*-value
Rheumatoid arthritis	Th17	3.26	<0.001
Type 1 diabetes	Th17	2.68	<0.001
Celiac disease	Th1	2.77	<0.001
Crohn's disease	Th1	2.44	<0.001
Ulcerative colitis	CD14+ monocytes	2.15	<0.001
Systemic lupus erythematosus	GM12878	2.58	<0.001
Psoriasis	CD8+ T	2.87	<0.001
Systemic sclerosis	CD3+ cord blood T	3.52	<0.001
Leprosy	CD14+ monocytes	2.72	0.0015
Colorectal cancer	HEPG2	1.84	0.021
Schizophrenia	Fetal brain	1.49	0.042

overlap with active enhancer and promoter elements of immortalized B-cell lines more than expected by chance and more than in non-immune cell types (19). We are now able to confirm that our previous findings were not biased by the immortalization process and that MS-associated variants considerably overlap with genomic regions that are active in B cells. These findings are in

agreement with a variety of observations, suggesting that B cells play a fundamental role in MS (20). The presence of oligoclonal IgG bands within the CSF is the most consistent immunological finding in MS patients and indicates abnormal B-cell activation in the CNS (21). Furthermore, B-cell abnormalities influence both conversion from clinically isolated syndrome to clinically definite MS, magnetic resonance imaging (MRI) activity, onset of relapses and disease progression (22–26). Notably, monoclonal antibodies targeting the B-cell population (rituximab and ocrelizumab) cause a substantial decrease in MRI activity and onset of relapses (4, 27).

We performed similar overlap analyses between cell-specific DHSs and genomic regions associated with 11 additional complex conditions. The method used was sensitive enough to discriminate between diseases driven by different components of the immune system. For example, genomic regions associated with systemic lupus erythematosus and psoriasis were particularly active in B and CD8+ T cells, respectively, while rheumatoid arthritis, Crohn's disease and celiac disease showed a marked T helper profile. The observation of a high enrichment of CD14+ and colorectal DHSs in ulcerative colitis associated regions adds further evidence for host–microbe interactions at the colonic level in shaping the risk of this condition (28). Genetic

variants active in CD14+ monocytes also influence the risk of leprosy, and this is expected given the role of the innate immune response against this infectious condition (29). Finally, we found that schizophrenia-associated variants tend to be active in the developing brain further confirming the neurodevelopmental origin of this disease (30).

DHSs data are strongly correlated between similar cell types such as different T-cell subsets. It is therefore not surprising that related cell types show similarly high enrichment scores for the same disease. For example, Th1, Th17 and Th2 DHSs are all highly enriched for type 1 diabetes and Crohn's disease associations. Therefore, to correctly interpret our results, one should examine the full list of enrichment and significance values across all cell types and treat small differences in enrichment with caution. To conclude, we show how the integration of GWAS and functional genomics information can help our understanding of complex diseases with relevant clinical implications. Genomic regions with confirmed and suggestive association with MS are active in immunological cell types and in particular Th1, Th17, T cytotoxic, B and NK cells. The same strategy applied to other complex traits highlights differences in the pathogenesis of diseases driven by different components of the immune system as well as of non-immune-mediated conditions. Future functional studies should aim to identify how MS-associated variants can modify the chromatin landscape and gene expression. Since these effects are likely to be cell specific, our data should be considered when planning these experiments.

MATERIALS AND METHODS

Defining associated regions

We collected disease-associated SNPs reported in the 'Catalog of Published Genome-Wide Association Studies' (http://www.genome.gov/gwastudies/) (31). Only associations reported from the GWAS with the largest sample size and replication cohort were included in our analysis (Supplementary Material, Table S5). We further analyzed MS data by manually extracting associated SNPs from the two largest and most recent GWAS performed to date (7, 8). In Sawcer *et al.*'s study, confirmed associated SNPs were defined as those that either were a replication of previous GWAS findings or had a discovery stage $P < 1 \times 10^{-4.5}$, a replication $P < 0.05$ and a P-combined $< 5 \times 10^{-7}$. SNPs suggestive of association were those having a discovery stage $P < 1 \times 10^{-4.5}$ and in which the same allele was overrepresented in cases compared with controls in both discovery and replication stages (8). In Patsopoulos *et al.*'s study, confirmed associated SNPs were defined as either those with P-values $< 5 \times 10^{-8}$ or previously identified associated SNPs that replicated. Suggestive SNPs were those with P-values ranging between 5×10^{-8} and 5×10^{-6} (7). Associated regions were defined as the genomic intervals of 100 kb centered on the associated SNP. The extent of overlap between regions associated with different diseases is shown in Supplementary Material, Table S6.

Estimating GWAS-DHSs overlap

All enrichment analyses were performed using the Genomic Hyperbrowser (http://hyperbrowser.uio.no/hb/) (32). In order to assess whether cell-specific DHSs fell more in the proximity of SNPs associated to a given disease than the general tendency of DHSs across all cell types to fall in the same proximity, we applied a three-track approach. A general DHSs track was created by aggregating all DHS sites across all cell types/tissues. Enrichment for a particular cell type was calculated as the ratio of cell-specific DHSs coverage to aggregated DHSs coverage in the proximity of disease-associated SNPs, divided by the corresponding ratio for the full genome. This is equivalent to testing whether cell-specific DHSs overlap with disease associated regions more than the general tendency of DHSs across all cell types to do this. The aggregated DHSs track contains overlapping regions (sites from different cell types that have overlapping genome coordinates), and in such cases the coverage for a given base pair were calculated as the count of regions covering the base pair. Instead of counting uniformly within such regions around SNPs, we used a kernel-based approach to assess chromatin accessibility in the proximity of SNPs. Conceptually, we slid a kernel along the DHSs track, computing a density estimate for each bp. We used a symmetric kernel where a position was assigned a weight inversely proportional to the distance of the position from the center of the kernel. This means that the cumulative weight up to a given distance (the integral) is inversely proportional to the logarithm of the distance. In order to avoid a too strong weighting on the immediate proximity of the kernel center, we used a slightly modified kernel where, for positions closer than 2 kbp from the kernel center, we still used a distance value of 2000 in the weight computations. We truncated the kernel at a maximum distance of 50 kb from the kernel center, since only a small proportion of SNPs in strong LD with GWAS SNPs ($r^2 = 1$) are located beyond this distance. This gave a kernel as shown in Supplementary Material, Figure S2. We then extracted the kernel density estimate at the SNP positions for both the cell type specific and the general (cross-cell type) DHSs track and normalized against the corresponding global averages for the same tracks. Since the density estimates are used only at the SNP positions, we only need to compute the kernel estimates with kernels centered on the SNPs. Total coverage in the proximity of disease-associated SNPs (as referred to above) was then defined as the sum of the kernel weighted DHSs densities for all SNPs. Since we are only interested in a ratio (cell-type specific to aggregate coverage), we no need to be concerned about the scale of the coverage values (as long as cell-type specific and aggregate coverage use the same scale). When computing DHSs coverage for the full genome, we do a plain count of bp coverage for the cell-type specific and aggregate DHSs tracks. Statistical hypothesis testing was performed using a permutation-based approach to compute P-values. We defined a null model for which the location of individual disease-associated SNPs were uniformly randomized within the chromosome arms they were located in, whereas DHSs were kept fixed. The kernel-based enrichment, as described above, was used as a test statistic in the hypothesis testing. This test statistic was calculated for the real data, as well as for several Monte Carlo samples from the null model. The P-value was calculated in the usual way, i.e. as the proportion of Monte Carlo samples being equal to or more extreme than the observed test statistic. The number of Monte Carlo samples needed to achieve sufficiently accurate P-values was automatically determined, individually for each cell type, based on the MCFDR sampling scheme (33).

Disease-associated regions are known to be preferentially located in the proximity of coding regions. To confirm that the significant disease–cell type relations are not a result of confounding effects from coding region proximity, we repeated the hypothesis testing of relations between MS-associated regions and cell types using a confounder handling approach where the proximity to coding regions is preserved in the null model. We denote a track of disease regions in the analysis as T, and a reference track of exon regions as R. Both tracks are of type segment (intervals of genome coordinates). We assume that segments in T have a certain distance relation to segments in R. We simulate new tracks for T from a null model denoted T1 where we condition on the empirical distances from each segment in T to its closest segment R. This means that the same number of segments in T and T1 has its midpoint inside a segment in R and the empirical distances between midpoints in segments in T to the nearest point in a segment in R is reproduced in T1. For midpoints that are inside segments in R, we sample new reference segments from R uniformly and then simulate a point inside the segment uniformly. When sampling new regions in the null model, some of these might end up overlapping each other. If this happens, we resample the whole track of segments to ensure that there is no intra-track overlap in the null model.

GRAIL analysis

The expression profile of the genes located within confirmed and suggestive MS-associated loci was analyzed using GRAIL and the methods have been previously described (http://www.broadinstitute.org/mpg/grail/) (9). In brief, for each input SNP (confirmed + suggestive) the software finds the furthest neighboring SNPs in the 3′ and 5′ direction in LD ($r^2 > 0.5$), and then proceeds in each direction to the nearest recombination hotspot. The expression of the genes located within these intervals is then tested in the Novartis Gene Expression Atlas, a database containing expression measurements for more than 30 000 probes across multiple tissues and cell types (10). The relatedness of each of these genes to every other human gene is assessed based on the correlation between expression vectors. Then, GRAIL estimates for each gene the presence of independent associated regions with at least one highly related gene. Finally, for each associated region the candidate gene is defined as the most connected within the network, and key cell types are identified as those that most strongly link the significant genes in each region (9).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

AUTHORS' CONTRIBUTIONS

Study concept and design: G.D., G.K.S., S.V.R.
Acquisition of data: G.D. and S.V.R.
Analysis and interpretation of data: G.D., L.H. and G.K.S.
Drafting of the manuscript: G.D.
Critical revision of the manuscript for important intellectual content: V.A.G.R., A.J.B.T., J.P., J.K., C.T.W., A.E.H., G.G., S.V.R.
Study supervision: S.V.R.

FUNDING

REFERENCES

1. Compston, A. and Coles, A. (2008) Multiple sclerosis. *Lancet*, **372**, 1502–1517.
2. Ramagopalan, S.V., Dobson, R., Meier, U.C. and Giovannoni, G. (2010) Multiple sclerosis: risk factors, prodromes, and potential causal pathways. *Lancet Neurol.*, **9**, 727–739.
3. Bielekova, B. and Becker, B.L. (2010) Monoclonal antibodies in MS: mechanisms of action. *Neurology*, **74**(Suppl 1), S31–40.
4. Kappos, L., Li, D., Calabresi, P.A., O'Connor, P., Bar-Or, A., Barkhof, F., Yin, M., Leppert, D., Glanzman, R., Tinbergen, J. *et al.* (2011) Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. *Lancet*, **378**, 1779–1787.
5. Gold, R., Kappos, L., Arnold, D.L., Bar-Or, A., Giovannoni, G., Selmaj, K., Tornatore, C., Sweetser, M.T., Yang, M., Sheikh, S.I. *et al.* (2012) Placebo-controlled phase 3 study of oral BG-12 for relapsing multiple sclerosis. *N. Engl. J. Med.*, **367**, 1098–1107.
6. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J. *et al.* (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science*, **337**, 1190–1195.
7. Patsopoulos, N.A., Esposito, F., Reischl, J., Lehr, S., Bauer, D., Heubach, J., Sandbrink, R., Pohl, C., Edan, G., Kappos, L. *et al.* (2011) Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann. Neurol.*, **70**, 897–912.
8. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L., Dilthey, A., Su, Z., Freeman, C., Hunt, S.E. *et al.* (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, **476**, 214–219.
9. Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D. and Daly, M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.*, **5**, e1000534.

10. Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

11. Kasper, L.H. and Shoemaker, J. (2010) Multiple sclerosis immunology: the healthy immune system vs the MS immune system. *Neurology*, **74**(Suppl 1), S2–8.

12. Ramagopalan, S.V. and Ebers, G.C. (2009) Multiple sclerosis: major histocompatibility complexity and antigen presentation. *Genome Med.*, **1**, 105.

13. Klein, L., Hinterberger, M., Wirnsberger, G. and Kyewski, B. (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat. Rev. Immunol.*, **9**, 833–844.

14. Babbe, H., Roers, A., Waisman, A., Lassmann, H., Goebels, N., Hohlfeld, R., Friese, M., Schroder, R., Deckert, M., Schmidt, S. *et al.* (2000) Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. *J. Exp. Med.*, **192**, 393–404.

15. Friese, M.A. and Fugger, L. (2009) Pathogenic CD8(+) T cells in multiple sclerosis. *Ann. Neurol.*, **66**, 132–141.

16. Skulina, C., Schmidt, S., Dornmair, K., Babbe, H., Roers, A., Rajewsky, K., Wekerle, H., Hohlfeld, R. and Goebels, N. (2004) Multiple sclerosis: brain-infiltrating CD8+ T cells persist as clonal expansions in the cerebrospinal fluid and blood. *Proc. Natl Acad. Sci. USA*, **101**, 2428–2433.

17. Kaur, G., Trowsdale, J. and Fugger, L. (2012) Natural killer cells and their receptors in multiple sclerosis. *Brain*, **136**, (Pt 9), 2657–2676.

18. Chanvillard, C., Jacolik, R.F., Infante-Duarte, C. and Nayak, R.C. (2013) The role of natural killer cells in multiple sclerosis and their therapeutic implications. *Front. Immunol.*, **4**, 63.

19. Disanto, G., Sandve, G.K., Berlanga-Taylor, A.J., Morahan, J.M., Dobson, R., Giovannoni, G. and Ramagopalan, S.V. (2012) Genomic regions associated with multiple sclerosis. *PLoS One*, **7**, e32281.

20. Disanto, G., Morahan, J.M., Barnett, M.H., Giovannoni, G. and Ramagopalan, S.V. (2012) The evidence for a role of B cells in multiple sclerosis. *Neurology*, **78**, 823–832.

21. Freedman, M.S., Thompson, E.J., Deisenhammer, F., Giovannoni, G., Grimsley, G., Keir, G., Ohman, S., Racke, M.K., Sharief, M., Sindic, C.J. *et al.* (2005) Recommended standard of cerebrospinal fluid analysis in the diagnosis of multiple sclerosis: a consensus statement. *Arch. Neurol.*, **62**, 865–870.

22. Brettschneider, J., Czerwoniak, A., Senel, M., Fang, L., Kassubek, J., Pinkhardt, E., Lauda, F., Kapfer, T., Jesse, S., Lehmensiek, V. *et al.* (2010) The chemokine CXCL13 is a prognostic marker in clinically isolated syndrome (CIS). *PLoS One*, **5**, e11986.

23. Brettschneider, J., Tumani, H., Kiechle, U., Muche, R., Richards, G., Lehmensiek, V., Ludolph, A.C. and Otto, M. (2009) IgG antibodies against measles, rubella, and varicella zoster virus predict conversion to multiple sclerosis in clinically isolated syndrome. *PLoS One*, **4**, e7638.

24. Howell, O.W., Reeves, C.A., Nicholas, R., Carassiti, D., Radotra, B., Gentleman, S.M., Serafini, B., Aloisi, F., Roncaroli, F., Magliozzi, R. *et al.* (2011) Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain*, **134**, 2755–2771.

25. Joseph, F.G., Hirst, C.L., Pickersgill, T.P., Ben-Shlomo, Y., Robertson, N.P. and Scolding, N.J. (2009) CSF oligoclonal band status informs prognosis in multiple sclerosis: a case control study of 100 patients. *J. Neurol. Neurosurg. Psychiatry*, **80**, 292–296.

26. Khademi, M., Kockum, I., Andersson, M.L., Iacobaeus, E., Brundin, L., Sellebjerg, F., Hillert, J., Piehl, F. and Olsson, T. (2011) Cerebrospinal fluid CXCL13 in multiple sclerosis: a suggestive prognostic marker for the disease course. *Mult. Scler.*, **17**, 335–343.

27. Hauser, S.L., Waubant, E., Arnold, D.L., Vollmer, T., Antel, J., Fox, R.J., Bar-Or, A., Panzara, M., Sarkar, N., Agarwal, S. *et al.* (2008) B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. *N. Engl. J. Med.*, **358**, 676–688.

28. Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A. *et al.* (2012) Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, **491**, 119–124.

29. Modlin, R.L. (2010) The innate immune response in leprosy. *Curr. Opin. Immunol.*, **22**, 48–54.

30. Piper, M., Beneyto, M., Burne, T.H., Eyles, D.W., Lewis, D.A. and McGrath, J.J. (2012) The neurodevelopmental hypothesis of schizophrenia: convergent clues from epidemiology and neuropathology. *Psychiatr. Clin. North. Am.*, **35**, 571–584.

31. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.

32. Sandve, G.K., Gundersen, S., Johansen, M., Glad, I.K., Gunathasan, K., Holden, L., Holden, M., Liestol, K., Nygard, S., Nygaard, V. *et al.* (2013) The Genomic HyperBrowser: an analysis web server for genome-scale data. *Nucleic Acids Res.*, **41**, W133–W141.

33. Sandve, G.K., Ferkingstad, E. and Nygard, S. (2011) Sequential Monte Carlo multiple testing. *Bioinformatics*, **27**, 3235–3241.