nature genetics

Interleukin 7 receptor α chain (*IL7R*) shows allelic and functional association with multiple sclerosis

Simon G Gregory[1,9], Silke Schmidt[1,9], Puneet Seth[2], Jorge R Oksenberg[3], John Hart[1], Angela Prokop[1], Stacy J Caillier[3], Maria Ban[4], An Goris[5], Lisa F Barcellos[6], Robin Lincoln[3], Jacob L McCauley[7], Stephen J Sawcer[4], D A S Compston[4], Benedicte Dubois[5], Stephen L Hauser[3], Mariano A Garcia-Blanco[2], Margaret A Pericak-Vance[8] & Jonathan L Haines[7], for the Multiple Sclerosis Genetics Group

Multiple sclerosis is a demyelinating neurodegenerative disease with a strong genetic component. Previous genetic risk studies have failed to identify consistently linked regions or genes outside of the major histocompatibility complex on chromosome 6p. We describe allelic association of a polymorphism in the gene encoding the interleukin 7 receptor α chain (*IL7R*) as a significant risk factor for multiple sclerosis in four independent family-based or case-control data sets (overall $P = 2.9 \times 10^{-7}$). Further, the likely causal SNP, rs6897932, located within the alternatively spliced exon 6 of *IL7R*, has a functional effect on gene expression. The SNP influences the amount of soluble and membrane-bound isoforms of the protein by putatively disrupting an exonic splicing silencer.

Multiple sclerosis is the prototypical human demyelinating disease, and numerous epidemiological, adoption and twin studies have provided evidence for a strong underlying genetic liability[1]. The disease is most common in young adults, with more than 90% of affected individuals diagnosed before the age of 55, and fewer than 5% diagnosed before the age of 14. Females are two to three times more frequently affected than males[2], and the disease course can vary substantially, with some affected individuals suffering only minor disability several decades after their initial diagnosis, and others reaching wheelchair dependency shortly after disease onset. The complex etiology of the disease and the currently undefined molecular mechanisms of multiple sclerosis suggest that moderate contributions from multiple risk loci underlie the development and progression of the disease.

Many different approaches, including genetic linkage, candidate gene association and gene expression studies, have been used (summarized in ref. 2) to identify the genetic basis of multiple sclerosis. However, genetic linkage screens have failed to identify consistent regions of linkage outside of the major histocompatibility complex (MHC). Candidate gene studies have suggested over 100 different associated genes, but there has not been consensus acceptance of any such candidates. Similarly, gene expression studies have identified hundreds of differentially expressed transcripts, with little consistency across studies. The alternative approach of genomic convergence[3], which requires multiple sources and types of positive evidence to implicate a candidate gene, can be used to integrate both statistical and functional data.

Using genomic convergence[3], we identified 28 genes that were differentially expressed in at least two of nine previous expression studies (**Supplementary Table 1** online). We focused on three genes (interleukin 7 receptor alpha chain (*IL7R*) [MIM: 146661], matrix metalloproteinase 19 (*MMP19*) [MIM: 601807] and chemokine (C-C motif) ligand 2 (*CCL2*) [MIM: 158105]) that had a previously published or inferred functional role in multiple sclerosis and that were not located within the MHC. Two of the three genes localize to previous regions of genetic linkage on 17q12 (*CCL2*)[4] and 5p13.2 (*IL7R*)[5]. We analyzed a large data set of 760 US families of European descent, including 1,055 individuals with multiple sclerosis, and identified a significant association with multiple sclerosis susceptibility only for a nonsynonymous coding SNP (rs6897932) within a key transmembrane domain of *IL7R*. We subsequently replicated this initial significant association in three independent European populations or populations of European descent: 438 individuals with multiple sclerosis and 479 unrelated controls ascertained in the United States, 1,338 individuals with multiple sclerosis and their parents ascertained in northern Europe (the UK and Belgium) and 1,077 individuals with multiple sclerosis and 2,725 unrelated controls also ascertained in northern Europe. We show that rs6897932 affects

[1]Center for Human Genetics, and [2]Center for RNA Biology and Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina 27710, USA. [3]Department of Neurology, University of California, San Francisco, California 94143, USA. [4]Department of Clinical Neurosciences, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, UK. [5]Section for Experimental Neurology, Katholieke Universiteit Leuven, 3000 Leuven, Belgium. [6]School of Public Health, University of California, Berkeley, California 94720, USA. [7]Center for Human Genetics Research, Vanderbilt University Medical Center, Nashville, Tennessee 37232, USA. [8]Miami Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida 33136, USA. [9]These two authors contributed equally to this work. Correspondence should be addressed to J.L.H. (jonathan@chgr.mc.vanderbilt.edu) or M.A.P.-V. (MPericak@med.miami.edu).

Received 22 February; accepted 18 June; published online 29 July 2007; doi:10.1038/ng2103

Table 1 Clinical characteristics of the family-based and case-control multiple sclerosis data sets

	US data set			European data set		
	Affected individuals from family-based data set (n = 1,055)	Cases (n = 438)	Controls (n = 479)	Affected individuals from family-based data set (n = 1,338)	Cases (n = 1,077)	Controls (n = 2,725)
Mean age (and range), in years			41.6 (21–60)			44.6 (24–61)
Mean age at onset (and range), in years	29.8 (3–63)	31.9 (11–60)		27.1 (9–53)	33.6 (10–62)	
Number of females (and percentage)	784 (74.3)	299 (68.3)	318 (66.4)	987 (74.0)	738 (68.5)	1350 (49.5)
Disease type: n (%)						
Relapsing-remitting	552 (52.3)	365 (83.3)	–	1,152 (86.1)	875 (81.2)	–
Secondary progressive	260 (24.6)	54 (12.3)	–			
Progressive-relapsing	16 (1.5)	3 (0.7)	–	NA	NA	–
Primary progressive	43 (4.1)	14 (3.2)	–	80 (6.0)	146 (13.6)	–
Missing	184 (17.4)	2 (0.5)	–	106 (7.9)	56 (5.2)	–
Mean disease duration (and range), in years	13.6 (0–50)	9.9 (0–44)	–	10.8 (0–39)	16.4 (0–57)	–

Family-based US data set: 760 families (197 with multiple affected relatives (multiplex families), 563 with a single affected individual). Family-based European data set: 1,338 trios. UK affected individuals from trios: 64.8% relapsing-remitting, 27.4% secondary progressive. Unrelated UK affected individuals: 45.9% relapsing-remitting, 32.6% secondary progressive. For Belgian affected individuals, relapsing-remitting and secondary progressive multiple sclerosis were grouped together as 'bout-onset'.

alternative splicing of exon 6, leading to increased skipping of the exon. This is predicted to increase production of soluble interleukin 7 receptor alpha chain (IL7Rα) for individuals carrying the risk allele at rs6897932. Our results strongly implicate the IL7Rα-related immune response pathway in the etiology of multiple sclerosis.

RESULTS

Initial analysis of *CCL2, MMP19* and *IL7R* in US families

None of the SNPs genotyped in *IL7R*, *CCL2* or *MMP19* strongly deviated from Hardy-Weinberg equilibrium in a sample of unrelated affected ($P > 0.04$) or unaffected ($P > 0.10$) individuals ($P = 0.05$ for rs6897932 in affected individuals). We did not find any statistically significant evidence for association with multiple sclerosis for SNPs in *CCL2* and *MMP19* (**Supplementary Table 2** online), although this does not exclude the possibility of a small effect of variants in either gene. In the 760 US families (**Table 1**), the SNP having the strongest association with multiple sclerosis was rs6897932 ($P = 0.0006$ by the pedigree disequilibrium test (PDT)[6]), a nonsynonymous coding SNP (T244I) in exon 6 of *IL7R*, followed by rs11567705 ($P = 0.002$) and rs13188960 ($P = 0.002$) (**Table 2** and **Fig. 1**). The latter SNPs are in strong linkage disequilibrium (LD) with rs6897932 ($r^2 = 0.97$ for rs11567705, $r^2 = 0.95$ for rs13188960). The result for rs6897932 was due to overtransmission of the major allele (C) to offspring affected with multiple sclerosis. The risk allele frequency was similar in individuals with multiple sclerosis who did or did not carry the multiple sclerosis–associated alleles HLA-DRB1*1501/1503 (0.79 and

Table 2 PDT analysis of *IL7R* SNPs

SNP	Position (bp)	Minor/major allele	MAF	Gene region	Type	Amino acid change	Overall PDT P value
rs13188960[a]	35889076	T/***G***	0.24	Upstream			***0.0023***
rs7718919	35891753	T/G	0.13	Upstream			0.8707
rs1389832[a]	35894478	T/C	0.34	Intron 1			0.0804
rs1494558[a]	35896825	A/G	0.34	Exon 2	Nonsynonymous	I66T	0.0249
rs11567705	35896909	G/***C***	0.24	Intron 2			***0.0018***
rs969128	35896916	G/A	0.13	Intron 2			0.9343
rs1494555[a]	35906947	C/T	0.33	Exon 4	Nonsynonymous	V138I	0.0327
rs7737000	35907030	T/C	0.13	Exon 4	Synonymous	H165H	0.6801
rs1494554[a]	35909629	C/A	0.29	Intron 5			0.7175
rs6897932[a]	35910332	T/***C***	0.24	Exon 6	Nonsynonymous	T244I	***0.0006***
rs987107[a]	35910984	T/C	0.29	Intron 6			0.7856
rs987106[a]	35911350	T/A	0.47	Intron 6			0.0134
rs3194051[a]	35912031	G/A	0.29	Exon 8	Nonsynonymous	I356V	0.7856
rs1494571	35915844	C/G	0.28	Downstream			0.9276

P values meeting the Bonferroni-corrected significance threshold and respective risk allele are shown in boldface italic. Minor allele frequency (MAF) estimated from genotyped founders.
[a]SNPs genotyped by the HapMap project.

0.78), suggesting that the *IL7R* effect on multiple sclerosis susceptibility is independent of the known HLA effect. A quantitative transmission disequilibrium test (QTDT) analysis[7,8] showed absence of association between genotypes at rs6897932 and age at onset of multiple sclerosis ($P = 0.88$) or multiple sclerosis severity ($P = 0.19$ for multiple sclerosis severity scores)[9,10].

The P value of 0.0006 for rs6897932 was significant after accounting for multiple comparisons using a Bonferroni correction for the effective number of independent SNPs evaluated across the three genes[11]. The effective number was 4 for *CCL2* (5 genotyped SNPs), 3 for *MMP19* (4 genotyped SNPs) and 8.7 for *IL7R* (14 genotyped SNPs), for an adjusted significance threshold of 0.05 / (4 + 3 + 8.7) = 0.003.

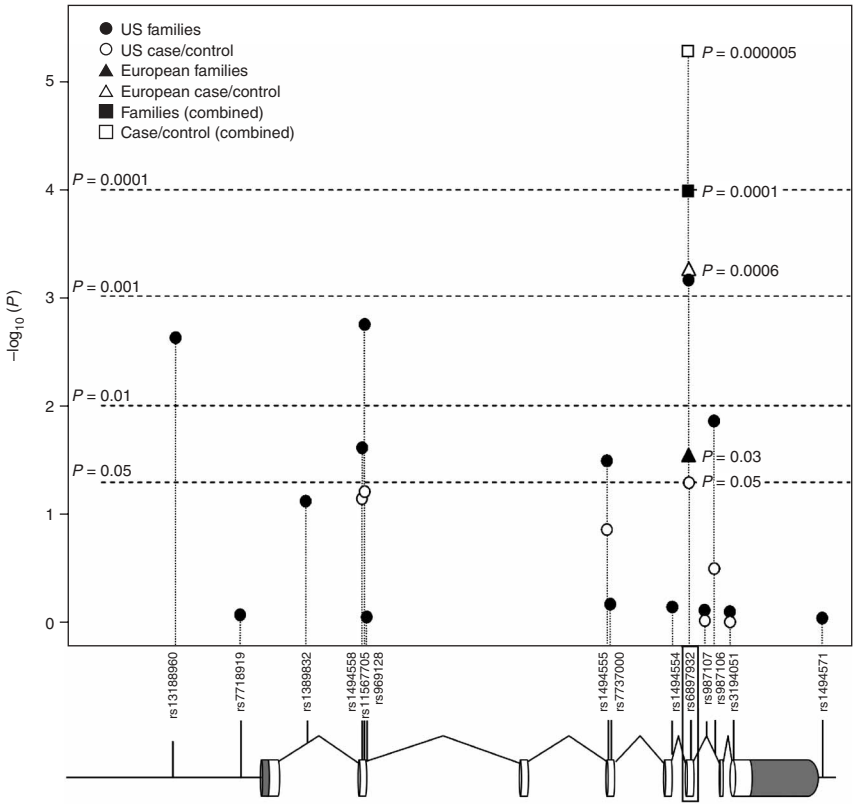

Figure 1 Physical map and association analysis of SNPs from *IL7R* in US, European and combined family-based and case-control data sets. Untranslated regions of *IL7R* are represented by gray filled regions and exons by unfilled regions in the schematic below the graph. The significant SNP, rs6897932, is outlined.

LD and haplotype analysis of *IL7R*

IL7R is located within a block of very high LD: all pairwise D′ values between SNPs in this block are ≥0.98. To identify common haplotypes and a subset of haplotype-tagging SNPs within *IL7R*, we analyzed haplotypes of all 14 genotyped SNPs for association with multiple sclerosis susceptibility. We found only four common haplotypes (>5% estimated frequency in family founders) in our data set, and a subset of four SNPs (rs1494555, rs6897932, rs987107, rs987106) was sufficient to distinguish between these four common haplotypes. We observed a significant association with multiple sclerosis susceptibility for haplotypes formed by these four tagging SNPs (global P value = 0.0007; **Table 3**). The association was due to overtransmission of haplotype C-C-C-A to offspring affected with multiple sclerosis (haplotype-specific $P = 0.006$) and undertransmission of haplotype T-T-C-T (haplotype-specific $P = 0.00005$). These two haplotypes are distinguished by the common risk allele (C) at the coding SNP rs6897932 (single-locus $P = 0.0006$), the C allele at rs1494555 (single-locus $P = 0.03$) and the T allele at rs987106 (single-locus $P = 0.01$). An evaluation of our four haplotype-tagging SNPs using the HapMap data for the population of European descent (CEU) showed that the SNPs genotyped in our study capture the majority (89%) of common sequence variation in *IL7R*, with a mean r^2 of 0.944. Furthermore, the HapMap data also showed that the very strong LD block that spans *IL7R* does not extend to any flanking genes, indicating that the only functional element within the LD block is *IL7R*. None of the HapMap-genotyped coding SNPs from any of the flanking genes is highly correlated ($r^2 > 0.8$) with the most strongly associated SNP rs6897932.

Replication of the rs6897932 effect

We genotyped an independent case-control data set of European descent ascertained in the US (438 cases and 479 controls; **Table 1**) for seven SNPs, including the four haplotype-tagging SNPs in *IL7R* identified by the family-based analysis (**Table 4**). Analysis by logistic regression, with adjustment for age and gender, demonstrated statistical significance only for rs6897932 ($P = 0.05$). In the subset

Table 3 Haplotype analysis of *IL7R* in 760 multiple sclerosis families of European descent, using four adjacent SNPs that distinguish all common haplotypes

	rs1494555	rs6897932	rs987107	rs987106	Freq.	Global *P*-value	No. informative families	*Z*	Haplotype-specific *P*-value
Haplotype 1	C	C	C	A	0.35	0.0007	284	2.99	0.006
Haplotype 2	T	C	T	T	0.27		264	0.05	0.96
Haplotype 3	T	T	C	T	0.23		245	–3.98	0.00005
Haplotype 4	T	C	C	A	0.14		180	0.79	0.43

Haplotype frequencies (freq.) were estimated from genotyped founders. The number of haplotype-specific informative families indicates how often transmission of haplotypes by heterozygous parents could be inferred. The *Z* statistic, calculated by the HBAT software using the 'hbat –e' command, is representative of overtransmission ($Z > 0$) or undertransmission ($Z < 0$) of the respective haplotype.

Table 4 Logistic regression analysis of four haplotype-tagging *IL7R* SNPs in 438 individuals with multiple sclerosis and 479 controls (US data set)

SNP	Allele	Frequency in individuals with MS US (Europe)	Frequency in controls US (Europe)	Genotype	Frequency in individuals with MS US (Europe)	Frequency in controls US (Europe)	*P* value[a] US	*P* value[a] Europe
rs1494555	C	0.349	0.312	CC	0.105	0.082	0.14	–
	T	0.651	0.688	CT	0.489	0.461		
				TT	0.406	0.457		
rs6897932	T	0.217 (0.238)	0.265 (0.283)	TT	0.039 (0.055)	0.065 (0.083)	0.05	0.0006
	C	0.783 (0.762)	0.735 (0.717)	CT	0.357 (0.368)	0.401 (0.399)		
				CC	0.605 (0.577)	0.534 (0.518)		
rs987107	T	0.284	0.274	TT	0.105	0.069	0.94	–
	C	0.716	0.726	CT	0.358	0.408		
				CC	0.537	0.522		
rs987106	T	0.495	0.473	TT	0.257	0.220	0.32	–
	A	0.505	0.527	AT	0.477	0.505		
				AA	0.266	0.275		

	Genotype	Odds ratio (95% c.i.) US	Odds ratio (95% c.i.) Europe	Odds ratio (95% c.i.) Combined	*P* value[b] US	*P* value[b] Europe	*P* value[b] Combined
rs6897932	TT, CT	1.0 (reference)	1.0 (reference)	1.0 (reference)	–	–	–
	CC	1.34 (1.03–1.74)	1.24 (1.08–1.44)	1.29 (1.14–1.46)	0.03	0.003	0.00008

Data for rs6897932 includes European affected individuals ($n = 1,077$) and controls ($n = 2,725$).
[a]Additive coding of SNP genotypes. [b]Recessive coding.

of 380 unrelated individuals with multiple sclerosis with available data on carrier status for HLA-DRB1*1501/1503, risk allele frequencies at rs6897932 were similar in carriers and noncarriers of the multiple sclerosis–associated haplotype (0.76 versus 0.81, $P = 0.11$). **Table 5** shows estimated joint odds ratios (ORs) for carriers of HLA-DRB1*1501/1503 and carriers of the C/C genotype at rs6897932, based on 380 individuals with multiple sclerosis and 428 controls with available data for both loci. In this data set, the population-attributable risk percent (PAR%)[12] was 40.1% for HLA-DRB1*1501/1503 and 16.4% for rs6897932 at *IL7R*, when adjusting for the other factor. The estimated summary PAR% for both loci was 49.6%, which is less than the sum of the adjusted PAR estimates, as exposures were not mutually exclusive.

Consistent with Wald χ^2 statistics from the logistic regression analysis, permutation-based single-locus tests for the four haplotype-tagging SNPs were significant only for rs6897932 ($P = 0.01$ with WHAP software[13]). In contrast to the family-based analysis, an omnibus haplotype test based on the four common haplotypes was not significant ($P = 0.17$). However, the same haplotype that was overtransmitted in the families (C-C-C-A) was more frequent in affected individuals (35.0%) than in controls (31.6%), whereas the undertransmitted haplotype (T-T-C-T) was less frequent in affected individuals (21.8%) than in controls (25.8%) (**Table 3**). These results suggest that rs6897932 is the strongest single-locus effect within *IL7R* and is responsible for the haplotype association observed in the family-based data set.

Given these findings, we concentrated further replication efforts on rs6897932, which we genotyped in two independent northern European populations ascertained in the UK and Belgium (**Table 1**). The PDT analysis confirmed a significant association with multiple sclerosis risk in 1,338 European families of individuals with multiple sclerosis and their parents (trios) ($P = 0.03$). Logistic regression analysis of 1,077 independent European individuals with multiple sclerosis and 2,725 unrelated controls also replicated the results from the US data sets ($P = 0.0006$). The minor allele frequencies were very similar in US and European individuals with multiple sclerosis (21.7% and 23.9%; **Table 4**) and US and European controls (26.5% and 28.3%; **Table 4**). A combined analysis of all 2,098 families by the PDT ($P = 0.0001$) and a combined logistic regression analysis of all 1,512 independent individuals with multiple sclerosis and 3,184 unrelated controls ($P = 0.000005$ for additive SNP genotype coding, adjusted for ascertainment site, US versus Europe) provided strong statistical support for a genuine association of *IL7R* with multiple sclerosis risk (**Table 4** and **Fig. 1**). A combined analysis of all the family-based and case-control data sets resulted in $P = 2.9 \times 10^{-7}$. With T/T and C/T genotypes coded as the reference group, the OR for C/C genotypes was 1.29 (95% c.i., 1.14–1.46; **Table 4**). For this effect size and a risk allele frequency of 0.72 in controls, a data set of 980 cases and 980 controls is needed to provide 80% power at significance level $\alpha = 0.05$ for replicating the association observed here. Although this effect was not seen in the subset of individuals ($n = 185$) with primary progressive multiple sclerosis ($P > 0.30$), the

Table 5 Estimated joint odds ratios and 95% confidence intervals for carriers of HLA-DRB1*1501/1503 and CC genotypes at rs6897932 in *IL7R*

Genotype at rs6897932 in *IL7R*	Noncarrier of HLA-DRB1*1501/1503	Carrier of HLA-DRB1*1501/1503
TT, CT	1.0 (reference)	3.95 (2.87–5.43)
CC	1.37 (1.02–1.84)	5.42 (3.45–8.51)

Results are from US data set of 380 affected individuals and 428 controls, with available data for both loci. Numbers in parentheses represent 95% c.i.

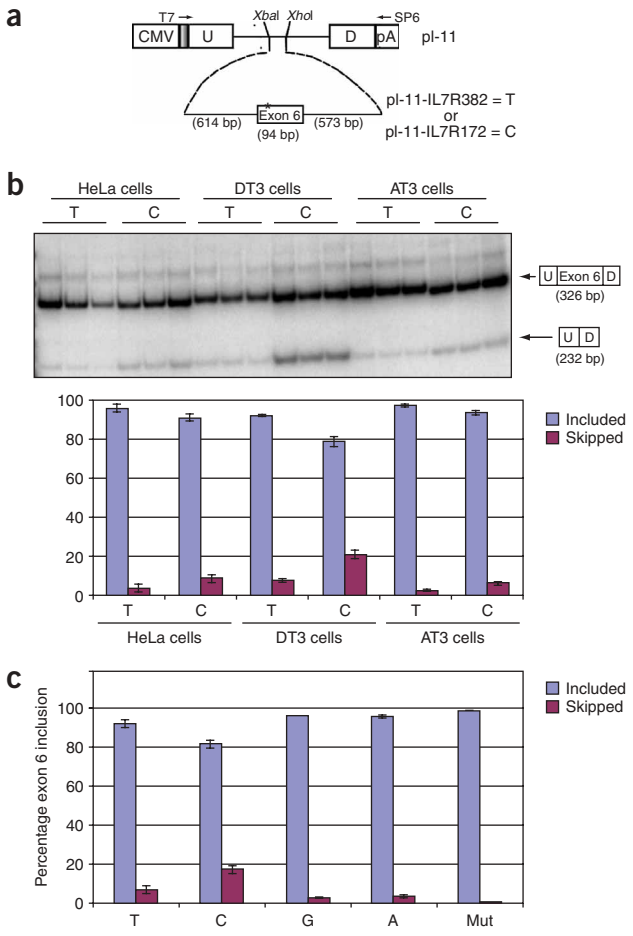

Figure 2 Transfections and splicing analysis of *IL7R* minigenes. (**a**) Splicing construct pI-11 and insertion of *IL7R* genomic sequence to create minigenes pI-11-IL7R382 and pI-11-IL7R172, which contain the 'T' and 'C' alleles of rs6897932, respectively. The location of rs6897932 is shown by an asterisk. CMV indicates the CMV promoter; pA indicates the bovine growth hormone polyadenylation sequence. *Xba*I and *Xho*I sites were used for cloning. The locations of T7 and SP6 vector-specific primers are indicated by arrows. U is the 5′ exon of pI-11, and D is the 3′ exon of pI-11. Sizes of exon 6 and parts of its flanking introns are shown in parentheses. (**b**) The *IL7R* minigenes pI-11-IL7R382 and pI-11-IL7R172 containing the 'T' allele and 'C' allele, respectively, of rs6897932, which were stably transfected in triplicate in each of HeLa, DT3 and AT3 cells. Mature minigene transcripts that either include or skip exon 6 produce 326-bp or 232-bp RT-PCR products. Error bars represent s.d. (**c**) The percentage of exon 6 inclusion is plotted against the nucleotide found at the rs6897932 locus. Exon 6 inclusion was also determined using a minigene where the 3 nt upstream and 3 nt downstream of the T at the rs6897932 were substituted by transversions (mut). Error bars represent s.d. of triplicate transient transfections.

power to detect the OR observed in relapsing-remitting multiple sclerosis is only 25%.

Sequence analysis of *IL7R*

We sequenced genomic DNA for exons 2, 3, 4, 5, 6, 7 and 8 and an upstream region (containing putative regulatory elements) from 16 phenotypically normal and 16 affected individuals to identify any new SNPs within *IL7R*. Although we did not identify any additional new sequence polymorphisms within the 64 chromosomes tested, we cannot discount the possibility that a rare variant with an allele frequency of $<5\%$ may contribute to the phenotype.

The associated T244I variation is located in an important transmembrane domain of IL7Rα. Multispecies sequence alignment in Ensembl of the portion of IL7Rα encoded by exon 6 showed that the polar threonine is conserved between chimp, macaque, dog, rabbit, elephant, hedgehog and bull, indicating that there is a level of evolutionary constraint in nucleotide diversity at the location of rs6897932 in the gene (data not shown).

The effect of rs6897932 on differential splicing of *IL7R*

The alternative splicing of *IL7R* has potent consequences for the function of the receptor, as transcripts that include exon 6 encode a membrane-bound IL7Rα, whereas transcripts that skip exon 6 encode a soluble form of the protein[14]. To test whether the alternate 'C' and 'T' alleles of rs6897932 could affect the inclusion of exon 6, we cloned the two allelic versions of exon 6 and the surrounding intronic sequences into the alternative splicing minigene construct pI-11 (ref. 15). Analysis of transcripts from the two minigenes showed that the multiple sclerosis–associated 'C' allele resulted in an approximately twofold increase in the skipping of exon 6 compared with transcripts containing the 'T' allele (**Fig. 2**). The data are consistent with at least two possible mechanisms: the multiple sclerosis–associated 'C' allele weakens an exonic splicing enhancer (ESE), or it augments an exonic splicing silencer (ESS). The latter is more likely, as transversions at rs6897932 to either a G or an A nucleotide substantially decreased exon skipping (for example, production of soluble IL7Rα), and substitutions adjacent to the SNP essentially abolished skipping (**Fig. 2**). Thus, our results are consistent with the presence of a weak ESS that mediates a low level of exon 6 skipping, resulting in about 10% soluble IL7Rα. Our data suggest a functional role for the alternative rs6897932 alleles on the splicing of exon 6 in *IL7R*.

Allele-specific expression of *IL7R*

We analyzed peripheral blood mononuclear cells from 94 healthy controls by quantitative real-time PCR to detect differential expression of *IL7R* using one amplicon that spanned exons 6–7 and two control amplicons spanning exons 4–5 and exons 7–8, respectively. Using analysis of variance with Bonferroni correction, we observed significantly lower expression of the exon 6–7 amplicon for carriers of the multiple sclerosis–associated 'C' allele ($P = 0.015$); the two control amplicons (exons 4–5 and 7–8) did not show significantly different expression by genotype. Consistent with our *in vitro* experiments, this suggests that carriers of the 'C' allele at rs6897932 produce less membrane-bound IL7Rα protein compared with carriers of the 'T' allele, leading to a further increase of the soluble form of IL7Rα. This hypothesis is supported by our previous report that interleukin-7 (IL7) function and IL7Rα expression in T cells is reduced in individuals with multiple sclerosis[16].

DISCUSSION

Traditional methods for identifying the genetic basis of complex human disease have often been successful when a single, relatively common locus contributes strongly to the phenotype[17,18]. HLA-DRB1 is one locus that carries such an effect in multiple sclerosis[19]; however, it explains less than 50% of the total genetic basis of the disease[20]. Identifying the remaining genetic effects in multiple sclerosis has been hampered by allelic, locus and phenotypic heterogeneity. Traditionally, genetic linkage, expression and association studies have been used

singly to identify these additional genetic components. In this study, we have used a potentially more powerful approach that integrates these previous results and have identified three candidate genes for detailed study. We have found significant association with multiple sclerosis susceptibility only for *IL7R*, which has been found to be differentially expressed in two independent multiple sclerosis studies[21,22]. This gene localizes to a region of genetic linkage on human chromosome 5p12–14 (refs. 5,23) and mouse chromosome 15A2 (ref. 24) and is functionally relevant to the etiology of the disease.

The common 'C' allele of the nonsynonymous coding SNP rs6897932 (T244I) in exon 6 of *IL7R* is strongly associated with increased multiple sclerosis risk, with consistent results in four independent data sets, including two family-based and two case-control studies. Although our sequencing of *IL7R* did not identify any new variants, it remains a formal possibility that a rare allele in LD with rs6897932 is the susceptibility allele. However, such a rare allele would occur in only a small minority of individuals with the rs6897932 risk allele and would have to have a very strong effect. In addition, our finding that rs6897932 changes the expression of the soluble versus membrane-bound form of the protein provides functional evidence for a direct, rather than indirect, disease association, given that the levels of these isoforms are important in regulating the IL7 signal transduction pathway.

Two previous expression analyses have found that *IL7R*, among many other genes, is overexpressed in peripheral blood mononuclear cells of individuals with multiple sclerosis, compared with controls[21,22]. It is unclear if these results describe expression of the membrane-bound receptor, the soluble protein or both isoforms of IL7Rα. However, our results suggest that these studies detected increased levels of the soluble form of IL7Rα due to a greater frequency of the 'C' allele of rs6897932 in individuals with multiple sclerosis. The seven additional expression studies included in our convergent approach did not report IL7Rα as being differentially expressed. This may be due to variability in the stage, type and number of individuals with multiple sclerosis that were sampled; the nonreproducibility of the disease etiology in the experimental autoimmune encephalomyelitis (EAE) model organism; the gene content of the expression array or the protocol used for generating the data.

Candidate gene analyses from two previous studies[25,26] have suggested a role for IL7Rα in multiple sclerosis. The first study[25] identified a suggestive but not statistically significant effect ($P = 0.1$) of a promoter region haplotype in *IL7R* when only DRB1*1501-positive multiple sclerosis cases were considered. The second study[26], published after initiation of this study, examined 66 candidate genes and found SNPs within *IL7R* to be statistically significant in a multiple sclerosis case-control population, both by single-locus and haplotype analysis. Our family-based analysis confirmed the previous result for rs987106 in intron 6 of *IL7R* ($P = 0.01$ in **Table 2**, $P = 0.04$ in ref. 26). However, the most significant nonsynonymous coding SNP from ref. 26, rs3194051 in exon 8 of *IL7R*, was not significant in our data set ($P = 0.79$ in **Table 2**). The most significant polymorphism from our study, rs6897932, was not genotyped in these two previous association studies. The combination of our genetic and molecular data thus suggests that the previously identified association of *IL7R* with multiple sclerosis is driven by the functional effect of rs6897932.

To investigate the role of membrane-bound and soluble isoforms of IL7Rα, others[27] have analyzed differential levels of IL7Rα expression in a sample of 12 individuals with multiple sclerosis and eight controls who had alternative alleles of a promoter-region polymorphism. IL7Rα was significantly differentially expressed in individuals with primary progressive multiple sclerosis, and the addition of more individuals with multiple sclerosis improved the significance of the associated promoter polymorphism that was suggested in the earlier study by the same group[25]. However, the promoter haplotype trended toward significance only when correlated with levels of soluble IL7Rα.

IL7R represents a plausible candidate gene for the development of multiple sclerosis, as it is important in B and T cell differentiation and maintenance via IL7 and thymic stromal lymphopoietin (TSLP) signaling pathways. IL7Rα mediates IL7 signaling by dimerizing with the cytokine receptor common γ (γ_c) chain to form the IL7 transmembrane receptor. The availability of membrane-bound IL7Rα is the limiting factor in IL7 receptor formation. Alternative splicing of exon 6, which leads to the mutually exclusive production of membrane-bound IL7 receptor or a soluble form of IL7Rα, also regulates IL7 signaling via cellular secretion of the soluble isoform and binding to IL7 in solution[14].

Functionally, IL7Rα is important in the body's innate (immediate but nonspecific) and adaptive (pathogen-specific 'memorized') inflammatory and immunological response. IL7 signaling is critical for T cell differentiation of $CD4^-$ $CD8^-$ thymocytes and has a role in survival of $CD4^+$ $CD8^+$ cells after positive selection[28,29]. IL7 signaling also modulates peripheral memory or naive T cell homeostasis (as reviewed in ref. 30). Resting T cells express low levels of IL7Rα that allow them to transduce the signal of trace amounts of IL7 to maintain survival and proliferation. In this context, it is believed that IL7 signaling contributes to the maintenance of low-affinity T cells in homeostatic peripheral expansion conditions[31]. Because of the very low levels of IL7Rα expression in T cells, any further decrease in IL7Rα expression or function may have a significant effect on IL7 signaling and T cell regulation and maintenance. Our results suggest that the multiple sclerosis–associated 'C' risk allele of *IL7R* would probably decrease membrane-bound expression of IL7Rα.

IL7Rα is also important in the TSLP signaling pathway, via dimerization with the product of the cytokine receptor–like factor 2 gene (*CRLF2*) to form the TSLP receptor[32]. TSLP is an IL-7–like cytokine that drives immature B cell development *in vitro* and, in myeloid dendritic cells, can promote naive $CD4^+$ T cells to differentiate into a T helper type 2 (Th2) phenotype and promote the expansion of $CD4^+$ Th2 memory cells[33]. TSLP is thought to trigger dendritic cell–mediated Th2-type inflammatory responses and is considered to be a master switch for allergic inflammation[34], which is relevant to the etiology of multiple sclerosis.

The complex role of IL7Rα in T cell development[28], combined with the hitherto undefined molecular mechanisms of multiple sclerosis, suggests several possible means by which aberrant or suboptimal functioning of IL7Rα may be implicated in the disease. A twofold increase in soluble IL7Rα could have marked consequences for immune activity, and even small changes in alternative splicing can explain substantial phenotypic diversity[35]. These small changes in IL7Rα isoform ratios may affect IL7 signaling and enhance antigenic T cell response to myelin basic protein and myelin oligodendrocyte glycoprotein[36], both of which have been implicated in the development of multiple sclerosis. Alternatively, individuals who carry the risk allele might have a compromised response to infection because of reduced B and T cell differentiation and initiation of multiple sclerosis, as there is evidence that the disease is mediated through bacterial or viral inflection[37]. As the risk allele is quite common in the general population (with a frequency of 0.72–0.74 in individuals of European

descent), it is plausible that additional triggers are required for the development and progression of multiple sclerosis, which is consistent with a complex disease model in which multiple genes and environmental factors contribute to the phenotype. On the basis of our findings, we hypothesize that additional genes involved in hematopoiesis, particularly genes such as the interleukin receptors, which mediate endogenous interleukin signaling, will also be important in the pathology of multiple sclerosis.

METHODS

US family data set. The data set consisted of 760 stringently ascertained families with multiple sclerosis (563 single-case and 197 multiple-case families), which included 1,055 sampled affected individuals with multiple sclerosis. All known ancestors were non-Hispanic individuals of European descent. Diagnostic criteria and ascertainment protocols have been summarized elsewhere[38,39]. All affected family members were examined or had their medical records reviewed by one of the authors (S.L.H.); 83% could be classified by disease subtype as having relapsing-remitting (RR), secondary progressive (SP), progressive-relapsing (PR) or primary progressive (PP) multiple sclerosis. SPMS was defined by six months of worsening neurological disability not explained by relapse and measured as a deterioration of either (i) one or more points on the expanded disability status scale (EDSS) in affected individuals with an EDSS <6 or (ii) a one-half point or more for those with an EDSS >6. PPMS was defined by the presence of both (i) progressive clinical worsening for more than 12 months from symptom onset without any relapses and (ii) abnormal cerebrospinal fluid (CSF), as defined by the presence of two or more oligoclonal bands or an elevated IgG index in the CSF. If acute relapses were superimposed on this steadily progressive course, individuals were considered to have PRMS.

The family collection has been ongoing for >20 years, and disease subtype information is not available on some of the first individuals with multiple sclerosis collected (17%). The protocol was approved by the Committee on Human Research at the University of California, San Francisco, and informed consent was obtained from all participants. A clinical and demographic description of the individuals with multiple sclerosis is shown in **Table 1**.

US case-control data set. The case-control data set was ascertained through a prospective study of phenotype-genotype-biomarker associations in a large cohort of individuals with multiple sclerosis followed longitudinally (**Table 1**). Individuals with multiple sclerosis (ages 18–65 years) who were evaluated in the Multiple Sclerosis Center at the University of California, San Francisco between July 2004 and September 2005 were invited to participate. This study preferentially recruited individuals with a recent onset of multiple sclerosis, although individuals with all clinical subtypes (RRMS, SPMS, PRMS and PPMS) participated. Therefore, there are more individuals with RRMS in the case-control data set, whereas many individuals with RRMS in the family-based data set have progressed to SPMS since the time at which they were first collected. Affected individuals were asked to invite a healthy control of the same sex and ancestry and similar age range to participate in the study. The protocol was approved by the Committee on Human Research at the University of California, San Francisco, and informed consent was obtained from all participants.

European family and case-control data sets. In the UK, trios (an affected individual and both parents) and unrelated affected individuals were recruited from across the country as part of an ongoing study of genetic susceptibility to multiple sclerosis. In Belgium, trios and unrelated cases were recruited from the region of Flanders. Diagnostic criteria were identical to those used in the US, except that progressive-relapsing multiple sclerosis was not distinguished from secondary progressive multiple sclerosis. In Belgium, relapsing-remitting and secondary progressive multiple sclerosis were grouped together as 'bout-onset'. All individuals gave written informed consent, and the study was approved by the Thames Valley Multicentre Research Ethics Committee (UK) and the ethics committee of the University Hospital Gasthuisberg in Leuven (Belgium), respectively. Unrelated UK control samples were obtained from two sources: the Human Random Control (HRC) DNA panel (derived from blood donors) supplied by the European Collection of Cell Cultures (479 individuals) and the British 1958 Birth Cohort (2,047 individuals). Belgian controls ($n = 199$) were hospital staff and individuals with noninflammatory neurological conditions.

Candidate gene and SNP selection. We compiled a list of 323 genes that were differentially expressed in nine previously published gene expression analyses[21,22]. Twenty-eight genes were differentially expressed in at least two of the nine studies (**Supplementary Table 1**), with three of these genes, matrix metalloproteinase 19 (*MMP19*) [MIM 601807], chemokine (C-C motif) ligand 2 (*CCL2*) [MIM 158105] and interleukin receptor 7 alpha chain (*IL7R*) [MIM 146661] having a previously published[25,26,40,41] or inferred functional role in multiple sclerosis. Two of the genes also localized to previous regions of genetic linkage: *CCL2* (17q12)[4] and *IL7R* (5p12–14)[5,23]. SNPselector[42], which incorporates genotype data and LD structure from the HapMap project, was used to identify SNPs within exons, untranslated regions and conserved (putative regulatory) regions within these genes.

Genomic DNA was extracted from whole blood using standard procedures. We genotyped five common SNPs within *MMP19*, four SNPs within *CCL2* and fourteen SNPs within *IL7R*, including all significant SNPs in *IL7R* from a previous study[26]. All SNP genotypes were generated by the TaqMan allelic discrimination assay on an ABI7900HT genotyping platform, using the Assay-on-Demand or Assay-by-Design service from Applied Biosystems. Duplicate quality control samples were placed both within and across plates, and matching genotypes were required for the genotype data to be included in the analysis. Laboratory personnel were unaware of pedigree structure, affection status and location of quality control samples. All SNPs were required to have at least 95% genotyping efficiency, and SNP rs6897932 had 97% efficiency.

Statistical methods. A small number of SNP genotypes causing unresolvable mendelian inconsistencies, as identified by the program PEDCHECK[43], and those implying greater-than-expected recombination rates given the small physical distances between SNPs, as identified by the MERLIN program[44], were eliminated from the analysis. From these data, we estimate that our residual error rate is $<0.3\%$. All SNPs were tested for Hardy-Weinberg equilibrium in separate samples of unrelated affected and unaffected individuals, using an exact test implemented in the GDA program[45]. Measures of LD were calculated with the GOLD software[46].

To compare our genotyped SNPs with the currently known common sequence variation in *IL7R*, we downloaded all SNP genotypes generated by the HapMap project on the CEU population, requiring a minor allele frequency $\geq 10\%$ and genotyping efficiency $\geq 90\%$. The algorithm implemented in Tagger software[47] indicated that our four haplotype-tagging SNPs were individually, or in combinations of up to three SNPs, highly correlated with 32 of the 36 variant alleles (89%) identified by the HapMap project, with a mean r^2 of 0.944. This suggests that these four SNPs captured the majority of common sequence variation in *IL7R*. Of the 36 variant alleles identified by the HapMap project, four were untaggable with these four SNPs, two were captured with $r^2 = 0.65$, one was captured with $r^2 = 0.54$, one was captured with $r^2 = 0.38$ and the other 28 were perfectly correlated with $r^2 = 1.0$.

For the family-based data sets, we used the PDT[6] for single-locus association testing, and the HBAT module of the FBAT package[48] for haplotype analysis (using the 'hbat –e' command line option). The SNP-specific maximum number of informative discordant sibling pairs in the US data set was 807 (395 from multiplex and 412 from singleton families), and the maximum number of informative parent-offspring trios in the US data set was 479 (151 from multiplex and 328 from singleton families). Multiple sclerosis severity scores (MSSS) were calculated as described previously[10]. Age of onset and MSSS were analyzed with the QTDT package[8], using the Monks-Kaplan test[7]. The case-control data sets were analyzed by logistic regression (SAS), using an additive allele coding of genotypes (0 for homozygous carriers of the major allele, 1 for heterozygous genotypes, 2 for homozygous carriers of the minor allele) as a screening test for all SNPs. The genotype frequencies in multiple sclerosis cases and controls suggested a recessive model for SNP rs6897932 at IL7Rα, which was used for the joint analysis of the US and European data sets (**Table 4**), for the combined analysis of HLA and *IL7R* (**Table 5**) and for the

power calculations performed with QUANTO[49]. The *P* values from the logistic regression model were very similar with and without adjustment for age (continuous) and gender (categorical). Permutation-based association testing and haplotype analysis were performed with the WHAP program[13]. The method implemented in WHAP uses a weighted maximum-likelihood approach to allow for ambiguity in statistically inferred haplotypes.

SNP detection by exon resequencing. We carried out *de novo* SNP identification by sequencing exons 2, 3, 4, 5, 6, 7 and 8 of *IL7R* in 16 affected individuals (14 RR, 1 PP and 1 PR), 5 unaffected sibs and 11 parents using an Applied Biosystems 3730 sequencer and manufacturer-recommended sequencing protocols. Using Primer3 software, we designed primers to completely incorporate the exons, intron boundaries and intronic sequence adjacent to each exon. Primer sequences are presented in **Supplementary Table 3** online.

Analysis of exon 6 inclusion in cell culture. To amplify exon 6 of *IL7R* via PCR, we used genomic DNA from two phenotypically normal individuals who were homozygous for the 'C' or 'T' alleles of rs6897932. The last 614 nucleotides of intron 5, the entire exon 6 and the first 573 nucleotides of intron 6 of *IL7R* were amplified and cloned between the *Xba*I and *Xho*I sites in the splicing reporter minigene pI-11 (ref. 15), to create splicing reporters pI-11-IL7R172 and pI-11-IL7R382.

We created mutants of the minigenes with either A or G at rs6897932 and also created a version of pI-11-IL7R382 in which 3 nt upstream and 3 nt downstream of this locus were mutated (TAA[T]CAT → GCC[T]ACG). All constructs and subclones were sequenced. The sequences of these constructs and subclones and the oligonucleotides used in their construction are available upon request.

Rat DT3 and AT3 cells were cultured in low-glucose DMEM (Gibco) with 10% FBS. HeLa cells were cultured in high-glucose DMEM with 10% FBS. Stable transfections (**Fig. 2b**) were as described in ref. 50. Transient transfections (**Fig. 2c**) used the same stable transfection protocol, except that 100 ng of DNA was transfected in cells, and cells were harvested 48 h post-transfection without any selection. All transfections were performed in triplicate for each construct.

Total RNA was extracted from stable cells using Trizol reagent (Invitrogen) according to the manufacturer's protocol, and exon 6 inclusion was determined by RT-PCR analysis as previously described[15,50]. PCR products were loaded directly on a 5% polyacrylamide nondenaturing gel, electrophoresed at 100 V for 3–4 h and exposed to Amersham Hyperfilm-MP or Molecular Dynamics phosphorimager screens. Quantification of PCR products was performed with ImageQuant software (Molecular Dynamics). In each case, RT-PCR products were normalized for molar equivalence.

Stable minigene transfection. Rat DT3 cells were cultured in low-glucose DMEM (Gibco) with 10% FBS. Stable transfections were as described in ref. 50. All transfections were performed in triplicate for each construct, as described above.

Allele-specific *IL7R* isoform expression in healthy controls. Real-time PCR was carried out in 94 control individuals using Applied Biosystems TaqMan one-step real-time PCR master mix reagents kit under the manufacturer's recommended conditions on an ABI 7900HT Sequence Detection System using SDS 2.0 software. Positive and negative controls, as well as a calibration curve spanning five orders of magnitude constructed with an *in vitro*–transcribed clone of *GAPDH*, all in triplicate, were also included in each reaction plate. Amplification values are expressed as $10^{12}/2^{ct}$. Three different gene expression TaqMan assays were tested: Hs00233682_m1, which covers the *IL7R* exon 4/5 boundary, Hs00904814_m1, which covers the exon 6/7 boundary, and Hs00902338_g1, which covers the exon 7/8 boundary.

Data access. Access to summary and primary data may be requested through the International Multiple Sclerosis Genetics Consortium website.

URLs. HapMap: http://www.hapmap.org; Ensembl: http://www.ensembl.org; European Collection of Cell Cultures (ECACC): http://www.ecacc.org.uk; British 1958 Birth Cohort: http://www.b58cgene.sgul.ac.uk; Primer3: http://fokker.wi.mit.edu/primer3/; International Multiple Sclerosis Genetics Consortium website: http://imsgc.org.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank all affected individuals and families who participated in this study and the collaborating clinics and physicians for referring individuals to the study. We also thank J. van der Walt, K. McDowell and W. Pope for their laboratory and technical assistance and C. DeLoa for clinical data management. We acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02. A.G. is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO - Vlaanderen). B.D. is supported by the Research Council of the University of Leuven. This research was supported by US National Institutes of Health grants NS32830 (J.L.H., M.A.P.-V.), NS26799 (S.L.H., J.R.O.), NS049477 (J.L.H., J.R.O., S.L.H., M.A.P.-V., S.J.S. and D.A.S.C.), GM63090 (M.A.G.-B.) and NMSS RG 2901C6 (J.R.O.) and a postdoctoral grant from the NMSS, FG 1718-A-1 (J.L.M.).

AUTHOR CONTRIBUTIONS

J.R.O., S.L.H., R.L., S.J.S., D.A.S.C. and B.D. collected and reviewed the participant samples and clinical information. J.L.H., M.A.P.-V. and S.G.G. designed the overall study. J.H. performed genotyping and sequencing using assays designed by and under the direction of S.G.G. M.B. and A.G. performed genotyping under the direction of S.J.S., and S.J.C. performed genotyping under the direction of J.R.O. S.J.C. performed expression assays under the direction of J.R.O. M.A.G.-B. and P.S. designed the differential splicing assays that were performed by P.S. M.A.G.-B., S.G.G. and P.S. interpreted the differential splicing assays, and M.A.G.-B. wrote the relevant portions of the manuscript. Statistical analysis and interpretation of the data was performed by S.S., A.P., L.F.B., J.L.M., J.L.H. and M.A.P.-V. Molecular analysis and interpretation was performed by S.G.G., J.R.O. and M.A.G.-B. The manuscript was written by S.G.G., S.S., J.L.H. and M.A.P.-V., with review and contributions by all authors. J.L.H. and M.A.P-V. contributed equally to this work.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Willer, C.J., Dyment, D.A., Risch, N.J., Sadovnick, A.D. & Ebers, G.C. Twin concordance and sibling recurrence rates in multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **100**, 12877–12882 (2003).
2. Fernald, G.H., Yeh, R.F., Hauser, S.L., Oksenberg, J.R. & Baranzini, S.E. Mapping gene activity in complex disorders: Integration of expression and genomic scans for multiple sclerosis. *J. Neuroimmunol.* **167**, 157–169 (2005).
3. Hauser, M.A. *et al.* Genomic convergence: identifying candidate genes for Parkinson's disease by combining serial analysis of gene expression and genetic linkage. *Hum. Mol. Genet.* **12**, 671–677 (2003).
4. Giedraitis, V. *et al.* Genome-wide TDT analysis in a localized population with a high prevalence of multiple sclerosis indicates the importance of a region on chromosome 14q. *Genes Immun.* **4**, 559–563 (2003).
5. Oturai, A. *et al.* Linkage and association analysis of susceptibility regions on chromosomes 5 and 6 in 106 Scandinavian sibling pair families with multiple sclerosis. *Ann. Neurol.* **46**, 612–616 (1999).
6. Martin, E.R., Monks, S.A., Warren, L.L. & Kaplan, N.L. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* **67**, 146–154 (2000).
7. Monks, S.A. & Kaplan, N.L. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. *Am. J. Hum. Genet.* **66**, 576–592 (2000).
8. Abecasis, G.R., Cardon, L.R. & Cookson, W.O. A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292 (2000).
9. Roxburgh, R.H. *et al.* Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity. *Neurology* **64**, 1144–1151 (2005).
10. Schmidt, S. *et al.* Allelic association of sequence variants in the herpes virus entry mediator-B gene (*PVRL2*) with the severity of multiple sclerosis. *Genes Immun.* **7**, 384–392 (2006).
11. Nyholt, D.R. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am. J. Hum. Genet.* **74**, 765–769 (2004).
12. Bruzzi, P., Green, S.B., Byar, D.P., Brinton, L.A. & Schairer, C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am. J. Epidemiol.* **122**, 904–914 (1985).

13. Sham, P.C. *et al.* Haplotype association analysis of discrete and continuous traits using mixture of regression models. *Behav. Genet.* **34**, 207–214 (2004).
14. Goodwin, R.G. *et al.* Cloning of the human and murine interleukin-7 receptors: demonstration of a soluble form and homology to a new receptor superfamily. *Cell* **60**, 941–951 (1990).
15. Carstens, R.P., McKeehan, W.L. & Garcia-Blanco, M.A. An intronic sequence element mediates both activation and repression of rat fibroblast growth factor receptor 2 pre-mRNA splicing. *Mol. Cell. Biol.* **18**, 2205–2217 (1998).
16. Cox, A.L. *et al.* Lymphocyte homeostasis following therapeutic lymphocyte depletion in multiple sclerosis. *Eur. J. Immunol.* **35**, 3332–3342 (2005).
17. Corder, E.H. *et al.* Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).
18. Haines, J.L. *et al.* Complement factor H variant increases the risk of age-related macular degeneration. *Science* **308**, 419–421 (2005).
19. Kenealy, S.J., Pericak-Vance, M.A. & Haines, J.L. The genetic epidemiology of multiple sclerosis. *J. Neuroimmunol.* **143**, 7–12 (2003).
20. Haines, J.L. *et al.* Linkage of the MHC to familial multiple sclerosis suggests genetic heterogeneity. The Multiple Sclerosis Genetics Group. *Hum. Mol. Genet.* **7**, 1229–1234 (1998).
21. Ramanathan, M. *et al.* In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. *J. Neuroimmunol.* **116**, 213–219 (2001).
22. Bomprezzi, R. *et al.* Gene expression profile in multiple sclerosis patients and healthy controls: identifying pathways relevant to disease. *Hum. Mol. Genet.* **12**, 2191–2199 (2003).
23. Ebers, G.C. *et al.* A full genome search in multiple sclerosis. *Nat. Genet.* **13**, 472–476 (1996).
24. Sundvall, M. *et al.* Identification of murine loci associated with susceptibility to chronic experimental autoimmune encephalomyelitis. *Nat. Genet.* **10**, 313–317 (1995).
25. Teutsch, S.M., Booth, D.R., Bennetts, B.H., Heard, R.N. & Stewart, G.J. Identification of 11 novel and common single nucleotide polymorphisms in the interleukin-7 receptor-alpha gene and their associations with multiple sclerosis. *Eur. J. Hum. Genet.* **11**, 509–515 (2003).
26. Zhang, Z. *et al.* Two genes encoding immune-regulatory molecules (LAG3 and IL7R) confer susceptibility to multiple sclerosis. *Genes Immun.* **6**, 145–152 (2005).
27. Booth, D.R. *et al.* Gene expression and genotyping studies implicate the interleukin 7 receptor in the pathogenesis of primary progressive multiple sclerosis. *J. Mol. Med.* **83**, 822–830 (2005).
28. Kondo, M., Weissman, I.L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
29. Akashi, K., Kondo, M., Freeden-Jeffry, U., Murray, R. & Weissman, I.L. Bcl-2 rescues T lymphopoiesis in interleukin-7 receptor-deficient mice. *Cell* **89**, 1033–1041 (1997).
30. Jameson, S.C. Maintaining the norm: T-cell homeostasis. *Nat. Rev. Immunol.* **2**, 547–556 (2002).
31. Fry, T.J. & Mackall, C.L. The many faces of IL-7: from lymphopoiesis to peripheral T cell maintenance. *J. Immunol.* **174**, 6571–6576 (2005).
32. Al Shami, A. *et al.* A role for thymic stromal lymphopoietin in CD4(+) T cell development. *J. Exp. Med.* **200**, 159–168 (2004).
33. Huston, D.P. & Liu, Y.J. Thymic stromal lymphopoietin: a potential therapeutic target for allergy and asthma. *Curr. Allergy Asthma Rep.* **6**, 372–376 (2006).
34. Koyama, K. *et al.* A possible role for TSLP in inflammatory arthritis. *Biochem. Biophys. Res. Commun.* **357**, 99–104 (2007).
35. Garcia-Blanco, M.A., Baraniak, A.P. & Lasda, E.L. Alternative splicing in disease and therapy. *Nat. Biotechnol.* **22**, 535–546 (2004).
36. Traggiai, E. *et al.* IL-7-enhanced T-cell response to myelin proteins in multiple sclerosis. *J. Neuroimmunol.* **121**, 111–119 (2001).
37. Correale, J., Fiol, M. & Gilmore, W. The risk of relapses in multiple sclerosis during systemic infections. *Neurology* **67**, 652–659 (2006).
38. Haines, J.L. *et al.* A complete genomic screen for multiple sclerosis underscores a role for the major histocompatability complex. The Multiple Sclerosis Genetics Group. *Nat. Genet.* **13**, 469–471 (1996).
39. Barcellos, L.F. *et al.* Linkage and association with the *NOS2A* locus on chromosome 17q11 in multiple sclerosis. *Ann. Neurol.* **55**, 793–800 (2004).
40. van Horssen, J. *et al.* Matrix metalloproteinase-19 is highly expressed in active multiple sclerosis lesions. *Neuropathol. Appl. Neurobiol.* **32**, 585–593 (2006).
41. Malmestrom, C., Andersson, B.A., Haghighi, S. & Lycke, J. IL-6 and CCL2 levels in CSF are associated with the clinical course of MS: implications for their possible immunopathogenic roles. *J. Neuroimmunol.* **175**, 176–182 (2006).
42. Xu, H. *et al.* SNPselector: a web tool for selecting SNPs for genetic association studies. *Bioinformatics* **21**, 4181–4186 (2005).
43. O'Connell, J.R. & Weeks, D.E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**, 259–266 (1998).
44. Abecasis, G.R., Cherny, S.S., Cookson, W.O. & Cardon, L.R. Merlin–rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* **30**, 97–101 (2002).
45. Zaykin, D., Zhivotovsky, L. & Weir, B.S. Exact tests for association between alleles at arbitrary numbers of loci. *Genetica* **96**, 169–178 (1995).
46. Abecasis, G.R. & Cookson, W.O. GOLD–graphical overview of linkage disequilibrium. *Bioinformatics* **16**, 182–183 (2000).
47. de Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.* **37**, 1217–1223 (2005).
48. Horvath, S. *et al.* Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet. Epidemiol.* **26**, 61–69 (2004).
49. Gauderman, W.J. Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* **155**, 478–484 (2002).
50. Baraniak, A.P., Lasda, E.L., Wagner, E.J. & Garcia-Blanco, M.A. A stem structure in fibroblast growth factor receptor 2 transcripts mediates cell-type-specific splicing by approximating intronic control elements. *Mol. Cell. Biol.* **23**, 9327–9337 (2003).