Systematic review of genome-wide expression studies in multiple sclerosis

A K Kemppinen,[1,2,3] J Kaprio,[1,4,5] A Palotie,[1,6,7,8] J Saarela[1,9,10]

For numbered affiliations see end of article.

Correspondence to
Anu Kemppinen;
ak635@medschl.cam.ac.uk

ABSTRACT

Background: Although recent genome-wide association studies have identified several genetic variants contributing to the complex aetiology of multiple sclerosis (MS), expression and functional studies are required to further understand its molecular basis.

Objectives: To identify genes and pathways with differential expression in MS.

Design: The authors conducted a systematic review of seven microarray studies, in which expression in immune cells was compared between MS patients and controls. These studies include a previously unpublished study, which is described here in detail.

Results and conclusion: Although in general the overlap between studies was poor, 229 genes were found to be differentially expressed in MS in at least two studies, of which 11 were in three studies and *HSPA1A* in four studies. After excluding the authors' unpublished experiment which may have been affected by certain confounding factors and inclusion of treated subjects, 135 genes were identified in at least two studies. The differentially expressed genes were significantly associated with several immunological pathways, including interleukin (IL)-4, IL-6, IL-17 and glucocorticoid receptor signalling pathways. 15 of the 229 loci have shown some association with MS in published genome-wide association studies (p<0.0001), including three loci with confirmed MS risk variants.

ARTICLE SUMMARY

Article focus

- To identify genes showing differential expression in multiple sclerosis through genome-wide expression profiling in peripheral blood mononuclear cells.
- To conduct a systematic review of genome-wide expression studies in multiple sclerosis in order to identify the most frequently reported genes.
- To identify pathways associated with genes most frequently reported as differentially expressed in multiple sclerosis.

Key messages

- The vast majority of all genes reported as differentially expressed were only identified in a single study.
- However, 229 genes were reported as differentially expressed in MS to the same direction in at least two of the seven studies reviewed, 12 genes of which were in at least three studies.
- After excluding our unpublished experiment. which may have been affected by confounding factors and inclusion of treated subjects, 135 genes were identified in at least two studies.
- The differentially expressed genes were significantly associated with several immunological pathways, including the IL-4, IL-6, IL-17 and glucocorticoid receptor signalling pathways.

INTRODUCTION

The aetiology of multiple sclerosis (MS) is complex and involves both genetic susceptibility and environmental factors. However, apart from the widely replicated association with human leukocyte antigen (*HLA*)-*DRB1*1501, genetic risk factors have remained unknown until genome-wide association studies (GWAS), which have recently led to identification of common MS risk variants in over a dozen loci.[1–6] Although further fine-mapping and functional studies are required in order to verify the causal variants and genes, a strong presence of immunological genes in these loci is evident. Expression and functional studies in immune cells can therefore elucidate the molecular mechanisms behind MS. Indeed, a number of studies have been conducted where genome-wide expression profiles in peripheral immune cells were compared between MS patients and unaffected controls. Together, these studies have reported a large number of genes with differential expression in MS. However, given that most of these studies have been conducted in small samples without replication, it is likely that many of the findings are false positives. Approaches are therefore needed to increase the probability of detecting the true signals from the vast number of reported genes. In order to extract the genes which are more likely to be true positives, we systematically reviewed results from seven microarray studies in MS, including our previously unpublished study. First, we identified genes

ARTICLE SUMMARY

Strengths and limitations of this study

- This is the first systematic review of genome-wide expression studies conducted in peripheral immune cells in multiple sclerosis.
- Strict criteria were applied for inclusion of studies, and clearly underpowered studies with fewer than 10 cases or controls were excluded.
- Many of the genes we found to be reported by at least two studies have interesting immunological functions and can be considered promising candidates for further studies.
- However, the studies included should not be considered directly comparable owing to differences in samples, platforms and analyses methods used. In addition, the majority of these studies are small and should be viewed with some caution.
- All studies were conducted in relatively heterogeneous cell populations, and some of the findings could therefore be explained by differences in numbers of different cell populations rather than differential transcriptional activity in MS.
- Finally, our previously unpublished microarray study may have been affected by differences between the labs where the patient and controls samples were prepared for arrays, as well as by the higher mean age of controls. Our study also included four patients who had received immunomodulatory treatment at the time of sample collection.

which had been found to be differentially expressed in MS to the same direction in at least two studies. In order to further examine the potential role of these most frequently reported genes, they were analysed using pathway tools. Finally, we searched these genes for evidence of association by making comparisons with top results from recent MS GWAS.

MATERIALS AND METHODS
Samples in the Finnish microarray experiment

Twelve female patients fulfilling Poser's criteria for clinically definite MS were recruited through the Seinä-joki Central Hospital. Fifteen healthy unrelated female controls were obtained from the Finnish Twin Study on Ageing (FITSA). The mean age was 54.2 in patients and 71.6 in controls. One patient was receiving cortisone treatment, two patients received β interferon, and one patient was being treated with both β interferon and cortisone at the time of sample collection. All subjects had provided their informed consent. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood using BD Vacutainer CPT Cell Preparation Tubes (Becton, Dickinson and Company, Franklin Lakes, New Jersy), and cells were disrupted and RNA extracted with TRIzol Reagent (Invitrogen, Carlsbad, California). RNA was then purified using Rneasy Mini Kit (Qiagen, Hilden, Germany), and the sample quality was examined using BioAnalyzer (Agilent, Santa Clara, California). The study was approved by the Committee on Ethics of the Central Hospital of Central Finland and by the Helsinki University Hospital Ethical Committee of Ophthalmology, Otorhinolaryngology, Neurology and Neuro-

surgery (permit 192/E9/02) for FITSA and patient samples, respectively.

Sample processing and microarrays

Eleven patient samples were prepared for hybridisation on the Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (Affymetrix, Santa Clara, California) according to the manufacturer's recommendations in our laboratory. In addition, one patient sample and technical replicates from two of the 11 patient samples were prepared according to the manufacturer's recommendations at the Helsinki Biomedicum Biochip Centre (BBC), where 15 control samples had been previously prepared. In brief, $1-2$ µg of total RNA was converted to biotin-labelled cRNA using the Affymetrix HT One-Cycle cDNA Synthesis Kit and the HT IVT Labelling Kit. Fifteen micrograms of cRNA was then fragmented and hybridised for 16 h at 45°C, washed in Affymetrix Fluidics Station 450 and scanned with Affymetrix Gene-Chip Scanner 3000. Hybridisation, washing, staining and scanning were conducted using the same instruments for all samples. All arrays had a present call percentage >40 $(42-47)$ and average background signal <50 $(36-44)$.

Microarray data analysis

Raw intensity data files were imported to GeneSpring 7.3 (Agilent Technologies, Santa Clara, California) and GC Robust Multi-array Average (GC-RMA) normalised. We then applied filtering steps in order to exclude probe sets with low signal intensity and probe sets potentially affected by differences between our laboratory and the BBC. First, we excluded all probe sets with a GC-RMA normalised signal <50 in at least 20 of the 27 arrays (N=35 203). Second, we excluded probe sets where the technical replicates showed a >1.4-fold difference in both replicate pairs (N=1668). Finally, we excluded probe sets where the signal in all three MS arrays prepared at the BBC ranked among the four lowest or four highest among the MS arrays (N=2469), after which 15 273 probe sets remained for analyses. After filtering, we discarded the two replicate arrays prepared in our laboratory and used only the replicates prepared at the BBC for final analyses, which thereby included 12 MS arrays and 15 control arrays. In order to identify genes with differential expression in MS, we first applied the fold-change filter in GeneSpring 7.3 using 1.5 as threshold. For probe sets showing a $\pm \geq 1.5$-fold difference in mean expression between MS patients and controls, we further determined non-parametric Mann−Whitney sum rank test p values with Benjamini−Hochberg correction for multiple testing. Probe sets with corrected a p value of ≤ 0.05 were considered to be differentially expressed. In order to annotate these probe sets, we compared the gene symbol obtained from GeneSpring 7.3 in October 2008 with the NetAffx annotation in December 2010 for each probe set (http://www.affymetrix.com/analysis/index.affx). If these were different, we first checked whether these were

alternative identifiers for the same gene. If not, we obtained the probe set nucleotide sequences from NetAffx and performed a Blat search in UCSC Genome Browser (hg19) to identify the correct target gene (http://genome.ucsc.edu/cgi-bin/hgGateway). Probe sets which recognise only intronic sequences or map to several loci were excluded from the list of differentially expressed probe sets. The Gene Expression Omnibus (GEO) accession number for the microarray dataset is GSE21942 (http://www.ncbi.nlm.nih.gov/geo/).

Pathway analysis

Pathway analyses were conducted using the Core Analysis option in the Ingenuity Pathway Analysis software (Ingenuity Systems, Redwood City, California). This option identifies canonical pathways associated with a given list of genes by calculating the Fisher exact test p value for the probability that association between this set of genes and a canonical pathway is explained by chance alone. In order to account for the fact that our input lists of genes were enriched for immunological genes and would therefore show association with immune-related pathways if compared with all genes, we restricted the analyses to genes expressed in immune cells by applying the Tissues & Cell lines filter in the analysis settings.

Systematic review of microarray studies in MS

We searched PubMed with keywords 'multiple sclerosis microarray' in November 2010 and obtained 156 records. These were complemented with two additional recent studies and studies identified through a review article.[7−9] Based on title, abstract and, if required, full text, we first identified all studies which had been conducted in peripheral immunological cells using a microarray platform and compared expression profiles between MS patients and unaffected controls. This led to the exclusion of 144 studies, which were not MS-related expression studies or investigated effects of MS treatments, were conducted in animal models for MS, or were performed in MS brain biopsies rather than immune cells. We then reviewed the remaining studies in further detail and excluded six studies with fewer than 10 MS patients and/or controls, a study which did not identify any differentially expressed genes, a study where only genes involved in T-cell mediated cytotoxicity were included in the analyses and a study where the list of identified genes was not available in the publication or upon contact with the corresponding author. The stages of the selection process are depicted as a flow chart in figure 1. Of the remaining eight studies which are listed in table 1, only six were independent, because the two studies by Satoh et al[15 16] and the studies by Achiron et al[10] and Mandel et al[13] had been conducted in the same set of patients and were from hereon considered to be single studies. After including our own unpublished experiment, we therefore had seven independent studies for the analysis. For each study, we listed all genes which had been reported to be differentially expressed in MS patients in comparison with healthy controls, and
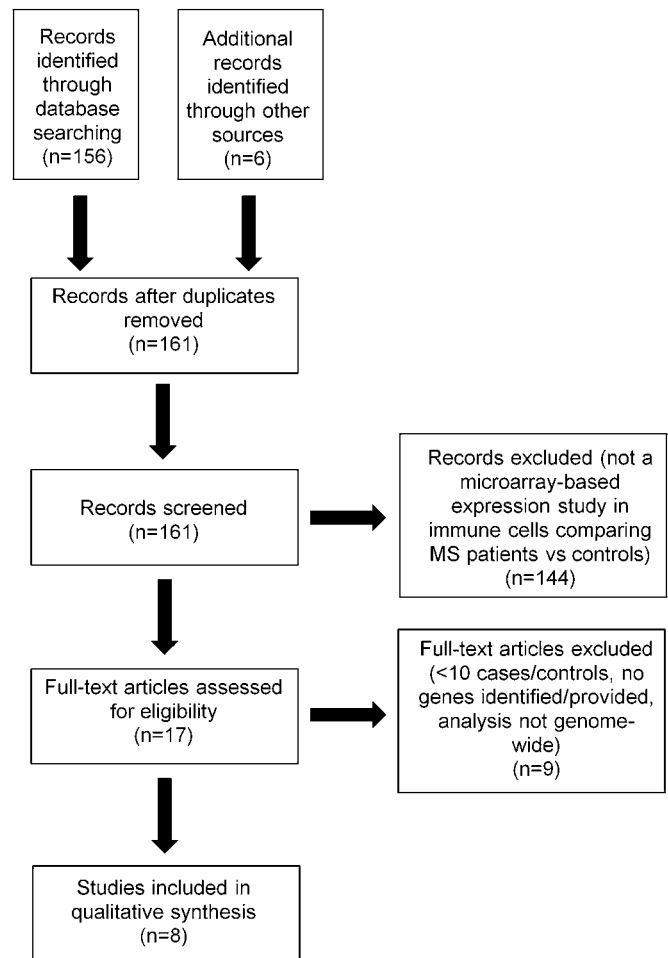

Figure 1 Flow chart showing stages in selecting studies for systematic review. MS, multiple sclerosis.

recorded whether their expression in MS was increased or decreased. The studies by Achiron et al[10] and Mandel et al[13] listed only a selected subset of the identified differentially expressed genes.[10 13] The corresponding author was contacted in order to obtain the full lists of differentially expressed genes, but we received no reply, and we therefore only included the reported genes. All gene symbols were mapped to the Human Gene Nomenclature gene symbols, and genes that were not unambiguously linked to a single gene symbol using the information available were excluded.

RESULTS

The previously unpublished microarray screen identifies 692 probe sets with differential expression in MS

We first analysed the data from our previously unpublished Finnish microarray screen and identified 692 probe sets, which showed a ≥1.5-fold difference in mean expression between 12 MS patients and 15 controls together with a non-parametric Mann−Whitney sum rank test p value of ≤0.05 after Benjamini−Hochberg correction (supplementary table 1). Three hundred and one probe sets showed increased expression, and 391 decreased expression in MS. Pathway analysis revealed that the differentially expressed genes were strongly

Table 1 Description of previous microarray expression studies included in the systematic review

Cells	Platform	No of cases (females/males)	No of controls (females/males)	Mean age (cases/controls)	No of reported differentially expressed genes with a Human Gene Nomenclature symbol	Original publication
PBMC	U95Av2 array (Affymetrix, California). Represents approximately 10 000 genes.	26 (20/6)	18 (16/2)	41/39.6	34	Achiron et al[10]
Whole blood	10.5 K Peter MacCallum array (The Peter MacCallum Cancer Institute, Australia). Represents 9381 unique cDNAs.	20 (11/9)	Pooled sample (20/0); five individual samples (0/5)	Not provided (range 30–66 in MS, 20–55 in controls)	2217	Arthur et al[11]
PBMC	cDNA arrays with 6500 or 7500 clones with partial overlap	24 (15/9)	19 (5/14)	42/37.5 (age for four controls unknown)	104	Bomprezzi et al[12]
Whole blood	HT-12 array (Illumina, California). Represents 48 804 transcripts.	99 (66/33)	45 (29/16)	54/48.5	779	Gandhi et al[9]
PBMC	U95Av2 array (Affymetrix, California). Represents approximately 10 000 genes.	13 (9/4)	5 (4/1) systematic lupus erythematosus (SLE), 10 healthy controls (16/2)	40.7/42.8 (SLE)/39.6	78	Mandel et al[13]
PBMC (monocytes depleted)	GeneFilters GF211 DNA array (Research Genetics, Alabama). Represents 5184 genes.	15 (11/4)	15 (gender matched to cases, but numbers not available)	42.1/41.7	28	Ramanathan et al[14]
T cells, non-T cells	A custom array with 1258 cDNAs (Hitachi Life Science, Saitama, Japan)	72 (55/18)	22 (16/6)	36.1/38.6	185	Satoh et al[15]
T cells	A custom array with 1258 cDNAs (Hitachi Life Science, Saitama, Japan)	72 (55/18)	22 (16/6)	36.1/38.6	281	Satoh et al[16]

PBMC, peripheral blood mononuclear cell.

associated with PI3K signalling in B lymphocytes (p=1.3E−06), and B cell development (p=2.6E−06) pathways. In addition, altered T cell and B cell signalling in rheumatoid arthritis, role of PKR in interferon induction and antiviral response, and production of nitric oxide and reactive oxygen species in macrophages pathways showed evidence of association with the differentially expressed genes (p<1E−04).

Systematic review identifies 229 genes reported in at least two studies

In order to perform a systematic review of expression studies in MS, we identified eight previous studies, which filled our selection criteria (table 1). Studies which had been conducted in the same set of patients were combined, and we therefore considered six previously published independent lists of differentially expressed genes. Together with our previously unpublished experiment, these studies identified 2017 unique genes with increased expression and 1860 with decreased expression in MS, of which 303 genes were reported to have both increased and decreased expression. However, this list is not comprehensive, as two of the studies did not provide full lists of the identified genes. Two hundred and twenty-nine of the genes were found to have differential expression in MS to the same direction in at least two studies (supplementary table 2) and will be referred to as 'in silico replicated' differentially expressed genes (DEGs). One hundred and eleven of these had increased expression, and 119 decreased expression in MS, with one gene (STK4) identified as both. Eleven of the 229 genes were differentially expressed in three studies, and one gene, HSPA1A, in four studies (table 2). We acknowledge that our unpublished experiment may have been confounded by technical differences in processing of the patient and control samples, as well as by the age difference between cases and controls. In addition, the study included four patients who had received treatment, which may also have had an impact on the results. After excluding

our study, 135 of the 229 DEGs were identified in at least two independent studies. These are indicated in supplementary table 2.

Several immunological pathways are significantly associated with the in silico replicated DEGs

In order to explore whether the in silico replicated DEGs are associated with specific pathways, we applied again the canonical pathway analysis option in the Ingenuity Pathway Analysis software. We conducted pathway analysis on the 135 DEGs identified in at least two independent studies excluding our unpublished study, and on the 229 DEGs identified when including our study. Unsurprisingly, many of the 20 most significantly associated pathways are related to immunological functions (tables 3, 4), with the most significantly associated pathway in both analyses being the glucocorticoid receptor signalling pathway. Other top-rated pathways include molecular mechanisms of cancer, NF-κB signalling and several interleukin (IL) signalling pathways (IL-4, IL-6, IL-17).

FIFTEEN IN SILICO REPLICATED DEGS ARE SUGGESTIVELY ASSOCIATED WITH MS

We then investigated whether the 229 in silico replicated DEGs showed any evidence for association in five previously published GWASs or in a recently published GWAS meta-analysis after excluding single nucleotide polymorphisms (SNP) in the human leukocyte antigen region (6p21−22).[1−6] All association p values were available for the GWASs by Jakkula et al[4] and International Multiple Sclerosis Genetics Consortium (IMSGC),[1] and we therefore first examined the p value distribution of SNPs mapping within 100 kb of the 229 DEGs by quantile−quantile plots. Interestingly, these SNPs showed an overall enrichment of p values in the range 0.1−0.001 in the IMSGC GWAS (figure 2). However, we did not see any evidence of enrichment in the Finnish GWAS. We then reviewed the reported SNPs in the five published GWASs and the GWAS meta-analysis, and found 15 of the 229 in

Table 2 Genes identified as up- or down-regulated in multiple sclerosis in at least three studies

Gene	Description	Direction of change in expression in multiple sclerosis
ATP7A	ATPase, Cu^{2+} transporting, alpha polypeptide	Decreased (FIN[11 15 16])
CCL3	Chemokine (C-C motif) ligand 3	Decreased[11−13]
CDKN1C	Cyclin-dependent kinase inhibitor 1C (p57, Kip2)	Decreased (FIN[9 11])
HSPA1A	Heat shock 70 kDa protein 1A	Decreased[9 12 13 15 16]
PLAUR	Plasminogen activator, urokinase receptor	Decreased[11−13]
EIF4A1	Eukaryotic translation initiation factor 4A1	Increased (FIN[9 11])
NEAT1	Nuclear paraspeckle assembly transcript 1 (non-protein coding)	Increased (FIN[9 11])
OGT	O-linked N-acetylglucosamine (GlcNAc) transferase	Increased (FIN[11 12])
PTGS2	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclo-oxygenase)	Increased[11−13]
RBBP6	Retinoblastoma-binding protein 6	Increased (FIN[9 11])
TNFAIP3	Tumour necrosis factor, alpha-induced protein 3	Increased (FIN[11 15 16])
ZMYND8	Zinc finger, MYND-type containing 8	Increased (FIN[11 12])

FIN, previously unpublished Finnish microarray screen described in detail here.

Table 3 Top 20 pathways associated with the 229 in silico replicated differentially expressed genes

Pathway	No of differentially expressed genes in the pathway/total no of genes in the pathway	Fisher exact test p value
Glucocorticoid receptor signalling	30/250	2.1E−16
IL-6 signalling	15/88	6.0E−11
Hepatic fibrosis/hepatic stellate cell activation	16/116	3.6E−10
Molecular mechanisms of cancer	24/303	1.8E−09
IL-17 signalling	12/70	4.9E−09
Pancreatic adenocarcinoma signalling	14/102	5.1E−09
Nuclear factor (NF)-κβ signalling	16/154	2.4E−08
IL-10 signalling	11/65	2.5E−08
Peroxisome Proliferator-Activated Receptor (PPAR) signalling	12/81	2.7E−08
Role of osteoblasts, osteoclasts and chondrocytes in rheumatoid arthritis	17/189	7.5E−08
Germ cell−Sertoli cell junction signalling	14/138	2.5E−07
Dendritic cell maturation	14/139	2.7E−07
Apoptosis signalling	11/86	4.9E−07
p38 Mitogen-Activated Protein Kinase (MAPK) signalling	11/89	6.9E−07
Colorectal cancer metastasis signalling	16/206	1.3E−06
Atherosclerosis signalling	10/77	1.4E−06
Tumor Necrosis Factor Receptor 1 (TNFR1) signalling	8/45	1.4E−06
Regulation of IL-2 Expression in activated and anergic T lymphocytes	10/79	1.8E−06
Peroxisome Proliferator-Activated Receptor Alpha (PPARα)/Retinoid X Receptor Alpha (RXRα) activation	13/142	2.2E−06
IL-4 signalling	9/64	2.4E−06

IL, interleukin.

silico replicated DEGs (including regions 100 kb up and downstream) to contain suggestively associated SNPs (p≤0.0001) (table 5). The risk variants near three of the genes, *CDK4*, *IL7R* and *TNFRSF1A*, have been confirmed with genome-wide significance in GWASs or their follow-up studies (p≤5×10^{-8}).[1 3 5 17]

Table 4 Top 20 pathways associated with the 135 in silico replicated differentially expressed genes after excluding the Finnish experiment

| Pathway	No of differentially expressed genes in the pathway/total no of genes in the pathway	Fisher exact test p value
Glucocorticoid receptor signalling	22/250	3.7E−14
IL-17 signalling	9/70	7.0E−08
Pancreatic adenocarcinoma signalling	10/102	1.8E−07
IL-6 signalling	9/88	5.2E−07
IL-2 signalling	7/52	1.5E−06
Phosphatase and Tensin Homolog (PTEN) signalling	9/101	1.7E−06
IL-15 signalling	7/60	4.1E−06
Agrin interactions at neuromuscular junction	7/60	4.1E−06
Tumor Necrosis Factor Receptor 1 (TNFR1) signalling	6/45	9.4E−06
p21-Activated Protein Kinase (PAK) signalling	7/75	1.9E−05
Phosphoinositide 3-Kinase (PI3K)/AKT signalling	8/108	2.5E−05
Regulation of IL-2 expression in activated and anergic T lymphocytes	7/79	2.6E−05
Molecular mechanisms of cancer	13/303	3.0E−05
PPAR signalling	7/81	3.1E−05
Aryl hydrocarbon receptor signalling	8/115	4.0E−05
Role of Janus Kinase 1 (JAK1) and Januse Kinase 3 (JAK3) in γc cytokine signalling	6/59	4.6E−05
Nuclear factor (NF)-κβ signalling	9/154	5.2E−05
Stress-Activated Protein Kinase (SAPK)/c-Jun N-terminal Kinase (JNK) signalling	7/90	6.1E−05
Renal cell carcinoma signalling	6/64	7.3E−05
IL-4 signalling	6/64	7.3E−05

IL, interleukin.
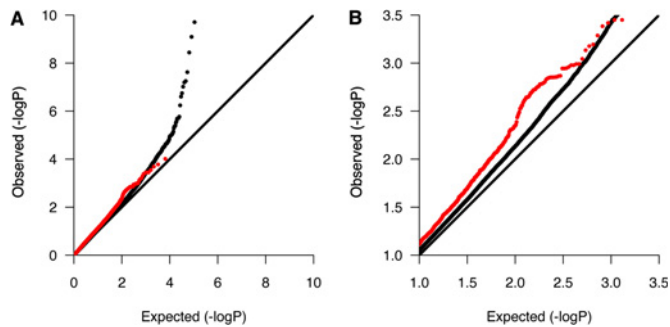
Figure 2 Quantile−quantile plot for Cochran−Mantel−Haenszel p values in the International Multiple Sclerosis Genetics Consortium genome-wide association studies highlighting single nucleotide polymorphisms (SNP) in genes showing differential expression in multiple sclerosis. SNPs in 6p21−22 were excluded. Red dots represent SNPs mapping within 100 kb of in silico replicated differentially expressed genes. The remaining SNPs are represented by black dots. Plot A shows the entire p value distribution, and plot B is a close-up showing the enrichment of observed p values in the range 0.1−0.001.

Suggestively associated variant correlates with *CXCR4* expression

To address the question, whether the above identified suggestively associated SNPs (p≤0.0001) mapping to 15 in silico replicated DEGs might contribute to the differential expression of these genes in MS, we first examined the genes using the mRNA by SNP Browser (http://www.sph.umich.edu/csg/liang/asthma/),[18] [19] which is a database of expression QTL (eQTL) SNP variants. However, none of the suggestively associated SNPs or their tagging SNPs (r^2>0.8) correlated significantly with the expression of the corresponding DEG (p≤0.001). We also investigated the expression of these genes in risk allele carriers versus non-carriers in lymphoblastoid cell lines of 60 Centre d'Étude du Polymorphisme Humain (CEPH)-derived HapMap samples (CEU) (GEO dataset GSE5859)[20] and obtained genotypes for the most strongly associated SNP, or SNPs if identified in several GWASs, from HapMap Release 24 (http://www.hapmap.org).[21] In *CXCR4*, we tested rs882300, which was the most strongly associated *CXCR4* SNP in a meta-analysis, instead of the two SNPs mapping within 100 kb of the gene. We did not identify any significant differences in expression between risk allele carriers and non-carriers after correcting for multiple testing. However, we found suggestive evidence for a higher expression of *CXCR4* in carriers of the associated G allele at rs882300 (uncorrected one-sided p value=0.04, fold change=1.36) (figure 3), which is in concordance with the higher expression of *CXCR4* observed in MS in our Finnish microarray study and three other studies, including two which were excluded from the systematic review owing to small sample sizes.[11] [22] [23] This effect was also seen with the same probe set (217028_at) in our microarray data in a combined sample of 11 MS patients and 14 controls for which a genotype was available

(fold change 1.19, one-sided Mann−Whitney test p value=0.06). However, no significant difference in expression levels was observed in two other probe sets measuring for the expression of *CXCR4* in our microarray data. Both of these probe sets recognize exonic sequences, while 217028_at identifies 3'UTR in *CXCR4*.

DISCUSSION

Despite extensive research and recent successes in identifying genetic risk variants predisposing to MS, the underlying molecular mechanisms still remain poorly understood. Given the suggested role of autoimmunity and predominance of immunological genes in loci associated with MS, genome-wide expression profiling in immune cells is a valid approach for further elucidating genes and pathways involved in the disease pathogenesis. Although several microarray experiments in MS have been conducted, and a large number of differentially expressed genes have been reported, in most cases the samples have been small, and replication has been lacking, making the findings difficult to interpret. While no obvious consistencies have emerged from these studies, there have not been any systematic attempts to evaluate the overlap between them. We therefore conducted a systematic review of seven microarray studies including our previously unpublished study and confirmed that the general overlap between the studies was indeed poor: only 229 of all 3574 (6%) genes reported to be differentially expressed in MS had been identified in at least two studies, and 94% were therefore unique to a single study. Only 12 of the 229 DEGs were identified in at least three studies. These include *NEAT1*, which encodes for a non-coding RNA and suppresses the expression of CIITA, an activator of genes within the major histocompatibility complex (MHC) class II locus.[24] Interestingly, the most frequently reported gene, *HSPA1A*, which showed decreased expression in MS in four studies, is also functionally connected to the MHC: it encodes for a heat shock protein, which is likely to be involved in MHC class I and II mediated antigen presentation.[25] The gene itself is located in the MHC class III region next to a highly homologous *HSPA1B* gene, and the measured expression levels may reflect the expression of both genes.

However, as is the case for expression studies in general, the studies are not directly comparable owing to differences in samples, sample sizes and platforms, as well as in criteria used for data quality control and for declaring differential expression. Furthermore, the majority of all reported genes came from the only two studies conducted in whole blood samples, and it is therefore not necessarily surprising that most of these genes were not identified in studies conducted in PBMCs or lymphocyte populations. Two studies also reported only a subset of the identified genes, and some studies were conducted using microarrays covering only a fraction of currently known human genes. However, perhaps the most likely explanation for poor overlap across

Table 5 Non-human leukocyte antigen single nucleotide polymorphisms (SNP) in silico replicated differentially expressed genes with p≤0.0001 in a genome-wide association study

Gene	SNP	Chr:bp position (hg18)	p Value	Genome-wide association study	Direction of change in expression in multiple sclerosis
ANXA1	rs13292677	9:74870791	0.0001	Baranzini et al[2]	Increased[9 11]
	rs1961830	9:74872500	0.0001		
	rs7863238	9:74886984	0.0001		
	rs1342022	9:74895327	0.0001		
	rs2310333	9:74897200	0.0001		
CD40	rs6131010	20:44157712	8.5E−07	ANZgene[5]	Increased (FIN[11])
	rs6074022	20:44173603	1.3E−07		
	rs1569723	20:44175471	2.9E−07		
	rs3746821	20:44188518	9.7E−05	IMSGC[1]	
	rs2425764	20:44233852	2.5E−05		
CDK4	rs10876994	12:56351004	5.4E−11	ANZgene[5]	Decreased[11 15 16]
	rs12368653	12:56419523	2.7E−10		
	rs703842	12:56449006	1.0E−07		
CXCR4	rs4954555	2:136509584	0.0001	Baranzini et al[2]	Increased (FIN[11])
	rs7574456	2:136606529	2.5E−05	De Jager et al[3]	
	rs1519529	2:136690727	4.1E−05		
C7orf54	rs1193335	7:127404040	3.0E−05	De Jager et al[3]	Increased (FIN[9])
GNG2	rs4468527	14:51394881	4.1E−05	De Jager et al[3]	Decreased[9 16]
IL7R	rs931555	5:35839334	2.7E−06	De Jager et al[3]	Increased[12 14]
	rs6897932	5:35910332	1.7E−06		
ITPR1	rs711663	3:4437774	1.6E−05	De Jager et al[3]	Increased (FIN[11])
NPEPPS	rs9901869	17:42930205	2.9E−06	De Jager et al[3]	Increased[11 12]
	rs4239162	17:43110809	6.9E−06		
	rs11079784	17:43057279	3.6E−05	Baranzini et al[2]	
PAK2	rs6583176	3:198009975	3.9E−05	De Jager et al[3]	Decreased[9 15 16]
TGFBR2	rs12490899	3:30649258	5.7E−05	De Jager et al[3]	Decreased[9 15 16]
TNFAIP3	rs892999	6:138180398	3.1E−06	De Jager et al[3]	Increased (FIN[11 15 16])
TNFRSF1A	rs1800693	12:6310270	1.6E−11	De Jager et al[3]	Decreased (FIN[15 16])
	rs4149576	12:6319376	1.0E−08		
TRIB2	rs7607490	2:12768571	1.0E−05	Baranzini et al[2]	Increased[11 12]
ZMIZ1	rs1250540	10:80706013	1.6E−06	De Jager et al[3]	Decreased[9 11]

The best p value is shown if several p values were provided.
FIN, previously unpublished Finnish microarray screen described in detail here.
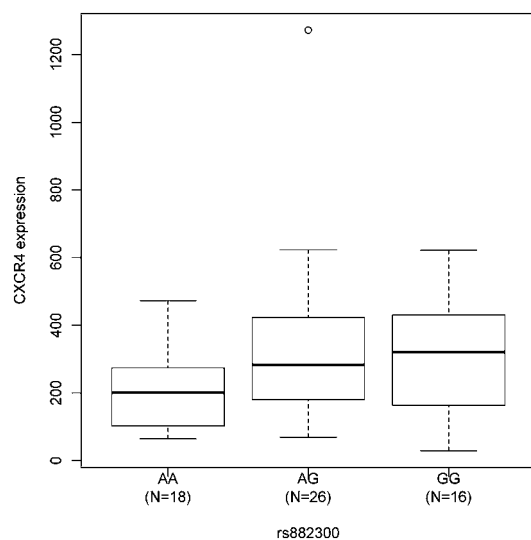

Figure 3 Box plot of *CXCR4* expression and rs882300 genotype in 60 Centre d'Étude du Polymorphisme Humain (CEPH) lymphoblastoid cell samples.

studies is a high rate of false positives and low power to detect true differences in small samples. Recent GWASs conducted in large samples have proven that most of the early genetic associations reported in candidate gene studies of at most a few hundred individuals seem to have been false positives. Small samples may be even more problematic in expression studies, which are susceptible to noise introduced by technical and biological factors. Large studies are required, especially if the aim is to identify expression changes which are due to genetic disease risk variants because the effects of common genetic variants on gene expression are in most cases relatively modest, even in rather homogeneous cell populations,[26] and in small samples the difference in risk allele frequency between cases and controls is not expected to be significant in the first place.

However, after excluding 13 genes in the MHC region, the 229 in silico replicated DEGs were enriched for variants showing a modest association in the IMSGC GWAS,[1] suggesting that at least some of these genes are likely to play a causative role in MS rather than showing

differential expression as a result of activation of immunological pathways secondary to MS. In addition, 15 of these genes have shown suggestive evidence for association with MS (p<0.0001) including *CDK4*, *IL7R* and *TNFRSF1A*, which are located in regions of genome-wide significance.[1] [3] [5] [17] These 15 also include *CD40*, which is associated with rheumatoid arthritis,[27] and *TNFAIP3*, which is associated with coeliac disease, SLE, psoriasis and rheumatoid arthritis.[28–31] *TNFAIP3* was also one of the genes identified as differentially expressed in three studies, including our own experiment.[11] [15] [16] It encodes for a zinc finger protein which inhibits nuclear factor (NF)-κβ activity and tumour necrosis factor (TNF)-mediated programmed cell death, and may therefore play an important role in regulating various immunological pathways.[32] However, apart from *CXCR4*, the expression of the 15 DEGs did not correlate with the proposed risk variants in lymphoblastoid cell lines, although potential eQTL effects should be further investigated in other immune cells populations. In *CXCR4*, the associated SNP correlated modestly with expression when measured by a probe set representing the 3′UTR, which may indicate differential usage of alternative polyadenylation signals. Interestingly, *CXCR4* promotes transendothelial migration of T cells in vitro together with its ligand, *CXCL12*,[33] while *CXCR4* and *CXCR3* antagonists reduce the accumulation of CD4+ T cells in the CNS and inhibit EAE pathogenesis.[34]

We acknowledge that results from our previously unpublished experiment may have been affected by technical factors as well as by the age difference between cases and controls. The study also included four patients who had received treatment. We therefore reviewed the list of DEGs after excluding our study and found that 135 of the 229 DEGs were identified in at least two independent studies. Pathway analysis on both the 229 and 135 in silico replicated DEGs showed that they were highly associated with several immunological pathways. Interestingly, the identified interleukin signalling pathways (IL-4, IL-6 and IL-17) are primarily related to Th2 and Th17 cells rather than Th1 cells, which are thought to mediate MS.[35–37] Further, IL-6 regulates the balance between regulatory T cell and Th17 cell differentiation together with TGF-β.[38] Th17 cells have been linked with autoimmunity, and several studies have provided evidence for their role in MS and EAE,[39] while regulatory T cells have been demonstrated to display loss of suppressive function in MS.[40] Further studies are needed to investigate whether changes in expression of genes in these interleukin signalling pathways are causative or secondary to MS. We also saw a strong association with cancer-related pathways, which may suggest some common molecular mechanisms behind cancer and autoimmunity, such as dysregulation of apoptosis signalling. Finally, the pathway showing most significant association with the in silico replicated DEGs was the glucocorticoid receptor signalling pathway, which is a central regulator of inflammation. Although one could speculate that this may reflect the usage of corticosteroids as a treatment for MS, patients in the included studies had reportedly been untreated shortly prior to sample collection apart from our study where two patients had been treated with cortisone. This pathway was also the most significantly associated when our study, which included treated patients, was excluded. This would suggest that the regulation of endogenous glucocorticoid receptor signalling pathway may be disturbed in MS, which is in concordance with previous evidence of reduced glucocorticoid receptor binding affinity and sensitivity in lymphocytes in MS patients.[41] [42] Furthermore, mice producing an antisense RNA for the glucocorticoid receptor do not develop EAE.[43] Interestingly, several of the genes in confirmed MS risk loci are connected to the glucocorticoid receptor signalling pathway, including *STAT3*, which was recently identified through a GWAS by our group[4] and acts as a co-activator of glucocorticoid receptor signalling.[44]

In conclusion, we have performed the first systematic review of microarray studies in MS. In general, there was little overlap between the seven studies investigated, most likely owing to the small sizes of these studies. However, 229 genes were reported to be differentially expressed in MS in at least two studies. After excluding our unpublished experiment, which may have been affected by confounding factors and inclusion of treated subjects, 135 genes were identified in at least two studies. Of the 229 genes, 12 were reported in at least three studies, including *TNFAIP3*, which is associated with several other autoimmune diseases, and *NEAT1* and *HSPA1A*, which are both functionally connected to MHC, the major MS susceptibility locus. Pathway analyses on the 229 and 135 DEGs provided support for the involvement of glucocorticoid receptor signalling and Th2, Th17 and regulatory T-cell-related interleukin signalling pathways in MS. Together with accumulating data from genetic association studies, our findings can be helpful in selecting genes and pathways for further functional studies in MS.

Author affiliations
[1]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland
[2]Helsinki Biomedical Graduate School, University of Helsinki, Helsinki, Finland
[3]Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
[4]Department of Public Health, University of Helsinki, Helsinki, Finland
[5]Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland
[6]Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[7]Department of Medical Genetics, University of Helsinki, Helsinki, Finland
[8]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK
[9]Department of Gynecology and Pediatrics, Helsinki University Central Hospital, Helsinki, Finland
[10]Department of Child Psychiatry, Helsinki University Central Hospital, Helsinki, Finland

REFERENCES

1. Hafler DA, Compston A, Sawcer S, et al; International Multiple Sclerosis Genetics Consortium. Risk alleles for multiple sclerosis identified by a genomewide study. *N Engl J Med* 2007;357:851—62.
2. Baranzini SE, Wang J, Gibson RA, et al. Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. *Hum Mol Genet* 2009;18:767—78.
3. De Jager PL, Jia X, Wang J, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. *Nat Genet* 2009;41:776—82.
4. Jakkula E, Leppä V, Sulonen AM, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. *Am J Hum Genet* 2010;86:285—91.
5. Australia and New Zealand Multiple Sclerosis Genetics Consortium (ANZgene). Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009;41:824—8.
6. Sanna S, Pitzalis M, Zoledziewska M, et al. Variants within the immunoregulatory CBLB gene are associated with multiple sclerosis. *Nat Genet* 2010;42:495—7.
7. Achiron A, Gurevich M. Peripheral blood gene expression signature mirrors central nervous system disease: the model of multiple sclerosis. *Autoimmun Rev* 2006;5:517—22.
8. Brynedal B, Khademi M, Wallström E, et al. Gene expression profiling in multiple sclerosis: a disease of the central nervous system, but with relapses triggered in the periphery? *Neurobiol Dis* 2010;37:613—21.
9. Gandhi KS, McKay FC, Cox M, et al; ANZgene Multiple Sclerosis Genetics Consortium. The multiple sclerosis whole blood mRNA transcriptome and genetic associations indicate dysregulation of specific T cell pathways in pathogenesis. *Hum Mol Genet* 2010;19:2134—43.
10. Achiron A, Gurevich M, Friedman N, et al. Blood transcriptional signatures of multiple sclerosis: unique gene expression of disease activity. *Ann Neurol* 2004;55:410—17.
11. Arthur AT, Armati PJ, Bye C; Southern MS Genetics Consortium. Genes implicated in multiple sclerosis pathogenesis from consilience of genotyping and expression profiles in relapse and remission. *BMC Med Genet* 2008;9:17.
12. Bomprezzi R, Ringnér M, Kim S, et al. Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. *Hum Mol Genet* 2003;12:2191—9.
13. Mandel M, Gurevich M, Pauzner R, et al. Autoimmunity gene expression portrait: specific signature that intersects or differentiates between multiple sclerosis and systemic lupus erythematosus. *Clin Exp Immunol* 2004;138:164—70.
14. Ramanathan M, Weinstock-Guttman B, Nguyen LT, et al. In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. *J Neuroimmunol* 2001;116:213—19.
15. Satoh J, Nakanishi M, Koike F, et al. Microarray analysis identifies an aberrant expression of apoptosis and DNA damage-regulatory genes in multiple sclerosis. *Neurobiol Dis* 2005;18:537—50.
16. Satoh J, Nakanishi M, Koike F, et al. T cell gene expression profiling identifies distinct subgroups of Japanese multiple sclerosis patients. *J Neuroimmunol* 2006;174:108—18.
17. International Multiple Sclerosis Genetics Consortium (IMSGC). Refining genetic associations in multiple sclerosis. *Lancet Neurol* 2008;7:567—9.
18. Dixon AL, Liang L, Moffatt MF, et al. A whole-genome association study of global gene expression. *Nat Genet* 2007;39:1202—7.
19. Moffatt MF, Kabesch M, Liang L, et al. Genetic variants regulating ORMDL3 expression are determinants of susceptibility to childhood asthma. *Nature* 2007;448:470—3.
20. Spielman RS, Bastone LA, Burdick JT, et al. Common genetic variants account for differences in gene expression among ethnic groups. *Nat Genet* 2007;39:226—31.
21. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005;437:1299—320.
22. Airas L, Nikula T, Huang YH, et al. Postpartum-activation of multiple sclerosis is associated with down-regulation of tolerogenic HLA-G. *J Neuroimmunol* 2007;187:205—11.
23. Avasarala JR, Chittur SV, George AD, et al. Microarray analysis in B cells among siblings with/without MS—role for transcription factor TCF2. *BMC Med Genomics* 2008;1:2.
24. Geirsson A, Paliwal I, Lynch RJ, et al. Class II transactivator promoter activity is suppressed through regulation by a trophoblast noncoding RNA. *Transplantation* 2003;76:387—94.
25. Stocki P, Morris NJ, Preisinger C, et al. Identification of potential HLA class I and class II epitope precursors associated with heat shock protein 70 (HSPA). *Cell Stress Chaperones* 2010;15:729—41.
26. Zeller T, Wild P, Szymczak S, et al. Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS One* 2010;5:e10693.
27. Raychaudhuri S, Remmers EF, Lee AT, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nat Genet* 2008;40:1216—23.
28. Graham RR, Cotsapas C, Davies L, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008;40:1059—61.
29. Nair RP, Duffin KC, Helms C, et al; Collaborative Association Study of Psoriasis. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* 2009;41:199—204.
30. Plenge RM, Cotsapas C, Davies L, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat Genet* 2007;39:1477—82.
31. Dubois PC, Trynka G, Franke L, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet* 2010;42:295—302.
32. Lee EG, Boone DL, Chai S, et al. Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science* 2000;289:2350—4.
33. Liu KK, Dorovini-Zis K. Regulation of CXCL12 and CXCR4 expression by human brain endothelial cells and their role in CD4+ and CD8+ T cell adhesion and transendothelial migration. *J Neuroimmunol* 2009;215:49—64.
34. Kohler RE, Comerford I, Townley S, et al. Antagonism of the chemokine receptors CXCR3 and CXCR4 reduces the pathology of experimental autoimmune encephalomyelitis. *Brain Pathol* 2008;18:504—16.
35. Fiorentino DF, Bond MW, Mosmann TR. Two types of mouse T helper cell. IV. Th2 clones secrete a factor that inhibits cytokine production by Th1 clones. *J Exp Med* 1989;170:2081—95.
36. Harrington LE, Hatton RD, Mangan PR, et al. Interleukin 17-producing CD4+ effector T cells develop via a lineage distinct from the T helper type 1 and 2 lineages. *Nat Immunol* 2005;6:1123—32.
37. Park H, Li Z, Yang XO, et al. A distinct lineage of CD4 T cells regulates tissue inflammation by producing interleukin 17. *Nat Immunol* 2005;6:1133—41.
38. Bettelli E, Carrier Y, Gao W, et al. Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* 2006;441:235—8.
39. Segal BM. Th17 cells in autoimmune demyelinating disease. *Semin Immunopathol* 2010;32:71—7.
40. Viglietta V, Baecher-Allan C, Weiner HL, et al. Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. *J Exp Med* 2004;199:971—9.
41. van Winsen LM, Muris DF, Polman CH, et al. Sensitivity to glucocorticoids is decreased in relapsing remitting multiple sclerosis. *J Clin Endocrinol Metab* 2005;90:734—40.
42. Ysrraelit MC, Gaitán MI, Lopez AS, et al. Impaired hypothalamic—pituitary—adrenal axis activity in patients with multiple sclerosis. *Neurology* 2008;71:1948—54.
43. Morale C, Brouwer J, Testa N, et al. Stress, glucocorticoids and the susceptibility to develop autoimmune disorders of the central nervous system. *Neurol Sci* 2001;22:159—62.
44. Zhang Z, Jones S, Hagood JS, et al. STAT3 acts as a co-activator of glucocorticoid receptor signaling. *J Biol Chem* 1997;272:30607—10.