The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis

Kaushal S. Gandhi1, Fiona C. McKay1, Mathew Cox2, Carlos Riveros2, Nicola Armstrong1, Robert N. Heard1, Steve Vucic1, David W. Williams2, Jim Stankovich3, Matthew Brown4, Patrick Danoy4, Graeme J. Stewart1, Simon Broadley5, Pablo Moscato2, Jeannette Lechner-Scott2, Rodney J. Scott2, David R. Booth1,* and ANZgene Multiple Sclerosis Genetics Consortium

1Westmead Millennium Institute, University of Sydney, Sydney, Darcy Rd, New South Wales 2145, Australia, 2Hunter Medical Research Institute, The University of Newcastle, Newcastle, New South Wales 2310, Australia, 3Menzies Research Institute, University of Tasmania, Hobart, Tasmania 7000, Australia, 4Diamantina Institute, University of Queensland, Brisbane, Queensland 4102, Australia and 5School of Medicine, Griffith University, Gold Coast, Queensland 4222, Australia

Received November 30, 2009; Revised February 16, 2010; Accepted February 24, 2010

Multiple sclerosis (MS) is an autoimmune disease with a genetic component, caused at least in part by aberrant lymphocyte activity. The whole blood mRNA transcriptome was measured for 99 untreated MS patients: 43 primary progressive MS, 20 secondary progressive MS, 36 relapsing remitting MS and 45 age-matched healthy controls. The ANZgene Multiple Sclerosis Genetics Consortium genotyped more than 300,000 SNPs for 115 of these samples. Transcription from genes on translational regulation, oxidative phosphorylation, immune synapse and antigen presentation pathways was markedly increased in all forms of MS. Expression of genes tagging T cells was also upregulated \((P < 10^{-12})\) in MS. A T cell gene signature predicts disease state with a concordance index of 0.79 with age and gender as co-variables, but the signature is not associated with clinical course or disability. The ANZgene genome wide association screen identified two novel regions with genome wide significance: one encoding the T cell co-stimulatory molecule, CD40; the other a region on chromosome 12q13-14. The CD40 haplotype associated with increased MS susceptibility has decreased gene expression in MS \((P < 0.0007)\). The second MS susceptibility region includes 17 genes on 12q13-14 in tight linkage disequilibrium. Of these, only 13 are expressed in leukocytes, and of these the expression of one, FAM119B, is much lower in the susceptibility haplotype \((P < 10^{-14})\). Overall, these data indicate dysregulation of T cells can be detected in the whole blood of untreated MS patients, and supports targeting of activated T cells in therapy for all forms of MS.

INTRODUCTION

Multiple sclerosis (MS) affects around 2.5 million people worldwide. It is an immune cell-mediated disease characterized by autoimmune destruction of the myelin sheath of the central nervous system (CNS). Extensive studies of population genetics have shown a major genetic component to the disease (1). Since 2007, there have been at least four genome wide association screens (GWAS) conducted for MS (2–5), and 14 genes have now been identified with genome-wide...
significance and international replication (Supplementary Material, Table S1).

The gene haplotypes contain genetic variants which potentially affect regulation of transcription and mRNA processing. This is the first level at which genetic variance can produce functional differences for genes, and the level where there is the least noise from other processes (6). SNP association with expression has been demonstrated in genomic and transcriptomic studies from combining HapMap genotype data with measurements of gene expression from lymphocytes (7). SNPs associated with diseases have also been associated with differences in gene expression (8). Such studies have been particularly successful where the gene expression is measured in tissues associated with pathogenesis. Emilsson et al. (9) compared gene expression in blood and adipose tissue and found expression for obesity genes was genotype dependent only in the latter.

In MS, nearly all genes so far identified as associated with disease are primarily expressed in immune cells, especially involving the dendritic cell/ T cell axis (Supplementary Material, Table S1), indicating the interaction of these leukocytes is critical in pathogenesis. The autoimmune attack on the myelin sheath is orchestrated by leukocytes which must come, at least initially, from the peripheral circulation (10). Current successful drug therapies target leukocytes. In particular, the clinical effect of the monoclonal antibody to α4 integrin, natalizumab, is thought to be due to its prevention of T cell migration across the blood brain barrier from the peripheral circulation. The mRNA of whole blood is largely of leukocyte origin, and differences in their gene expression between MS and controls are likely to reveal pathogenically significant processes. Correlation of MS-associated SNP genotype with gene expression in these cells may reveal the functional basis for such associations.

There are three main clinical subtypes of MS: relapsing remitting MS (RRMS), the most common, in which there is recovery between relapses of CNS dysfunction; secondary progressive MS (SPMS), where there is an accumulation of disability following a period of relapses and primary progressive MS (PPMS), where disease progression occurs without remission. Patients with PPMS are considered less responsive to immunomodulatory therapies (10,11), and they are usually ineligible for funding support for such therapies in national health schemes. It is difficult to measure drug response in PPMS. The usual clinical parameters measured in clinical trials are relapse rates (patients with PPMS do not have any by definition), gadolinium enhanced lesions (in PPMS these are rare) or T2 lesion load (less in PPMS, and fewer lesions). For these reasons, patients with PPMS are usually excluded from trials aimed at demonstrating drug efficacy. Subsequently they are then excluded from using the new therapies in PPMS patients. The known genes for susceptibility are not known to be different between clinical phenotypes. Gene expression differences between MS clinical subtypes in large cohorts have not been reported. Characterization of gene expression in whole blood may lead to identification of useful clinical biomarkers for MS, including PPMS.

We measured mRNA expression for all known genes (48 000 probes) in whole blood from 144 individuals, 99 with MS (43 PPMS, 36 RRMS and 20 SPMS). Most of these individuals were also genotyped for 317 000 SNPs tagging 80–90% of the common haplotypes, as part of the recently published ANZgene GWAS (5). Analysis of gene expression and genotype in these individuals has indicated that the dominant genetic association with T cell regulation is matched by the dominant pattern of dysregulation of T cell gene expression in MS and its clinical subtypes.

### RESULTS

#### Gene and gene pathway dysregulation

To test whether there were differences in gene expression in whole blood cells between people with MS and healthy controls (Con), we used PAXgene tubes to collect whole blood from age- and sex-matched patients and controls (Table 1). PAXgene collections prevent degradation of labile mRNAs by inhibiting RNases at the moment of phlebotomy (12). Blood collections were taken in a 4 h window (9 am to 1 pm) to minimize the effect of diurnal variation (13); patients and controls were not using immunomodulatory drugs, and had not for at least 3 months. Over 19 000 genes were expressed in the blood cells. By SAM analysis (14), many genes were differently expressed between MS and controls, between each MS type and controls, but few differences were detected in the within MS comparisons (Table 2, Supplementary Material, Table S2). This is partly due to increased statistical power in MS/Con compared with subtype comparison, but there were as many samples in the PPMS/Con, RRMS/Con and PPMS/RRMS comparisons, with virtually no significant differences detected in the latter compared with the former two. This indicates there is gene dysregulation in MS, and that it is common to the MS subtypes. There was also substantial overlap in the dysregulated genes between the MS subtypes (Supplementary Material, Fig. S1). It is striking

<table>
<thead>
<tr>
<th>Sample</th>
<th>n</th>
<th>M/F</th>
<th>Age</th>
<th>Age at diagnosis</th>
<th>Duration</th>
<th>EDSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthy</td>
<td>45</td>
<td>16:29</td>
<td>48.5 (23–77)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>99</td>
<td>33:66</td>
<td>54 (29–87)</td>
<td>35.5 (16–59)</td>
<td>16.5 (1–54)</td>
<td>4.5 (0–9)</td>
</tr>
<tr>
<td>PPMS</td>
<td>43</td>
<td>21:22</td>
<td>55.3 (30–87)</td>
<td>42 (21–58)</td>
<td>13 (5–31)</td>
<td>5.6 (2–8)</td>
</tr>
<tr>
<td>RRMS</td>
<td>36</td>
<td>7:29</td>
<td>48.5 (29–65)</td>
<td>33 (19–56)</td>
<td>15 (1–36)</td>
<td>2.7 (0–6.5)</td>
</tr>
<tr>
<td>SPMS</td>
<td>20</td>
<td>5:15</td>
<td>57.5 (34–73)</td>
<td>35.5 (16–59)</td>
<td>22 (2–54)</td>
<td>6.5 (4–9)</td>
</tr>
</tbody>
</table>

M/F, male/female; EDSS, expanded disability scale; mean and range are shown.

### Table 2. Number of genes dysregulated in MS and its clinical types

<table>
<thead>
<tr>
<th>Comparison (n v n)</th>
<th>Up</th>
<th>Down</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSvCon (99 v 45)</td>
<td>552</td>
<td>344</td>
<td>10%</td>
</tr>
<tr>
<td>PPMSvCon (43 v 45)</td>
<td>250</td>
<td>334</td>
<td>10%</td>
</tr>
<tr>
<td>RRMSvCon (36 v 45)</td>
<td>243</td>
<td>37</td>
<td>10%</td>
</tr>
<tr>
<td>SPMSvCon (20 v 45)</td>
<td>92</td>
<td>9</td>
<td>10%</td>
</tr>
<tr>
<td>PPMSvRRMS (43 v 36)</td>
<td>4</td>
<td>1</td>
<td>10%</td>
</tr>
</tbody>
</table>

Numbers of individuals compared are in parenthesis.

FDR, false discovery rate derived using SAM (Tusher et al. (14)). No differences were found between SPMS and RRMS or PPMS.
that there are more genes underexpressed in PPMS compared with controls, than for the other two MS subtypes. This may be due, at least partially, to the increased statistical power of this comparison.

We then tested the gene ontology pathways represented by the dysregulated gene sets (Fig. 1). There was a striking over-representation of genes regulating translation and oxidative phosphorylation. If a less stringent false discovery rate (FDR) is used, which then includes a much larger number of genes, this dysregulation can be seen to be consistent across even more of the genes in these two pathways, such that nearly all genes on the dysregulated list are overexpressed, in MS and its clinical subtypes. mRNA collected by PAX tubes is mainly from leukocytes and red blood cells. No genes tagging the latter were differentially expressed (see next section). This suggests that leukocytes are more active in MS, producing more energy (by oxidative phosphorylation) and protein (translation): 68 of 74 ($P < 10^{-13}$ by sign test) dysregulated genes from these linked pathways were overexpressed in MS (Fig. 2). Of the other over-represented gene ontology pathways, virtually all are from immune gene ontologies (Fig. 1, Supplementary Material, Table S3), especially in RRMS, but also from PPMS and SPMS. SPMS and RRMS shared dysregulated genes and gene ontologies.

Although there was little difference in gene expression between the MS clinical subtypes, it is striking that two genes are under-expressed in RRMS compared with PPMS, with an FDR set to zero. These genes are insulin-growth factor binding protein 7 (IGFBP7) and F-box only protein 33 (FBXO33). The Bromodomain PHD finger transcription factor (BDFT) is upregulated in RRMS compared with PPMS, also with FDR of zero.

**Cell subset gene expression**

Genes specific for immune cell subsets have been identified by others (15,16) and are described on the immunological genome website (17). Genes in the dysregulated gene set which are predominantly expressed in one cell subset were identified from this data, and the percentage tagging the cell subsets from up- and downregulated genes is shown in Figure 3. In a second approach, we compared expression of all cell tagging genes (not just those in the dysregulated gene set) identified by Abbas et al. (15) (Supplementary Material, Table S4). In both approaches, mRNAs from lymphocyte-specific genes were generally over-expressed in MS, while myeloid cells were relatively underexpressed. Of the lymphocyte cell subsets, T cell tagging genes were mainly upregulated ($P < 4.3 \times 10^{-12}$, Fig. 3), whereas B and NK cells were similar to controls. The T cell gene over-expression, such as the over-representation of T cell ontologies in the pathway analysis, was similar in each clinical subtype of MS. Of the myeloid cell subsets, neutrophil tagging genes were over-represented in the set of downregulated genes ($P < 1.8 \times 10^{-16}$, Fig. 3), and mainly downregulated in MS (Supplementary Material, Table S3), also with a
similar pattern being seen in each MS subtype. In whole blood, the two dominant sources of mRNA are T cells and neutrophils. The reduction of relative expression of the latter in these normalized data may reflect the aberrant presence of the former. For 30 of the PPMS samples used in this study, the proportion of neutrophils and lymphocytes was not different from controls (Supplementary Material, Fig. S2), even though the cell tag expression was. Abnormalities in the T cell/neutrophil ratio have not been identified in MS in general. This suggests the T cell/neutrophil mRNA difference is due to a change in cell state, not numbers. Enhancement of T cell activity seems likely, rather than reduction in neutrophil activity, because the vast majority of dysregulated genes and gene ontologies are related to T cell functions, and their role in pathogenesis is well-documented (10). Studies based on CD3 cells alone have also indicated excess transcriptional and immune gene activity in MS (18). Genes from the translation regulation pathway are upregulated in T cells in blood of healthy controls (19), and in Th0 rather than terminally differentiated Th1 or Th2 cells (16).

Within the CD3 T cells, genes expressed by CD4 and CD8 activated and inactivated cells have recently been characterized (20). We found no evidence for an imbalance between CD4 and CD8 cell gene expression in MS—the dysregulated genes are evenly represented between these T cell subsets. However, genes upregulated in MS are mainly from the list of genes upregulated on activation of CD4/8 cells; and genes downregulated in MS are mainly from the genes downregulated on T cell activation: 43/54 (P = 1.4 x 10^-5 by sign test) of the MS upregulated genes expressed in CD8s are upregulated in activated CD8s; whereas 16/23 genes [not significant (NS)] in the MS downregulated genes from CD8s are downregulated in activated CD8s, and 18/25 (P = 0.029) of the downregulated CD4 genes are downregulated in activated CD4s. Genes predominantly expressed in regulatory T cells compared with other T cells have also been recently identified (21). Most of these genes are in the list of genes underexpressed in MS (13 versus 3 in the list of overexpressed genes, P = 0.021), suggesting this T cell subset is under-active in MS whole blood. Further refinement of the T cell type contributing to the T cell activation and other signatures will be possible from new databases currently being established (17).

**T cell signature in MS**

Given this evidence of enhanced expression of T cell activation genes in all types of MS, and the importance of these cells in MS pathogenesis, we tested how well the genes upregulated in MS and predominantly expressed in T cells (113 genes) predicted disease state. We used a Support Vector Machine (SVM) with radial function kernel in a cross-validation framework to identify the set of genes with the highest specificity and sensitivity to predict MS (Supplementary Methods). This was obtained from the 86 most dysregulated genes. From a heat map (Fig. 4), this T cell signature can be seen in most MS cases, but also in some controls. Some MS cases do not have it. The trained SVM and its predictions were used in a logistic regression singly, with age, with gender and with age and gender. The concordance index was 0.705, 0.727, 0.749 and 0.797, respectively.
part of the ANZgene GWA for MS susceptibility (5). That expression study were also genotyped for 317 000 SNPs as
One hundred and fifteen of the samples used in this gene GWAS and gene expression consilience: CD40 and
score, suggesting the T cell signature may be innate rather
females. The degree of disease progression did not affect the slightly less likely to be MS in older individuals and in
(Supplementary Material, Fig. S3). The T cell signature was
subset observed and expected in either up or down regulated genes. Only T
cells, dendritic cells (DCs), neutrophils, B-cells, natural killer (NK)-cells or
other myeloid cells determined from the human immune cell module of the
immunological genome project (Heng and Painter (17)). A much higher percen-
tage of the genes upregulated in MS are predominantly expressed in T-cells, and
immunological genome project (Heng and Painter (17)). A much higher percen-
tage of the genes upregulated in MS are predominantly expressed in T-cells, and
versely most downregulated genes are mainly expressed in neutrophils. 
$P$-values were determined by comparing the number of genes from each cell subset observed and expected in either up or down regulated genes. Only T cell ($P < 4.3 \times 10^{-12}$) and neutrophil ($P < 1.8 \times 10^{-16}$) genes were signifi-
cantly different between up- and downregulated gene lists (Chi-squared test). Other myeloid cells: macrophages, mast cells, eosinophils or basophils.

**GWAS and gene expression consilience: CD40 and FAM119B**

One hundred and fifteen of the samples used in this gene expression study were also genotyped for 317 000 SNPs as part of the ANZgene GWA for MS susceptibility (5). That study identified two novel associations with MS at genome-level significance. The minor allele of CD40 SNP rs6074022 was more common in MS. CD40 is a co-stimulatory molecule, predominantly expressed on antigen presenting cells, and engagement with CD40L contributes to T cell activation and B cell immunoglobulin switching. Surprisingly, the MS-susceptibility haplotype is less expressed (Fig. 5, $P < 0.0009$), especially so in MS ($P < 0.6$ in healthy controls, $P < 0.0007$ in MS samples). This was confirmed by quantitative RT–PCR (Supplementary Material, Fig. S4). Of the 30 SNPs genotyped from the genomic region of CD40 in the ANZgene GWAS, SNP rs6074022, which has the strongest association with MS, also has the strongest association with CD40 expression (Supplementary Material, Table S5). Two SNPs in LD with each other but not with rs6074022 were associated with CD40 expression. These SNPs were not associated with MS. This may prove a useful tool in determining microenvironments important in pathogenesis. Microen-
vironments in which these latter SNPs are not associated with expression, but where rs6074022 is, are more likely to be of pathogenic significance in MS. It is interesting that some data points appear as outliers, and these may turn out to be due to rare variants with large effects tagged by the common SNP rs6074022.

The second association ($P < 10^{-10}$ for SNP rs703842) from the ANZgene GWAS identified a cluster of genes on chromosome 12q13-14 in strong linkage disequilibrium as potentially associated with MS. Fine-mapping will need to be employed to test association independent of linkage. An adjunct approach is to look for genes which are good candidates for association on the basis of their function. Most of the MS-associated genes so far detected are expressed predominantly in leukocytes, consistent with immune dysregulation contributing to MS susceptibility. Genes from this cluster that were identified as expressed in the whole blood leukocytes are shown in Figure 6A. Of the 17 genes in this genomic cluster, 13 were expressed. We then tested if the MS-associated SNP was also associated with gene expression. It was, and most strikingly with the gene FAM119B (Fig. 6B and C) ($P < 10^{-14}$). The MS susceptibility allele is the low-expressor of FAM119B. This was confirmed by quantiative RT–PCR (Supplementary Material, Fig. S5). Expression of other genes was also associated with this SNP, but to a lesser extent. This points to co-ordinated regulation of this region. This may be due, for example, to reduced methylat-
on or increased histone deacetylization of a common regulatory region which is tagged by the ‘G’ variant of SNP rs703842. The other 21 SNPs genotyped from this region in the ANZgene GWAS were tested for association with expression (Supplementary Material, Table S6). The SNP rs703842 was most associated, and the other SNPs were associated only in proportion to their linkage with rs703842. Gene expression analysis supports the polymorphisms of FAM119B as the most likely of the 12q13-14 cluster to be the basis for the increased MS suscepti-
bility of SNP rs703842. However, it should be noted that CYP27B1, which is a very strong candidate gene from this region [5], is mainly expressed in renal tissue, and scarcely if at all in whole blood. It catalyses production of the active form of vitamin D, which then plays a pivotal role in immune regu-
lation. The genome wide association study of global gene expression analysis by CD40 and FAM119B.

**DISCUSSION**

These data indicate that whole blood gene expression in MS differs from healthy controls. T cell genes and pathways are over-represented in the dysregulated gene sets, and T cell
specific genes are over-expressed. This is consistent with, and extends, previous studies using smaller sample sizes and different sources of blood and microarray platforms (23,24). It is also consistent with the postulated pathogenesis of MS (10), and the success of therapies which reduce T cell activation gene expression in peripheral blood (25,26).

Aberrant T cell activation is also consistent with the genetic associations identified in the ANZgene and other GWAS. Of
the 14 MS genes so far identified, most are predominantly expressed on cells from the antigen presenting cell (APC): T cell complex (Supplementary Material, Table S1). Our transcriptomic data suggest excessive T cell activity in whole blood is a hallmark of MS. This pattern of expression could flag aberrant regulatory processes which are constitutive (genetically programmed) and lead to predisposition to disease; or could be as a result of inflammatory and other factors accompanying the disease state, which then cause the T cell activation signature. Work by others (27) suggests the former is likely: unaffected siblings of probands with autoimmune diseases have the same whole blood gene expression patterns. Also, there was no correlation of T cell signature with disability or clinical course, which would be expected if the latter was true. No correlation with disability and clinical course may also be useful in identifying dysregulated genes between MS and controls.

Therapies aimed at reducing T cell activation and trafficking across the blood-brain barrier have proved effective in RRMS and SPMS, but have largely not been tested in PPMS. Here we have also compared gene expression in the MS clinical subgroups, and found they share evidence of excessive T cell activation, but that this is most emphasized in RRMS and SPMS. These data suggest there is immune dysregulation in PPMS that may be responsive to the same therapies as RRMS and SPMS.

There is evidence of altered regulation of T cells from immune cell phenotype and functional studies in MS and other autoimmune diseases. Tr1s (29,30) and Tregs (31) cells, which regulate T cell activity, have been shown to have reduced suppressive activity but not reduced numbers in MS. A population of activated T cells, CD4+CD25HiCD127Hi, which are more resistant to suppression, have been identified as more active in MS (32). Cells sorted on a comprehensive array of CD markers have identified CD8lowCD4+ cells, particularly NK cells, as under-represented in MS (33). Such cells must have specific transcriptional programs which, when characterized, may allow their identification in whole blood. Comparison of transcriptomes from such cell subsets may also be useful in identifying dysregulated genes between MS and controls.

The reduced production of CD40 mRNA in the MS susceptibility allele identified here is intriguing. CD40–CD40L interactions play a crucial role in T cell activation, and contribute to germinal centre formation, memory B cell development, Ig isotype switching and affinity maturation (34). Higher expression might be expected to correlate with increases in signalling, so more T cell and B cell activation, proliferation and maturation. However, mutations or deletions in CD40, its ligand or downstream signalling molecules, can result in the immune deficiency disease in humans, hyper IgM syndrome, which is often accompanied by autoimmune disease (35). The MS susceptibility haplotype is associated with lower protein expression in healthy controls (36), but in other autoimmune diseases associated with the CD40 genotype, Graves (36) and rheumatoid arthritis (37), the higher expressing haplotype (tagged by SNPs rs1883832 and rs4810485, respectively, both of which are in complete linkage with the MS SNP) is associated with disease.

The dominant source of CD40 mRNA in whole blood is B cells (16). Most of the CD40–CD40L functions detailed above would be expected to increase the risk of autoimmunity. However, there are a number of examples of CD40 function...
where higher expression of CD40 in B cells may be protective. First, antigen specific B cells are important for reversing EAE, and it has been speculated that CD40 signalling may regulate suppressive B cells (38). Secondly, control of class switching via CD40 is crucial in immune regulation, and may be under strict regulation by tissues, especially at sites of immune privilege, such as the CNS (39). If MS B cells are less likely to class switch due to lower expression of CD40, the outcome may be suboptimal regulation of this important function. Thirdly, variable ligand binding can have unpredictable effects. For example, reduced ligand binding to CD40 can stimulate B cells to secrete more immunoglobulin than higher levels of ligand binding (40). Fourthly, the lower expression of the MS-susceptibility haplotype identified here may not be the effect of the haplotype that is pathogenic. Haplotype-specific expression of CD40, in the many cell types that express it, and in particular microenvironments, needs to be investigated, to assess conditions where haplotype differences are maximal, and so more likely to contribute to pathogenesis (6,41). The impact of anti-CD40 therapies would be particularly difficult to predict until the basis for the haplotypic association with MS is better understood.

The linkage group on chromosome 12 identified as associated with MS contains 17 genes, including such promising candidates as CYP27B1, the activator of vitamin D (5). However, of the genes in this LD group expressed in lymphocytes, expression of FAM119B is clearly the most associated with the rs703842 genotype. This gene is of unknown function, and so could not be evaluated from a candidate gene perspective. Further sequencing and fine mapping of this region is needed to identify the gene/s contributing to the association of this SNP with MS.

Determining the basis for pathogenic significance of SNPs associated with MS will be of advantage in studying autoimmunity in general: several are also associated with other diseases such as type 1 diabetes (T1D) (CD25, CD127, CD226 and CLEC16A (42) celiac disease (43) and rheumatoid arthritis (37), although sometimes the susceptibility haplotype for one disease is the protective haplotype for the other (e.g. CD25 for MS and T1D) (44). At a clinical level, this is supported by the co-occurrence of autoimmune disease in kindreds and in individuals which occurs more often than would be expected by chance (45). These suggest that some genes have a common effect on autoimmunity, whereas others may be disease specific, each acting in concert with MHC genotype and environmental factors.

The MS gene expression and genotyping data presented here suggest that discovery driven approaches will be useful in producing a further consilience of genomics, transcriptomics and immune cell phenotype data. This will support hypothesis driven studies testing genetic control of T cell regulation in vitro, and functional studies of T cells ex vivo, and should empower studies identifying the immune cell dysregulation(s) which lead to the development of MS.

**MATERIALS AND METHODS**

**Patients and sample collection for gene expression**

Peripheral blood samples were collected from 43 PPMS, 36 RRMS, 20 SPMS and 45 HCs in PAXgene™ Blood RNA tubes (Qiagen, Germany). These patients were recruited from Sydney, Newcastle (NSW) and Gold Coast (QLD). The demographics of the participants are shown in Table 1. Samples were collected between 9 am and 1 pm. Patients had no immunomodulatory therapy in the last 3 months. MS diagnosis was according to Poser and McDonald criteria (28,46) as previously described (47). Controls were laboratory personnel matched for age, gender, ethnicity and not on any immunomodulatory medications. Human Research Ethics Committees from the Sydney West Area Health Service, The John Hunter Hospital and the Gold Coast Hospital approved this study. Written informed consent was given by all participants.

**RNA extraction and microarray**

All the steps were followed according to the manufacturer’s instruction unless otherwise stated. Total RNA was extracted from the whole blood in PAXgene tubes using PAXgene™ Blood RNA kit (Qiagen, Germany). Total RNA quality was assessed and its concentration was measured using the Agilent RNA 6000 series II Nano kit (Agilent Technologies, CA, USA). Each sample had an RNA Integrity Number (RIN) score of >7.250 ng total RNA from each sample was biotinylated and amplified using Illumina® TotalPrep RNA Amplification Kit (Ambion, TX, USA). Total RNA was reverse transcribed to synthesize the first strand of cDNA, followed by a second-strand synthesis. Double-stranded cDNA was then transcribed and amplified in vitro to synthesize biotin labelled complementary mRNA (cRNA). The cRNA yield was measured at 260 nm using the NanoDrop 1000A Spectrophotometer. Seven hundred and fifty nanogram of cRNA sample was hybridized on a human HT-12 expression beadchip (Illumina, Inc., CA, USA) profiling 48 804 transcripts per sample. The chips were stained with streptavidin-Cye3 conjugate and scanned using an Illumina BeadArray Reader (Illumina, Inc.). All microarray data have been submitted to the Gene Expression Omnibus.

**Differential expression analysis**

The average signal intensity for each gene was measured using Beadstudio v3. The sample signals were normalized with cubic spline in order to minimize variation due to non-biological factors. Around 19 000 genes were selected for differential expression analysis with detection P-value less than 0.01 for at least one of all samples. The detection P-value measures the probability of observing signal without specific probe-target hybridization. The differentially expressed genes between MS and HC and between different types of MS were identified using significance analysis of microarray (SAM) (14). Using a FDR of 5%, 312 genes were identified as upregulated in MS. Of these, 167 were expressed more in one leukocyte subset than others (17), and of these 113 were expressed predominantly by T cells (Fig. 3). From these, the T cell signature was identified using a SVM with radial basis function kernel. The optimal number of genes required in the signature was identified using a stepwise forward filter approach. One hundred iterations of 5-fold cross validation were performed and the
average sensitivity and specificity recorded. In order to determine the concordance index for the T cell gene signature, the trained SVM and its predictions were used in a logistic regression. The concordance index was calculated for the following logistic regressions: including only the SVM predictions as an explanatory variable (‘Signature’); SVM predictions + Age, SVM predictions + Sex, SVM predictions + Sex + Age. This was repeated 500 times (100 iterations of 5-fold CV) and the results summarized.

Pathway analysis

The differentially expressed genes from the SAM analysis were further analysed to identify the biological networks using the GeneGo Maps module of Metacore (GeneGo, MI, USA). Metacore conducts functional analysis in the form of network pathways based on a manually curated database of human protein–protein, protein–DNA and protein–compound interactions, metabolic and signalling pathways and the effects of bioactive molecules in gene expression.

Real-time RT–PCR

Whole blood mRNA from 28 MS and 20 HC samples genotyped for the CD40 SNP rs6074022, and 20 MS and 13 HC for the FAM119B SNP rs703842 were further used for the QRT–PCR validation of the microarray gene expression measurements. Forward and reverse primers were designed for CD40 (Fwd: 5′-TCCTCTTGTTGCTGTCTTT-3′; Rev: 5′-AGGAAGATCGTGGGAAAT-3′) and FAM119B (Fwd: 5′-CTGTGGGACGCGGCCC-3′; Rev: 5′-CCTCGGAAATCCA-CATTGG-3′) using the Primer3 software (48). Semi-quantitative RT–PCR was used to determine relative CD40 and FAM119B mRNA expression using SYBR green (Applied Biosystems) on a Corbett Rotor-gene 3000 (Corbett, Australia). The expression levels of CD40 and FAM119B mRNA were normalized to median expression of the housekeeping gene GAPDH, as described by Livak and Schmittgen (49).

ACKNOWLEDGEMENTS

We would like to thank people with MS in Australia and New Zealand for supporting this research. We are grateful to Mr Jeremy Wright and Ms Christine Remediakis from MS Research Australia for expediting this research.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by Multiple Sclerosis Research Australia and the Australian Research Council under the Linkage Projects Scheme (LP0776744). K.S.G. is supported by an Australian Research Council Biogen Idea PhD award, M.B.C. is supported by a grant from the John Hunter Hospital Charitable Trust Fund and a special grant from Macquarie Bank.

REFERENCES


