The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease

Grant P. Parnella, Prudence N. Gatta, Malgorzata Krupa, Dorothee Nickles, Fiona C. McKay, Stephen D. Schibeci, Marcel Batten, Sergio Baranzini, Andrew Henderson, Michael Barnett, Mark Slee, Steve Vucic, Graeme J. Stewart, David R. Booth

Institute for Immunology and Allergy Research, Westmead Millennium Institute University of Sydney, Sydney, New South Wales 2145, Australia
School of Medicine, Flinders University of South Australia, South Australia 5042, Australia
Department of Neurology, University of California at San Francisco, CA 94158, USA
Garvan Institute of Medical Research, Darlinghurst, Sydney, NSW 2010, Australia
Brain and Mind Research Institute, University of Sydney, Sydney, NSW 2050 Australia
Westmead Clinical School, University of Sydney, Westmead Hospital, Sydney, New South Wales 2145, Australia

Received 2 December 2013; accepted with revision 7 January 2014
Available online 15 January 2014

Abstract We have identified a marked over-representation of transcription factors controlling differentiation of T, B, myeloid and NK cells among the 110 MS genes now known to be associated with multiple sclerosis (MS). To test if the expression of these genes might define molecular subtypes of MS, we interrogated their expression in blood in three independent cohorts of untreated MS (from Sydney and Adelaide) or clinically isolated syndrome (CIS, from San Francisco) patients. Expression of the transcription factors (TF) controlling T and NK cell differentiation, EOMES, TBX21 and other TFs was significantly lower in MS/CIS compared to healthy controls in all three cohorts. Expression was tightly correlated between these TFs, with other T/NK cell TFs, and to another downregulated gene, CCL5. Expression was stable over time, but did not predict disease phenotype. Optimal response to therapy might be indicated by normalization of expression of these genes in blood.

© 2014 The Authors. Published by Elsevier Inc. All rights reserved.

KEYWORDS
Transcription factors; Multiple sclerosis; EOMES; TBX21; Gene expression

☆ This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.
* Corresponding author. Fax: +61 298913889.
E-mail address: david.booth@sydney.edu.au (D.R. Booth).
1. Introduction

Genome-wide association studies (GWAS) and the internationally collaborative ImmunoChip experiment involving 10 autoimmune diseases including Multiple Sclerosis (MS) have recently identified hundreds of genetic variants that confer increased risk to various autoimmune diseases. MS is a common, partially heritable autoimmune disease, characterized by plaques of sclerotic tissue in the central nervous system. Over 100 genetic variants have now been identified, with many shared with other autoimmune conditions [1,2]. Genes from most of the shared and other loci are predominantly expressed in leukocytes, and most of the variants are in regulatory regions of the DNA [1,3], indicating that regulation of genes in leukocytes underpins MS susceptibility.

Transcription factors are master regulators of the transcriptome, controlling lineage differentiation and cell survival. In this study we identified an over-representation of transcription factor genes in the recently published list of 110 non-MHC MS susceptibility loci. Several other lines of evidence suggest that dysregulation of gene expression in leukocytes, including in the peripheral circulation, may drive MS. Histological studies have shown that leukocytes from the peripheral circulation infiltrate the central nervous system and are associated with MS plaques [4]. Therapies designed to prevent these cells becoming activated and crossing the blood brain barrier, have proven highly effective in reducing relapse rates and disease burden as indicated by the lesion load on magnetic resonance imaging (MRI) scans in relapsing remitting MS [5–7]. Although early effective therapeutic intervention for autoimmune diseases has been shown to be critical in delaying progression [8], and individuals respond to some therapies better than others, practical biomarkers have not yet been identified to guide clinical management.

Assessment of therapeutic success is based on prevention of relapses and reduction of gadolinium enhanced lesions on MRI of the brain. Relapses can be highly variable in presentation and MRI scans are infrequently assessed, due to cost and logistical considerations. Much neuronal damage can precede identification of therapeutic failure, so it is critical to identify biomarkers of therapeutic response that can be employed to monitor disease progress more sensitively and frequently [9]. Many of the GWAS genes expressed in leukocytes are transcription factors controlling their differentiation [1,3,10,11], implicating variation in regulation of immune cell differentiation as underpinning disease susceptibility and driving MS pathogenesis. We reasoned that expression of these genes in blood leukocytes should indicate disease state and type, and be useful in clinical practice. We therefore interrogated their expression in whole blood using RNAseq, microarray analysis and quantitative RTPCR in three independent cohorts.

As a discovery cohort, we analyzed the expression of MS-associated transcription factors in whole blood from 72 individuals (32 untreated MS, 40 controls) using RNAseq. The transcription factors controlling T and NK cell differentiation [12], EOMES and TBX21, were highly dysregulated in untreated MS. We then tested this association in two independent replication cohorts of 71 individuals from Australia (41 untreated MS, 31 controls) and in 52 people from San Francisco (15 untreated CIS/MS, 37 controls). The stability of EOMES and TBX21 expression over time was assessed from longitudinal blood collections. The association of this expression with disease progression, disability and clinical course was determined in each cohort. The consilience of their dysregulation, genetic association, temporal stability, genotype-dependent expression and the pathogenic pathways on which they function, suggest measurement of expression levels of these genes may be clinically useful in MS and other autoimmune diseases.

2. Material and methods

2.1. Sample collection

Single PAXgene blood RNA tubes (PreAnalytiX, Switzerland) were collected from subjects with MS who were not receiving any treatment, and had not received any immunomodulatory therapy in the previous three months, from clinics in Sydney, Adelaide and the University of California, San Francisco (UCSF). Blood was also collected from healthy controls. Each cohort was matched for age, gender and for season, and time of day of blood draw. Demographics for patients and controls for each cohort are shown in Table 1.

Informed written consent was obtained from each donor, and the study was approved by the relevant Human Research Ethics Committee.

2.2. Quantitation of gene expression

mRNA from the discovery cohort was interrogated for whole transcriptome gene expression using the Illumina Hiseq 2000 as described previously [13]. RTPCR has been used to validate measurements in this cohort [13]. Total mRNA from the UCSF cohort was analyzed using the Affymetrix Exon 1.0 ST array as previously described, and Nanostring assays were used to validate measurements in that cohort [14]. For the RT PCR cohort cDNA was prepared from mRNA using SuperScript™ III Reverse Transcriptase (Invitrogen, California) and Taqman Gene Expression Assays (Life Technology, California) were employed for the GWAS genes EOMES (Hs00172872_m1), TBX21 (Hs00203436_m1), cMAF (Hs04185012_s1), ZNF1 (Hs00393480_m1), STAT3 (Hs00374280_m1); and other transcription factors controlling T cell differentiation: RUNX3 (Hs00231709_m1), TOX (Hs01055573_m1), ZBTB7 (Hs00757087_s1), GATA3 (Hs00231122_m1), FOXP3 (Hs01085834_m1), RORC (Hs01076112_m1) and GAPDH (house keeping comparator, Hs0275899_g1). Derivation of cell subset mRNA and subsequent RNAseq has been described previously [13].

2.3. CCL5 quantitation in serum

Serum separated from whole blood was collected in a single CAT BD Vacutainer from subjects with MS who were not receiving any treatment, and had not received any immunomodulatory therapy in the previous three months, from clinics in Sydney, and also from healthy controls. The serum was diluted 1/320 and quantified for CCL5 according to the RayBio Human RANTES ELISA kit (RayBiotech Inc., Georgia). The ELISA plate was read at 450 nm on the Victor X3 (PerkinElmer, Massachusetts) and the absorbance readings
were analyzed using SoftMaxPro 4.3.1 (Molecular Devices, California).

2.4. Statistics

The association of transcription factor (TF) expression with disease and season was tested using an ANOVA. Age was tested using Pearson's correlation. Two tailed $T$ tests were then used to assess differences in gene expression by disease state, gender and by season. Effect of disease was determined in the RNAseq cohort using the EdgeR Exact Test [15]. Season was defined as “winter” months 4–9 and “summer” months 1–3, and 10–12, corresponding to the 6 months of lowest and highest UV radiation respectively, in Sydney and Adelaide. For the UCSF cohort these were “summer” and “winter”. Correlation between TFs, longitudinal collections, TFs and other genes, was determined using Pearson correlation.

3. Results

3.1. T and NK GWAS transcription factors are underexpressed in MS

For the 110 non-HLA loci recently identified as MS-associated [1], the closest genes to the signal for 26 of them were DNA binding transcription factors, and 25 of these were predominantly expressed in immune cells (Fig. 1A), using cell subset transcriptomes we had derived previously from RNAseq data [13]. In addition, TBX21 is in linkage disequilibrium (LD) with NPEPPS as the closest gene to the most-associated SNP in the locus. Considering that only about 300 of the 20,000 known genes are TFs expressed in immune cells [16], they are enriched over 15 fold in the MS GWAS/ImmunoChip list ($p < 1E-05$, chi-square test). We determined the relative expression of these MS associated transcription factors using RNAseq for 40 healthy controls and 32 MS samples (cohort described in Table 1). In the discovery cohort, EOMES ($p = 0.0004$), TBX21 ($p = 0.0147$), ZMIZ1 ($p = 0.0016$), ZNF438 ($p = 0.03$), and TCF7 ($p < 0.04$; all $p$ values uncorrected for multiple testing) were differently expressed in MS. EOMES ($p = 0.0060$), TBX21 ($p = 0.0034$), ZMIZ1 ($p = 0.0002$) and ZFP36L2 ($p = 0.0003$) were also associated with disease in the UCSF study involving 52 individuals [14], in all cases with lower expression in MS in both cohorts (Fig. 1B).

We confirmed that these genes are predominantly expressed in NK or T cells (EOMES, TBX21 and ZFP36L2), or myeloid cells (ZMIZ1) (Fig. 1). We then tested expression of EOMES, TBX21 and ZMIZ1 by RTPCR with disease in a second Australian replication cohort. Again reduced expression of these genes was significantly associated with MS (Fig. 2, cohort details in Table 1, association of transcription factors in Supp Table 1). Gender and age were not associated with differences in expression of these genes (data not shown). Time of day [17] and season [18–20] are also known to affect immune cell subset representation in blood. All samples were collected between 0830 and 1300; and matched for season. However, it is notable that in one Australian cohort expression of TBX21, EOMES, ZMIZ1 and other transcription factors were affected by season (Supp Table 2) in healthy controls.

3.2. EOMES and TBX21 expression is correlated with each other and with expression of other NK/T cell transcription factors

Given that EOMES and TBX21 have a similar pattern of immune cell subset distribution (Fig. 1), we tested if their expression was correlated with each other. Indeed, they were positively correlated in all three cohorts (Fig. 3). Several T and NK cell transcription factors regulating differentiation have been characterized [21,22]. We had earlier identified that one of the core T cell transcription factors controlling CD4/CD8 differentiation, RUNX3, was underexpressed in MS [13]. We tested if this and other transcription factors known to regulate T and NK cells [22] were also correlated in the three cohorts. A set comprising EOMES, TBX21, RUNX3, and TOX (and to lesser

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Demographics of Australian and American cohorts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIS — clinically isolated syndrome, RRMS — relapsing remitting MS, SPMS — secondary progressive MS, PPMS — primary progressive MS, EDSS — expanded disability status scale, MSSS — Multiple Sclerosis Severity Score.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RNAseq</td>
</tr>
<tr>
<td></td>
<td>HC</td>
</tr>
<tr>
<td>Number</td>
<td>40</td>
</tr>
<tr>
<td>Age</td>
<td>44.7</td>
</tr>
<tr>
<td>Age range</td>
<td>25–78</td>
</tr>
<tr>
<td>MS</td>
<td></td>
</tr>
<tr>
<td>Age of onset</td>
<td>29.5</td>
</tr>
<tr>
<td>CIS</td>
<td>0</td>
</tr>
<tr>
<td>RRMS</td>
<td>21</td>
</tr>
<tr>
<td>SPMS</td>
<td>1</td>
</tr>
<tr>
<td>PPMS</td>
<td>10</td>
</tr>
<tr>
<td>EDSS</td>
<td>4.19</td>
</tr>
<tr>
<td>MSSS</td>
<td>6.96</td>
</tr>
</tbody>
</table>
extent, GATA3) was highly positively correlated with each other (Fig. 3, Supp Table 3; RNAseq: R ranged from 0.73 to 0.80, p < 1E-08 for all; UCSF: R = 0.61–0.75, p < 1E-06 for all; RTPCR: R = 0.44–0.72, p < 1E-03 for all). The GWAS TFs ZMIZ1 and ZFP36L2 were also correlated in the two cohorts tested (RNAseq and UCSF), with highly significant p values (RNAseq: R = 0.64, p < 2.5E-07; UCSF: R = 0.47, p < 4E-04). For Set 1 (EOMES, TBX21, RUNX3, TOX) these correlations are in agreement with the cell subset representation of these transcription factors (Fig. 4C), all being predominantly expressed in T and NK cells, but transcripts from Set 2 are only expressed in lymphocytes (ZFP36L2) and myeloid cells (ZMIZ1) (Fig. 1A).

We had earlier identified that NK cell genes were overrepresented in the list of genes underexpressed in MS as compared to controls [13]. We tested if other NK cell genes underexpressed in the MS RNAseq data were correlated with Set 1 transcription factors. Strikingly, several were, with CCL5/EOMES being the most tightly correlated (RNAseq: R = 0.87, p < 2.9E-17) Figs. 4A–B, Supp Table 4). CCL5 expression was also correlated with expression of these transcription factors in the UCSF cohort (UCSF: CCL5/EOMES R = 0.60, p < 1.9E-06). Serum CCL5 protein was not correlated with EOMES mRNA in whole blood in the PCR cohort (Fig. 4D).

### 3.3. EOMES and TBX21 expression is stable over time

The temporal stability of TBX21 and EOMES expression was assessed in the UCSF cohort, where collections from the same individuals from up to 3 yearly time points had been made [14]. Expression levels from consecutive yearly time points were highly correlated for both EOMES and TBX21 in controls and CIS/MS (Fig. 5). The MS patients with lower levels of expression than most controls were consistently lower at subsequent time points. This was also true for the other Set 1 TFs and Set 2 TFs.

This observed stable expression of the transcription factors over time suggested expression might be under genetic control. We sought in silico evidence that SNPs in cis and trans affect whole blood expression of these transcription factors and CCL5. The GWAS associated SNP for EOMES, rs11129295, is highly associated with EOMES expression, and expression of all of the Set 1 transcription factors (EOMES, TBX21, RUNX3, TOX), the Set 2 transcription factor ZFP36L2, and CCL5 is genotype-dependent (Supp Table 5,[23]). The minor allele of rs11129295, which is associated with increased risk of MS, is the more highly expressed. The expression of EOMES is lower in MS, but we would expect the expression level of EOMES in MS to be the net effect of many genetic variants, not just rs11129295.

---

**Figure 1**  Expression of transcription factors associated with MS by GWAS/ImmunoChip. (A) Immune cell subset expression as measured by RNAseq and (B) differential expression in MS (RNASeq) or CIS (UCSF) patients compared to healthy controls for two independent cohorts. Orange color indicates increased expression level; blue color indicates decreased expression level. p value for RNASeq cohort is for EdgeR Exact Test [15], p value for UCSF cohort is for two tailed T test.
3.4. Association with disability, clinical course, and progression

We computed the association of the low expression MS phenotype with age of onset, disability, progression and clinical course. Disability was assessed using the expanded disability status scale (EDSS, [24]). Progression was the rate of EDSS change (MSSS, [25]). Clinical course was relapsing remitting MS (RRMS), secondary progressive MS (SPMS) or primary progressive MS (PPMS). There was no correlation between transcription factor expression and any of these phenotypes.

4. Discussion

The GWAS studying genetic variants affecting risk of autoimmune diseases have identified shared risk factors, indicating a common molecular and cellular architecture of these diseases [2,11]. The genetic variants mainly occur in regulatory regions, indicating that the control of gene expression which alters immune cell differentiation underpins variation in heritability [26]. Here we have shown that the expression of the transcription factors identified as MS risk genes: EOMES, TBX21, ZFP36L2 and ZMIZ1, is low in MS. EOMES and TBX21
expression is positively correlated, as is expression of ZFP36L2 and ZMIZ1. EOMES and TBX21 expression is also positively correlated with the T and NK cell transcription factors RUNX3, TOX, and GATA3, and with genes under-expressed in MS, notably CCL5. These data are replicated in three independent cohorts. Expression of these transcription factors is stable over time in healthy controls and CIS patients, and known to be genotype dependent, indicating that the gene expression level is characteristic for individuals and so defines a molecular phenotype. The lower expression phenotype, ETlow, predicts increased risk of MS, would be stable without intervention, but may normalize with successful therapy. Increased expression might be useful in assessing positive response to therapy.

The association of altered expression of transcription factors regulating T, NK and myeloid cell subsets in all three cohorts is consistent with other lines of evidence supporting a role for altered immune cell differentiation in whole blood as contributing to MS pathogenesis [27–32]. Since EOMES, TBX21, RUNX3 and TOX are predominantly expressed in NK and CD8CD45RO T cells, the ETlow molecular phenotype identified here likely represents a lower contribution to the whole blood transcriptome from these cells, or major subsets of them. Both CD8CD45RO and NK cell subsets have previously been reported as aberrant in MS (reviewed in [33]). Immunophenotyping with a comprehensive array of CD markers has identified CD8<sup>low</sup>CD4<sup>−</sup> cells, particularly NK cells, as under-represented in peripheral blood in MS [34]. It is notable that the genes identified as underexpressed in whole blood in MS are predominantly from NK cells [13]. From this study we could hypothesize that those with the MS risk phenotype also have an altered NK/CD8CD45RO cellular phenotype. Note the stable ETlow phenotype observed here is consistent with other studies showing stable whole blood immunophenotype [20], chemokine and cytokine profiles [35] and flow cytometry observations on longitudinal variation in CD4 cell subsets [36].

The association of the ETlow phenotype with low CCL5 is consistent with other lines of evidence suggesting that increased immune cell trafficking from the blood contributes to, or even drives, MS pathogenesis [35,37]. The major source of CCL5 is immune cells, so that reduced expression of mRNA observed in MS would lead to less extracellular CCL5 in the blood, which would lead to a chemokine gradient out of the blood. The cells expressing receptors for CCL5 (CCR5, CCR3 and CCR1) would traffic to the tissues, and so be more likely to promote an immune response at the target site. Blocking this trafficking may be beneficial, especially in those with the ETlow phenotype. The level of CCL5 has been

---

**Figure 4**  Expression level of CCL5 and EOMES RNA is correlated in (A) RNAseq cohort of healthy control and MS patient samples and (B) UCSF cohort of healthy control and CIS patient samples. R is Pearson correlation coefficient with corresponding p value (p). (C) Immune cell subset expression of transcription factors not identified as associated with MS from GWAS. Orange color indicates increased expression level; blue color indicates decreased expression level. (D) Correlation of serum CCL5 protein with EOMES mRNA in whole blood (RTPCR cohort).
reported as higher in the CSF of MS compared to controls [38], an increased level of CCL5 associated with different clinical forms of MS [39], and with lower levels of receptor in different MS forms reported in another study [40]. Antagonists for CCR5 have been reported to reduce severity of EAE [41,42]. People with HIV on long term antiretroviral therapy, and so reduced CCR5 availability, have been reported to have a reduced risk of MS [43]. Since these aspects of trafficking are likely to be tissue agnostic, the ETlow phenotype identified here may also predict risk of other autoimmune diseases, which may respond to agents blocking CCL5 interactions.

The transcription factors shown not to be aberrantly expressed here, and the other MS susceptibility genes, also regulate immune cell subset differentiation [1,10], as do GWAS-identified genes for other autoimmune diseases [2]. Monoclonal antibodies targeting all leukocytes, or particular subsets of leukocytes, can greatly reduce relapse rates and numbers of gadolinium enhanced lesions detected by MRI. These include Alemtuzumab, which targets a receptor, CD52, expressed on all leukocytes [44]; Tysabri which targets CD49D, expressed on T cells [6]; or Daclizumab, which targets CD25, expressed on T and NK cells [45]. Notably CD25 (IL2Ralpha) itself, and the ligand for CD49D, VCAM1, are in MS risk factor loci. Small molecule antagonists for sphingosine 1 phosphate receptors, which block trafficking between the lymphoreticular system and the blood, are also effective [46]. Many of the transcription factors regulate B cells, and monoclonal antibodies to surface markers specific for B cells are also effective in MS [47,48]. The cross-protection, where more than 50% of patients respond to both B and T cell specific therapies, evident from these studies, suggests that targeting subsets has benefits to overall modulation of immune response. However, therapies are not universally effective, nor without adverse reactions [49], and because early control of disease is crucial in delaying the progressive phase of the disease [8], personalizing therapy is crucial. The ETlow phenotype may be useful for this purpose.

We found no association between MS clinical course or disease severity and the ETlow phenotype, suggesting this phenotype is unlikely to have predictive value for disease progression in the absence of therapy. However, the many different types of drugs now available for MS target different immune cell subsets and processes, and may have differential effects on disease progression in individuals with different underlying pathologies. Such differences in response may be detectable by the extent of normalization of the ETlow phenotype on particular therapies. Assessing this, and identification of the immune cell subset driving this

Figure 5  Expression of EOMES and TBX21 is stable over time in the whole blood of healthy controls and CIS patients.
gene expression phenotype, may lead to diagnostic tools to choose therapy and evaluate its success, and to develop new therapies.

Conflict of interest statement

The authors declare that there are no conflicts of interest.

Acknowledgments

We would like to thank the people with MS and controls for donating blood to support this research. This work was funded by grants from the Multiple Sclerosis Research Australia (MSRA). David Booth is the MSRA Hunt Family Senior MS Research Fellow.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.clim.2014.01.003.

References


D.M. Evans, I.H. Frazer, N.G. Martin, Genetic and environmental causes of variation in basal levels of blood cells, Twin Res. 2 (1999) 250–257.


