AJHG, Volume *86*

Supplemental Data

GWA in a High-Risk Isolate for MS Reveals Associated Variants in STAT3 Gene

Eveliina Jakkula, Virpi Leppä, Anna-Maija Sulonen, Teppo Varilo, Suvi Kallio, Anu Kemppinen, Shaun Purcell, Keijo Koivisto, Pentti Tienari, Marja Liisa Sumelahti, Irina Elovaara, Tuula Pirttilä, Mauri Reunanen, Arpo Aromaa, Annette Bang Oturai, Helle Bach Søndergaard, Hanne F. Harbo, Inger-Lise Mero, Stacey B. Gabriel, Daniel B. Mirel, Stephen L. Hauser, Ludwig Kappos, Chris Polman, Philip L. De Jager, David A. Hafler, Mark J. Daly, Aarno Palotie, Janna Saarela, and Leena Peltonen

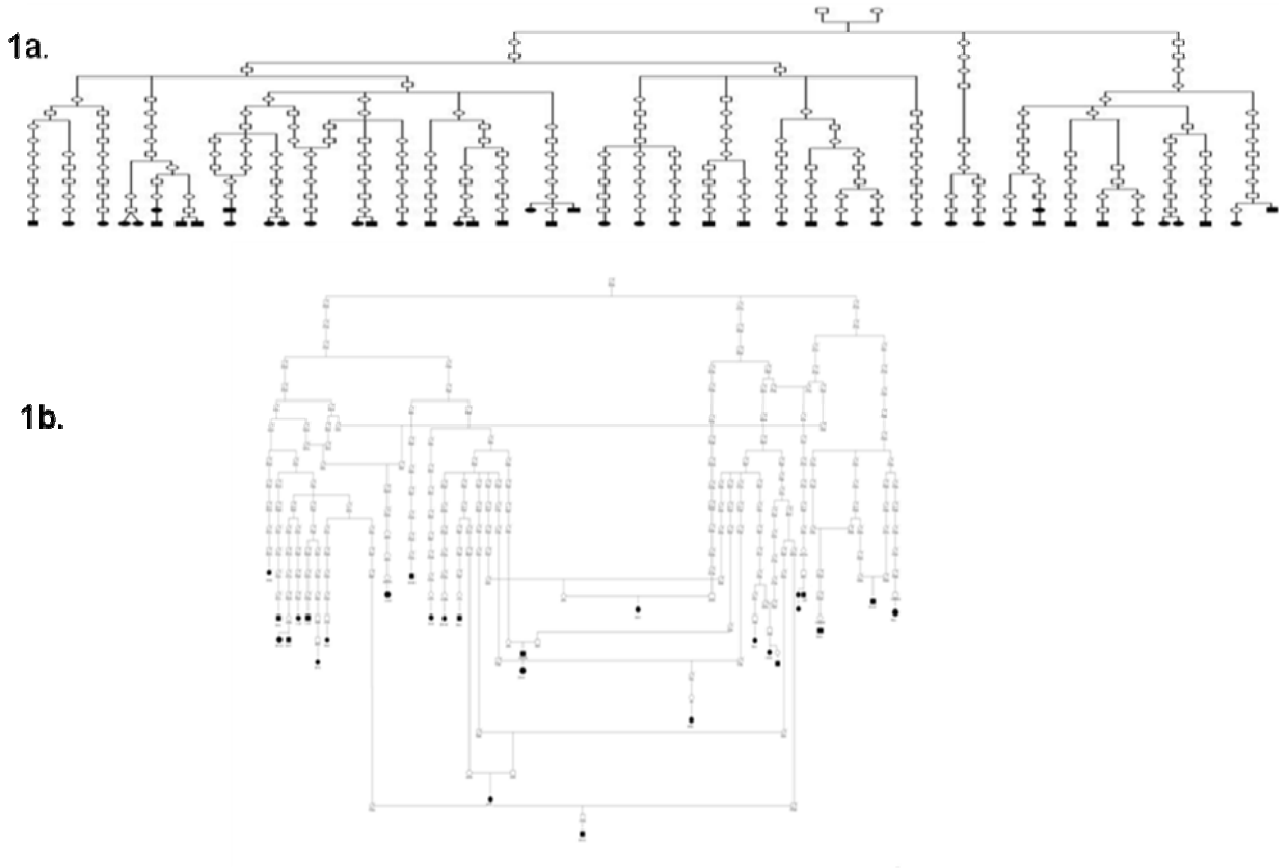

Figure S1. The Extended Pedigrees for the Isolate Study Set
Figure 1a and b present the extended pedigrees of the Southern Ostrobotnia originating cases reaching up to 18 generations back. We constructed the pedigrees using genealogical information from population registries in accordance with published criteria. Only the shortest connections of relatedness within extended families are shown, although there is extensive interrelatedness in the pedigrees. These pedigrees are supportive of founder effect.

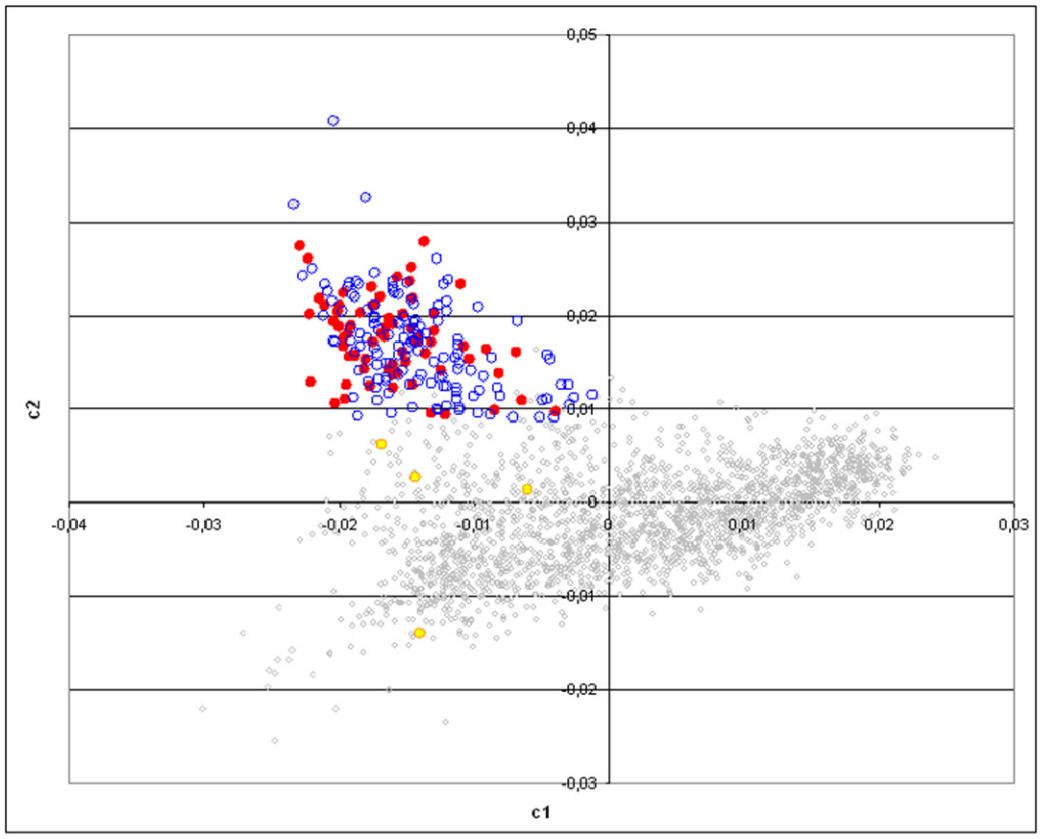

Figure S2. The MDS Plot of the GWA Sample Set
The two first dimensions (c1, c2) of the multidimensional scaling analysis of IBS sharing distances between individuals are plotted. MS cases included in the study (n = 68) are indicated with red dots, MS cases excluded from the study (n = 4) based on the results of MDS (not clustering with Southern Ostrobothnia originating cases) are indicated with yellow dots, selected controls (n = 136) with blue open circles and population controls by grey circles. The first and the second dimensions of MDS-analysis as well as principal component analysis have recently been shown to correlate with east-west (c1) and South - North geographical axes within Finland[23].

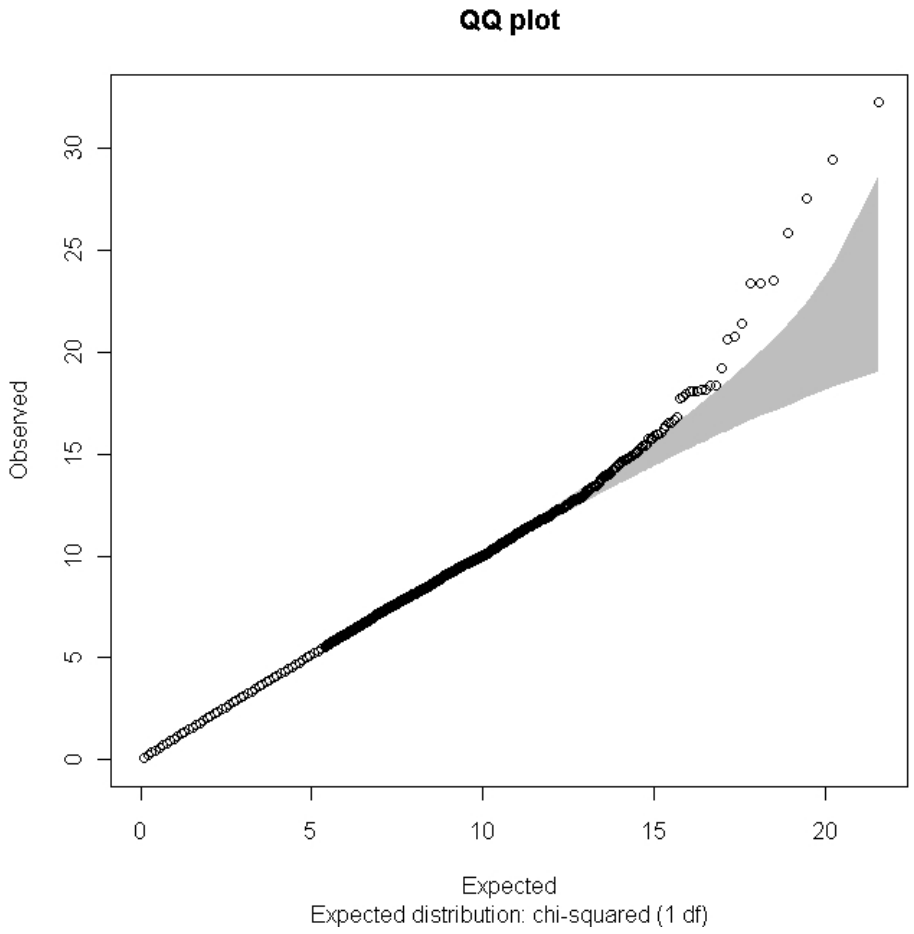

Figure S3. Quantile-Quantile (QQ) Plot of χ^2 Expected vs. Observed Statistics of Single SNP Allele Association Analysis

QQ-plot follows null-hypothesis and only end-of-tail distribution differs showing possible real association signals. No major inflation of χ^2-statistics is present and thus cases and controls are well-matched.

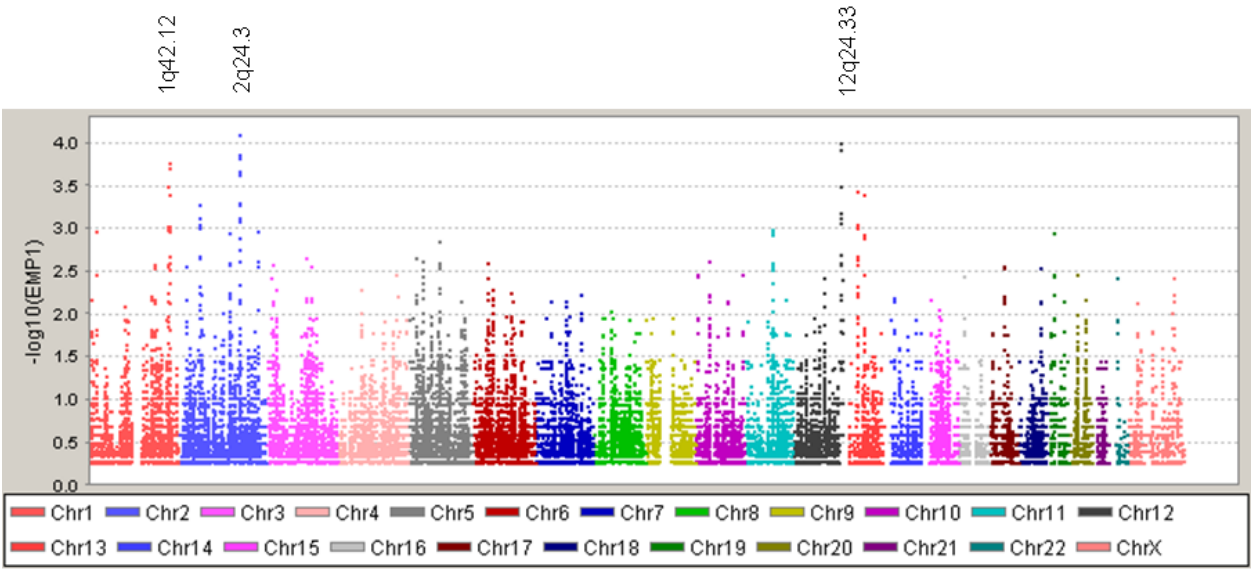

Figure S4. The Regions of Homozygosity (ROHs) Enriched in Cases
Regions in chromosomes 1, 2 and 12 illustrate more homozygous segments in cases compared to controls. Empirical p-values (EMP1) obtained by comparing ROHs in cases and controls and evaluation of statistical significance was obtained by permutation. X-axis depicts each chromosome in distinguishable colour and Y-axis represents $-\log_{10}(P)$ of empirically obtained P-values.

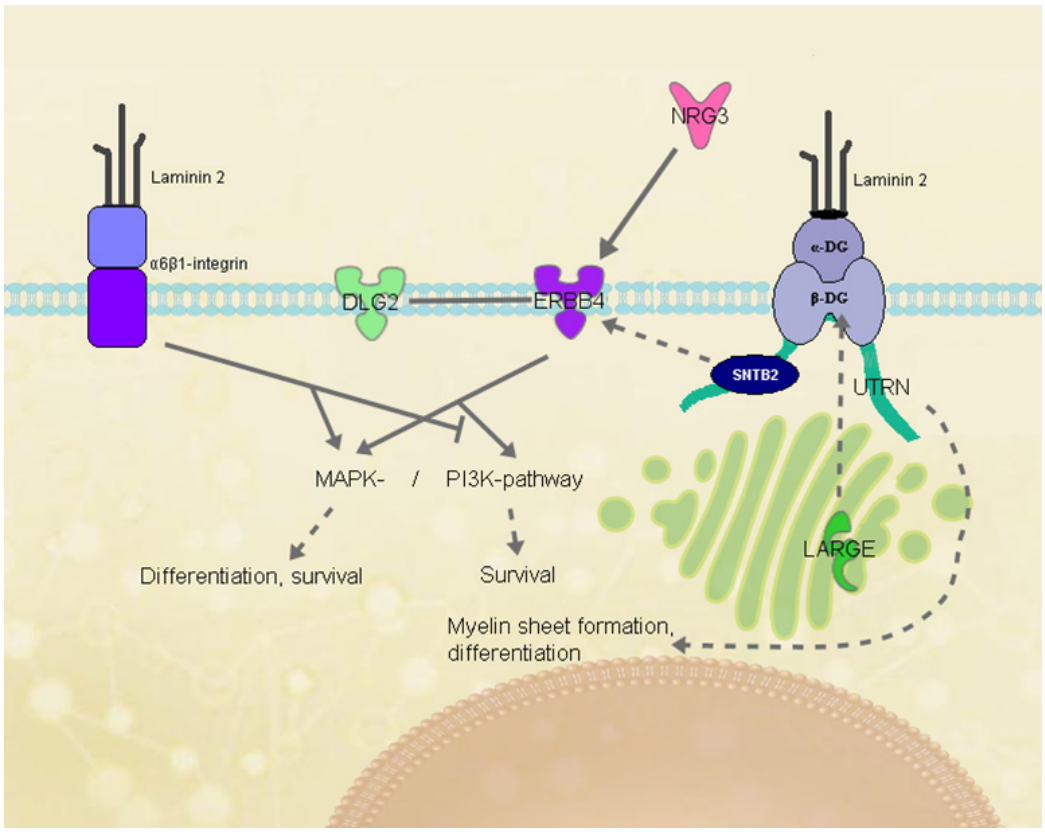

Figure S5. The Pathway Involving the Genes *NRG3*, *ERBB4*, *DLG2*, *UTRN*, and *LARGE*
The pathway involving the genes *NRG3*, *ERBB4*, *DLG2*, *UTRN* and *LARGE* detected with Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) potentially regulates oligodendrocyte differentiation and myelin sheet formation. Neuregulin 3 (coded by NRG3-gene) is a neural growth factor which inhibits apoptosis of oligodendrocyte precursor cells trough its receptor ERBB4 and PI3-K-pathway[40]. DLG2 (PSD-93, chapsyn-110) is likely to be involved in the assembly of ERBB4 to the cell membranes, in ERBB4 turnover and neuregulin signaling[41,38]. The differentiation of oligodendrocyte precursors and the proper formation of myelin sheets by mature oligodendrocytes seems to be regulated by interactions between neuregulin-ERBB –mediated signaling, laminin-2-integrin signaling and laminin-2-dystroglycan (DG) signaling[36,37]. Though the exact intracellular pathways in laminin-2-dystroglycan signaling mediated myelin sheet formation are unknown, dystroglycan is known to bind utrophin (UTRN). Utrophin in turn binds β2-syntrophin (SNTB2), which has a binding site for ERBB4[38], thus potentially enabling direct interactions between ERBB4 and dystroglycan-utrophin-complex. Like-glycosyltransferase (LARGE) is an important glycosyl transferase for dystroglycan. Mutations in LARGE-gene cause congenital muscular dystrophy type 1D (MDC1D |MIM #608840]), a muscular dystrophy with profound mental retardation, white matter changes and subtle structural abnormalities on brain MRI[39]. Figure modified from Ingenuity Pathway Analysis.

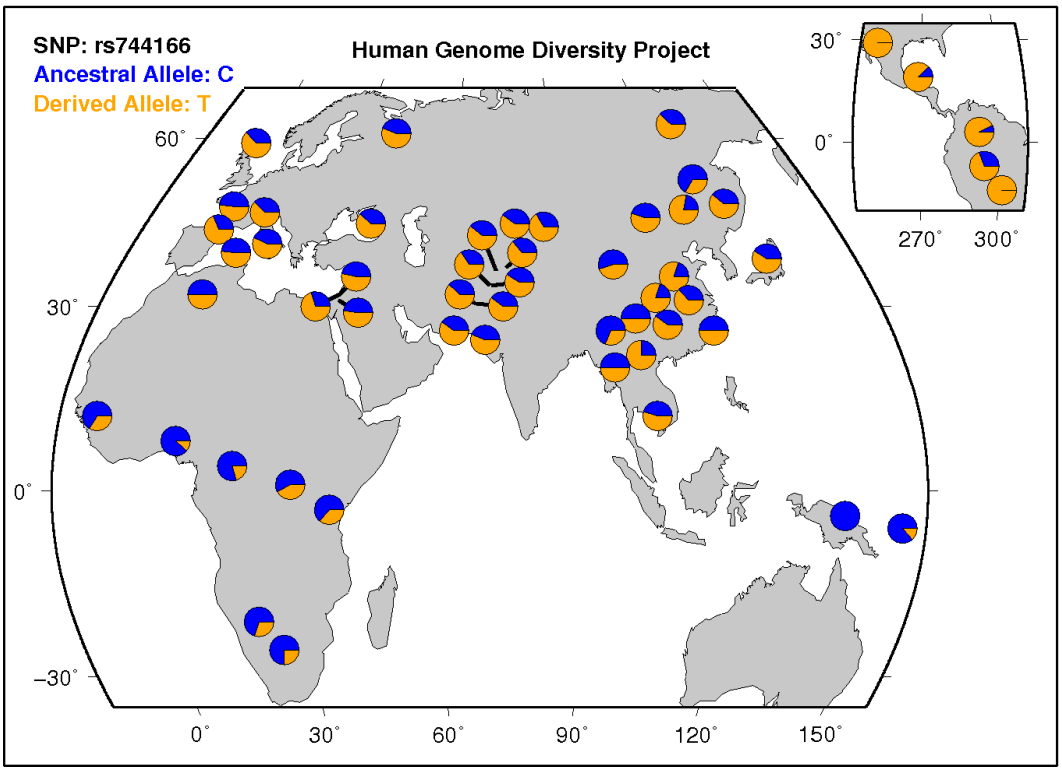

Figure S6. The Worldwide Distribution of rs744166 Alleles
The distribution of the rs744166 alleles as courtesy of the Human Genome Diversity Panel team[45]. The derived A-allele (here T-allele) tags the MS-protective haplotype in populations outside Africa and therefore the picture illustrates the proportion of the protective haplotype in different populations. The protective allele has become more common in populations outside Africa, but seems to be absent in the Oceanian populations near the Equator.

Table S1. Overlapping Extended Regions of Homozygosity (ROHs) Enriched in Cases (Empirical P <10^{-3} Evaluated by Permutation)

Overlapping Consensus Region Identified in Cases						ROH of 50 SNPs and 500 kb Extending within Each Region			Identified Overlapping Region as Homozygous	
Chr	**Start (Mb)**	**End (Mb)**	**Length (kb)**	**N of SNPs**	**Gene(s)**	**N of Cases (%)**	**N of Controls (%)**	**EMP1**	**N of Cases (%)**	**N of Controls (%)**
1	220,967	221,209	242	24	*CNIH3*	11 (16%)	3 (2%)	3 x 10^{-4}	11 (16%)	11 (8%)
2	167,634	168,146	512	39	*CMYA3*	30 (44%)	26 (19%)	8 x 10^{-5}	27 (40%)	27 (20%)
12	130,783	131,355	573	48	multiple	9 (13%)	1 (0.4%)	3 x 10^{-4}	7 (10%)	2 (1%)

ROHs with a minimum length of 50 SNPs and 500 kb were identified first in each individual. Then, ROHs in cases were compared to controls to expose regions of homozygosity enriched in cases. Overlapping consensus region of ROHs in cases were then identified within these enriched regions and boundaries for consensus region were determined. The frequencies for ROHs within each region (as defined by boundaries identified in cases) are presented in the table for both cases and controls both as defining ROH as 50 SNP and 500 kb and only for the overlapping consensus region (fewer SNPs and shorter length). EMP1 corresponds for empirical P-value obtained by permutation.

Table S2. Identical-by-Descent (IBD)-Sharing Estimates between Cases and Controls

	N of Case-Case Pairs (%)	N of Case-Control Pairs (%)	N of Control-Control Pairs (%)
pi-hat = 0	1543 (67.73 %)	6204 (67.08 %)	5960 (64.92)
0 < pi-hat < 0.01	435 (19.10 %)	1829 (19.78 %)	1935 (21.08 %)
0.01 > pi-hat > 0.03	286 (12.55 %)	1168 (12.63 %)	1214 (13.22 %)
0.03 > pi-hat > 0.06	7 (0.31 %)	45 (0.49 %)	66 (0.72 %)
0.06 > pi-hat > 0.12	1 (0.04 %)	2 (0.02 %)	2 (0.02 %)
0.12 > pi-hat > 0.25	4 (0.18 %)	0 (0 %)	2 (0.02 %)
pi-hat > 0.25	2 (0.09 %)	0 (0 %)	1 (0.01 %)
Total n of pairs	2278	9248	9180

IBD-sharing estimates were calculated between case-case, case-control and control-control pairs using genome-wide SNP data. Pi-hat corresponds to the proportion of the genome that is estimated to be shared by IBD between two individuals. Theoretically, 1st degree relatives (parent-child and sibling-sibling pairs) share 50 % of their genome by IBD, 2nd degree relatives share 25 % of their genome by IBD, 3rd degree relatives (eg 1st cousins) 12.5 %, 4th degree relatives 6.3 %, 5th degree relatives 3.1 %, 6th degree relatives 1.6 % and 7th degree relatives 0.78 %. Thus, as suggested by genealogical research, results of these IBD-estimations show that most relationships between cases and between cases and controls are distant and no 1st degree relatives were observed in any of the groups. The closest known relatives included in case group were 2nd degree relatives and their genetic relationship based on IBD-estimates correlated well with their known relationships. No unknown close relationships were observed in case groups, no genealogical information was available for controls.

Table S3. The Description of All CNVs Found in the Isolate GWA Samples

Chr	Start (b35)	End (b35)	Size (bp)	N of SNPs	Copy Number Type[a]	Cases (n=68)	Cases Freq.	Db of Genomic Variants, August 5th 2009 Freeze	Genes
1	2058523	2312823	254301	21	gain	3	4.4%	Known variation	PRKCZ, C1orf86, SKI, MORN1
1	12354566	12763485	408920	36	gain	1	1.5%	Known variation	VPS13D, DHRS3, AADACL4, AADACL3, C1orf158, after TNFRSF1B
1	102379409	102651711	272303	21	loss	1	1.5%	Known variation	
1	120776649	120923841	147193	7	gain	2	2.9%	Known variation	
1	193569717	193628745	59029	6	loss	4	5.9%	Known variation	CFHR4
2	19443	189678	170236	14	gain	1	1.5%	Known variation	FAM110C
2	35725730	35999293	273564	40	loss	1	1.5%	Known variation	
2	49511191	49664240	153050	21	loss	2	2.9%	Known variation	
2	53071353	53103920	32568	4	loss	1	1.5%	Known variation	
2	54354348	54404322	49975	7	loss	1	1.5%	NEW	ACYP2 intron, TSPYL6
2	86983132	87615568	632437	3	gain	1	1.5%	Known variation	CD8B1, PLGLB1, RGPD1
2	89772948	89932893	159946	3	loss	7	10.3%	Known variation	
2	99215459	99306408	90950	4	gain	3	4.4%	Known variation	TSGA10, C2orf15, LIPT1, MITD1, MRPL30, LYG2, LYG1
2	110214618	110292098	77481	7	loss	1	1.5%	Known variation	MALL, NPHP1
2	212981040	213018166	37127	5	loss	3	4.4%	Known variation	ERBB4 intron
2	241106342	241131279	24938	6	gain	1	1.5%	Known variation	GPC1
3	41894	108412	66519	8	gain	1	1.5%	Known variation	CHL1 promoter
3	4021691	4329697	308007	35	loss	1	1.5%	Known variation	SETMAR
3	5383302	5411345	28044	6	loss	1	1.5%	Known variation	
3	24134391	24172277	37887	8	loss	1	1.5%	NEW	THRB, last 6 exons
3	65166887	65187636	20750	3	loss	1	1.5%	Known variation	
3	152993783	153028739	34957	7	loss	3	4.4%	Known variation	AADAC
4	84915835	84969552	53718	10	loss	1	1.5%	Known variation	
4	132303980	132685819	381840	24	gain	1	1.5%	Known variation	

4	167425804	167482371	56568	8	gain	1	1.5%	Known variation	After TLL1
4	190200031	191131631	931601	93	gain	1	1.5%	Known variation	
5	19055301	19305295	249995	22	loss	1	1.5%	Known variation	
5	97074222	97121798	47577	7	loss	5	7.4%	Known variation	RIOK2 near
5	104465860	104510644	44785	3	loss	1	1.5%	Known variation	
5	178661436	178854467	193032	25	gain	1	1.5%	Known variation	ADAMTS2
6	5496231	5522147	25917	7	loss	1	1.5%	Known variation	FARS2, intron
6	29076909	29287216	210308	25	loss	1	1.5%	Known variation	ZNF311, OR2W1, OR2B3, OR2J3, OR2J2
6	67044129	67104015	59887	11	loss	5	7.4%	Known variation	BAI3 promoter
6	95472452	95649511	177060	9	loss	1	1.5%	Known variation	
6	137942200	138003365	61166	21	gain	1	1.5%	Known variation	OLIG3 promoter, TNFAIP3 promoter
6	144743809	145071977	328169	32	loss	1	1.5%	Known variation	UTRN, many exons
7	3175715	3252673	76959	8	loss	1	1.5%	Known variation	SDK1, intron
7	6643875	7123923	480049	54	gain	1	1.5%	Known variation	C1GALT1
7	8882102	8996708	114607	23	loss	3	4.4%	Known variation	After NXPH1
7	9790639	9877397	86759	9	loss	2	2.9%	Known variation	
7	12560343	12685347	125005	17	loss	1	1.5%	Known variation	
7	69761016	70029858	268843	22	gain	1	1.5%	Known variation	After AUTS2, WBSCR17 promoter
7	70919613	71427575	507963	52	gain	1	1.5%	Known variation	CALN1, exons 1-3
7	75690975	76059191	368217	6	gain	1	1.5%	Known variation	ZP3, DTX2, UPK3B, POMZP3
7	119899100	119906697	7598	3	loss	1	1.5%	Known variation	KCND2, intron
7	152959823	153145033	185211	20	gain	1	1.5%	Known variation	DPP6
8	5590045	5591685	1641	3	loss	1	1.5%	Known variation	
8	16170358	16306880	136523	11	loss	1	1.5%	Known variation	MSR1 promoter
8	87256166	87403084	146919	11	gain	1	1.5%	Known variation	SLC7A13, WWP1 promoter
8	92192384	92243291	50908	4	loss	5	7.4%	Known variation	Hypothetical protein FLJ27355, after OTUD6B, SLC26A7 promoter
8	137757412	137919630	162219	13	loss	1	1.5%	Known variation	
9	5296824	5325470	28647	6	loss	1	1.5%	Known variation	RLN1
9	10651370	10676202	24833	7	loss	1	1.5%	Known variation	PTPRD promoter
9	21743138	21761241	18104	4	gain	1	1.5%	NEW	MTAP promoter

9	28534375	28556849	22475	5	loss	3	4.4%	Known variation	LINGO2
9	69329605	69348516	18912	4	loss	1	1.5%	Known variation	APBA1, intron
9	70946694	71169744	223051	32	gain	1	1.5%	Known variation	TRPM3 promoter & exon1
9	71310230	71589788	279559	48	gain	1	1.5%	Known variation	TMEM2
10	15030375	15100889	70515	6	loss	1	1.5%	Known variation	DCLRE1C, MEIG1
10	47013328	47173619	160292	12	gain	12	17.6%	Known variation	ANXA8 promoter
10	51457810	51805221	347412	28	gain	1	1.5%	Known variation	FAM21A, ASAH2, SGMS1
10	66102105	67710331	1608227	209	loss	1	1.5%	Known variation	CTNNA3
10	81567594	82006206	438613	40	gain	1	1.5%	Known variation	SFTPD, C10orf57, PLAC9, ANXA11
10	82869699	82875955	6257	3	loss	2	2.9%	Known variation	NRG3 promoter
11	38188336	38965147	776812	58	loss	1	1.5%	Known variation	
11	84219051	84245672	26622	4	loss	1	1.5%	Known variation	DLG2, intron
11	89843914	91455249	1611336	121	loss	1	1.5%	Known variation	
11	133873030	134225383	352354	69	gain	1	1.5%	Known variation	
12	7884583	8017012	132430	14	gain	5	7.4%	Known variation	SLC2A14, SLC2A3
12	31157554	31293957	136404	12	gain	1	1.5%	Known variation	OVOS2
12	31898373	31954269	55897	12	gain	2	2.9%	Known variation	
12	42212399	42288535	76137	12	loss	2	2.9%	Known variation	ADAMTS20
12	81671084	81707620	36537	5	loss	1	1.5%	Known variation	TMTC2, intron
12	98788860	98966181	177322	18	gain	1	1.5%	Known variation	ANKS1B exon1, KIAA0701 last 7 exons
12	130255197	130339814	84618	11	loss	1	1.5%	Known variation	
13	67542909	67552774	9866	3	gain	1	1.5%	NEW	
13	83009774	83055928	46155	8	loss	8	11.8%	Known variation	After SLITRK1
13	94733590	94779857	46268	11	gain	1	1.5%	Known variation	ABCC4 promoter & exon1
13	94797740	94825508	27769	7	gain	1	1.5%	Known variation	
14	73494691	73538366	43676	4	loss	1	1.5%	NEW	COQ6 last 8 exons, ENTPD5 exons 1-13
15	20347960	20777695	429736	50	gain	1	1.5%	Known variation	TUBGCP5, CYFIP1, NIPA2, NIPA1
15	38624662	39476524	851863	53	loss	2	2.9%	Known variation	CCDC32, RPUSD2, CASC5, RAD51, FAM82C, GCHFR, DNAJC17, ZFYVE19, PPP1R14D, SPINT1, RHOV, VPS18, DLL4, CHAC1, INOC1, EXDL1, CHP, OIP5,

									NUSAP1, NDUFAF1
15	98736980	98770528	33549	5	loss	1	1.5%	Known variation	LASS3
16	21515973	21647775	131803	12	loss	1	1.5%	Known variation	METTL9, IGSF6, OTOA
16	81289305	81385638	96334	27	loss	1	1.5%	Known variation	CDH13, intron
17	30708148	30792312	84165	14	loss	1	1.5%	Known variation	SLFN11 promoter & exons 1-4, SLFN12, SLFN13 exon1
17	38782929	38958044	175116	13	gain	1	1.5%	Known variation	ARL4D, DHX8, near BRCA1
17	74878104	74905197	27094	9	gain	1	1.5%	Known variation	HRNBP3
18	1917798	1970668	52871	9	loss	1	1.5%	Known variation	
18	56949141	56950768	1628	3	loss	1	1.5%	NEW	
18	62947038	63159299	212262	31	gain	1	1.5%	Known variation	After DSEL
18	64897188	64909977	12790	3	loss	5	7.4%	Known variation	After hypothetical proteins CCDC102B & TXNDC10. DOK6 promoter
19	3755701	3786177	30477	9	gain	1	1.5%	Known variation	KIAA1086, ATCAY promoter, MATK promoter
19	7658063	7691713	33651	6	gain	1	1.5%	Known variation	FCER2 whole gene & promoter
19	48066441	48350666	284226	6	loss	2	2.9%	Known variation	PSG1, PSG6, PSG7, PSG11, PSG2, PSG5, PSG4, PSG9
19	59994795	60069820	75026	5	loss	1	1.5%	Known variation	KIR3DP1, KIR2DL4, KIR3DL1, KIR2DS4, KIR3DL2, FCAR & NCR1 promoter
20	5219710	5345328	125619	22	loss	1	1.5%	Known variation	PROKR2 whole gene and promoter
20	14874333	15089959	215627	38	loss	1	1.5%	Known variation	MACROD2
21	18981497	19000295	18799	6	gain	1	1.5%	Known variation	PRSS7 promoter
21	43646295	43663581	17287	7	gain	1	1.5%	Known variation	SNF1LK
22	20654301	20920813	266513	48	gain	1	1.5%	Known variation	TOP3B, IGLC1
22	23974960	24235221	260262	10	gain	2	2.9%	Known variation	LRP5L
22	32490635	32503091	12457	6	gain	1	1.5%	Known variation	LARGE, intron
X	6317902	7961924	1644023	92	gain	1	1.5%	Known variation	HDHD1A, STS, VCX, PNPLA4, VCX2
X	91173351	91262603	89253	5	loss	1	1.5%	Known variation	PCDH11X intron & exon3
X	140076396	140280802	204407	30	loss	2	2.9%	Known variation	SPANXA2 intron

Table S4. SNPs with P < 10^{-4} in the Genome-wide Association Analysis Were Replicated in Independent Finnish Sample Sets

			Isolate GWA n(MS)=68, n(ctrl)=136					Isolate (Independent) n(MS)=83, n(ctrl)=365				Finland n(MS)=628, n(ctrl)=668				Combined CMH n(MS)=711, n(ctrl)=1033	
Chr	**bp**	**SNP**		**MAF MS**	**MAF ctrl**	**p**	**OR**		**MAF MS**	**MAF ctrl**	**p**	**OR**	**MAF MS**	**MAF ctrl**	**p**	**OR**	**p**	**OR**
1	246044043	rs10736372	G	0.162	0.040	2.27E-05	4.58	G	0.085	0.088	0.9039	0.96	0.064	0.066	0.8707	0.97	0.8421	0.97
3	180874014	rs4854945	C	0.250	0.460	5.16E-05	0.39	C	0.325	0.366	0.3265	0.84	0.332	0.316	0.3802	1.08	0.6759	1.03
4	138364575	rs7696692	A	0.500	0.302	8.83E-05	2.32	A	0.361	0.375	0.7382	0.94	0.342	0.341	0.9236	1.01	0.9553	1.00
4	182400130	rs13121504	G	0.145	0.033	4.07E-05	4.96	G	0.048	0.045	0.8792	1.07	0.048	0.062	0.1003	0.75	0.1402	0.79
4	186744181	rs1046236	A	0.441	0.246	6.15E-05	2.42	A	0.307	0.326	0.6400	0.92	0.260	0.265	0.7369	0.97	0.6239	0.96
6	32509195	rs3135338	A	0.544	0.298	1.35E-06	2.81	A	0.333	0.138	5.4E-10	3.12	0.291	0.128	1.8E-25	2.80	1.6E-25	3.43
6	42169711	rs7749023	C	0.507	0.309	9.43E-05	2.31	C	0.368	0.378	0.7989	0.96	0.385	0.379	0.7223	1.03	0.8357	1.02
7	129421525	rs12533403	A	0.331	0.537	8.54E-05	0.43	A	0.488	0.473	0.7340	1.06	0.486	0.447	0.0482	1.17	0.0474	1.15
8	16712180	rs2134111	A	0.368	0.574	8.82E-05	2.31	A	0.415	0.448	0.4397	0.87	0.429	0.420	0.6268	1.04	0.9193	1.01
8	15188615	rs12675001	A	0.243	0.445	7.13E-05	0.40	A	0.384	0.355	0.5116	1.13	0.326	0.340	0.4170	0.94	0.630	0.96
11	69515796	rs1994776	G	0.470	0.261	2.51E-05	2.51	G	0.277	0.318	0.3062	0.82	0.362	0.331	0.1013	1.15	0.2804	1.09
11	70119917	rs720629	G	0.053	0.228	1.18E-05	0.19	G	0.171	0.178	0.8215	0.95	0.154	0.165	0.4662	0.92	0.4511	0.93
11	74672874	rs480174	A	0.382	0.165	1.22E-06	3.12	A	n.a.	n.a.	n.a.	n.a.	0.287	0.275	0.5047	1.06	0.4765	1.06
11	74694735	rs611908	A	0.419	0.202	3.69E-06	2.85	A	0.319	0.280	0.3058	1.21	0.327	0.306	0.2359	1.11	0.1426	1.12
11	74801780	rs571632	A	0.331	0.540	6.4E-05	0.42	A	0.434	0.496	0.1847	0.78	n.a.	n.a.	n.a.	N.a.	0.08217	0.87
12	10785930	rs17809949	G	0.471	0.276	9.05E-05	2.34	G	0.354	0.315	0.3441	1.19	0.315	0.326	0.5729	0.95	0.8772	0.99
12	51096513	rs1610791	A	0.493	0.235	1.53E-07	3.16	A	0.343	0.313	0.4514	1.15	0.283	0.269	0.4018	1.08	0.291	1.09
12	68684239	rs1152944	G	0.184	0.051	1.82E-05	4.15	G	0.115	0.103	0.6729	1.12	0.113	0.117	0.7470	0.96	0.9572	0.99
12	111574024	rs232921	A	0.617	0.390	2.08E-05	2.53	A	0.494	0.423	0.0974	1.33	0.363	0.356	0.6987	1.03	0.3091	1.08
13	61230371	rs6562206	A	0.397	0.199	1.83E-05	2.66	A	0.271	0.282	0.7736	0.95	n.a.	n.a.	n.a.	n.a.	0.01884	1.21
15	27170361	rs11070500	A	0.118	0.022	5.58E-05	5.91	A	0.072	0.045	0.1541	1.64	0.058	0.073	0.1369	0.79	0.4215	0.89
16	58696642	rs1364194	A	0.169	0.044	2.14E-05	4.41	A	0.121	0.062	0.0086	2.08	0.076	0.058	0.0553	1.35	0.00465	1.48
16	77850583	rs270443	G	0.544	0.335	4.79E-05	2.37	G	0.488	0.398	0.0342	1.44	0.362	0.362	0.9919	1.00	0.3675	1.07
17	37767727	rs744166	G	0.559	0.353	7.19E-05	2.32	G	0.470	0.406	0.1288	1.30	0.446	0.391	0.0044	1.26	0.00120	1.27
17	13978033	rs8079640	A	0.081	0.007	6.71E-05	11.9	A	0.048	0.032	0.3251	1.54	0.028	0.021	0.2840	1.31	0.1613	1.36

17	37932924	rs647397	A	0.294	0.129	4.76E-05	2.82	A	0.183	0.176	0.8296	1.05	0.205	0.178	0.0776	1.19	0.0816	1.17
20	34443365	rs6130176	A	0.096	0.265	7.33E-05	0.29	A	0.217	0.204	0.7140	1.08	0.196	0.220	0.1295	0.86	0.2313	0.90
21	27006619	rs2952414	A	0.324	0.143	2.03E-05	2.86	A	0.206	0.195	0.7406	1.07	0.256	0.239	0.3264	1.10	0.302	1.09

We identified 27 loci with $P < 10^{-4}$ in the genome-wide association analysis of 68 MS cases from the Southern Ostrobothnia and 136 IBS matched controls and at least one SNP per loci was genotyped in the replication phase. This table shows the original GWAS association result and the replication results of these SNPs in an independent isolate sample (83 MS cases, 365 controls) and a more heterogenous general Finnish sample set (628 MS cases and 664 controls). The replication results were combined using Cochran-Mantel-Haenszel population stratified analysis.

Table S5. Summary of Results for rs744166 (chr17) in STAT3 and rs1364194 (chr16) by Population

	Chr17, rs744166 (A)					Chr16, rs1364194 (A)				
Study Population	**Freq MS**	**Freq Controls**	**P**	**OR**	**95% CI**	**Freq MS**	**Freq Controls**	**P**	**OR**	**95% CI**
Finland SO isolate GWA	0.441	0.647	0.0000719	0.43	(0.28-0.66)	0.169	0.044	0.0000214	4.41	(2.12-9.17)
Finland SO replication	0.530	0.594	0.129	0.77	(0.55-1.08)	0.121	0.062	0.00861	2.08	(1.19-3.63)
Finland	0.554	0.609	0.000444	0.80	(0.68-0.93)	0.077	0.058	0.0553	1.35	(0.99-1.84)
Norway	0.551	0.589	0.0439	0.86	(0.74-0.99)	0.077	0.080	0.720	0.72	(0.72-1.25)
Denmark	0.511	0.572	0.000556	0.78	(0.68-0.90)	0.078	0.087	0.383	0.89	(0.69-1.15)
GeneMSA Switzerland	0.591	0.601	0.757	0.96	(0.74-1.25)	0.097	0.088	0.682	1.10	(0.70-1.72)
GeneMSA Netherland	0.513	0.588	0.0214	0.74	(0.57-0.96)	0.104	0.084	0.291	1.27	(0.81-1.98)
GeneMSA US	0.537	0.608	0.00222	0.75	(0.62-0.90)	0.080	0.072	0.503	1.13	(0.80-1.59)
IMSGC UK[a]	0.541	0.56	0.267	0.93	(0.80-1.07)	0.043	0.039	0.576	1.10	(0.78-1.56)
IMSGC US[a]	0.575	0.582	0.7124	0.97	(0.82-1.15)	0.028	0.040	0.138	0.69	(0.43-1.13)
BWH[a]	0.557	0.597	0.00623	0.85	(0.76-0.95)	0.036	0.033	0.513	1.11	(0.81-1.52)
Combined replication[b]	**0.546**	**0.581**	**2.753×10^{-10}**	**0.87**	**(0.83-0.91)**	**0.081**	**0.058**	**0.0793**	**1.10**	**(0.99-1.23)**

The SNP rs744166 in STAT3 was associated with MS susceptibility in multiple independent populations, but we gained no additional support for rs1364194 in non-Finnish populations.

[a]Genotypes imputed for rs1364194

[b]Combined analysis of the replication sets (original analysis set Finland SO isolate GWA not included) was done using the Cochran-Mantel-Haenszel analysis implemented in PLINK[33]. Each population was treated as a separate cluster. The US samples were analyzed as three separate clusters, since they were all genotyped with different platforms (Table 1) and the Gene MSA sets in different centers. The combined MAFs are weighed averages of the population specific minor allele frequencies.

Table S6. Association for rs744166 G Allele in chr 17: 37767727 Results Using Different Models

Model	OR	SE	L95	U95	STAT	P
Additive	1.157	0.02561	1.1	1.216	5.683	1.33E-08
Dominant	1.184	0.03836	1.098	1.276	4.398	1.09E-05
Recessive	1.243	0.04475	1.138	1.357	4.858	1.19E-06

This table shows the results of a logistic regression analysis in PLINK using population as a covariate. OR stands for odds ratio, SE for standard errors, L95 and U95 are the 95 % confidence intervals, STAT is the coefficient t-statistic and P is the asymptotic p-value for t-statistic.

Table S7. Haplotype Diversity in the STAT3 Locus in Hapmap 23a Populations

Frequencies				**SNPs**																							
JPT	CHB	CEU	YRI	rs8075442	rs12721576	rs12721583	rs2306580	rs7215104	rs10153241	rs9890802	rs9889396	rs8069645	rs7217655	rs3816769	**rs6503695**	rs6503696	rs6503697	rs4103200	rs9891119	rs9912773	rs17593222	**rs744166**[a]	rs3785898	**rs957970**	rs1905340	rs129449918	rs4796644	rs1026916	rs7211777
0	0	0	0.08	T	C	A	C	C	T	C	G	A	C	T	**T**	C	A	G	A	C	C	**T**	G	**T**	C	T	G	G	G
0	0	0	0.06	T	C	T	C	C	T	C	G	A	C	T	**T**	C	A	G	A	C	C	**T**	G	**T**	C	T	G	G	G
0.57	***0.66***	***0.56***	***0.07***	***C***	***C***	***T***	***C***	***C***	***T***	***C***	***G***	***A***	***C***	***T***	***T***	***C***	***A***	***G***	***A***	***C***	***C***	***T***	***G***	***T***	***C***	***T***	***G***	***G***	***A***
0	0	0	0.01	T	C	T	C	C	T	C	G	A	C	T	**T**	C	A	G	A	C	C	**T**	G	**T**	C	T	G	G	A
0.40	*0.33*	*0.23*	*0.31*	*C*	*C*	*T*	*C*	*C*	*T*	*C*	*G*	*G*	*T*	*C*	*C*	*T*	*T*	*C*	*C*	*G*	*C*	*C*	*T*	*C*	*A*	*C*	*G*	*A*	*G*
0	0	0.02	0.24	C	C	T	C	C	T	C	G	G	T	C	**C**	T	T	C	C	C	C	**C**	T	**C**	A	C	G	A	G
0	0	0	0.06	T	C	T	C	C	T	C	G	G	T	C	**C**	T	T	C	C	C	C	**C**	T	**C**	A	C	G	A	G
0	0		0.06	T	G	T	G	T	C	A	A	A	C	T	**T**	C	A	G	A	C	C	**C**	G	**T**	C	T	C	A	G
0	0		0.03	C	C	T	C	C	T	C	G	A	C	T	**T**	C	A	G	C	C	C	**C**	T	**C**	A	T	G	A	G
0	0		0.03	T	G	T	G	T	C	A	A	A	C	T	**T**	C	A	G	A	C	C	**C**	G	**T**	C	T	G	G	G
0	0		0.02	T	G	T	G	T	C	A	A	A	C	T	**T**	C	A	G	A	C	C	**C**	G	**T**	C	T	G	A	G
0	0	0.08	0	na	na	na	na	na	na	na	na	A	C	T	**C**	C	A	G	A	C	C	**C**	G	**T**	na	C	na	G	A
0.03	0.01	0	0.01	C	C	T	G	C	T	C	G	A	T	C	**T**	C	A	G	C	C	C	**C**	G	**C**	C	C	G	A	G
0	0	0.11	0	C	C	T	G	C	T	C	G	A	T	C	**T**	C	A	G	C	C	G	**C**	G	**C**	C	C	G	A	G

The three SNPs (rs6503695m, rs744166 and rs957970) selected to tag the STAT3 LD-block haplotypes can differentiate 4 of the 5 haplotypes observed in the CEU population in Hapman 23a build. The increase in the MS protective haplotype marked with bold and italics is at least eight times more common in the non-African populations. The SNPs used in following haplotype analyses are bolded. The putative predisposing haplotype is in italics.

[a]The SNP rs744166 is in C/T orientation in the HapMap genotype database.

Table S8. The Haplotype Frequencies and Association Results by Study Set

	SO (93 MS, 342 ctrl)			Fin (534MS, 575 ctrl)			Meta[a] (2624 MS, 7220 ctrl)		
Haplotype	**MS**	**CTRL**	**P**	**MS**	**CTRL**	**P**	**MS**	**CTRL**	**P**
TAA[b]	0.516	0.592	0.256	0.569	0.622	0.0119	0.552	0.578	6.02E-05
CGG[c]	0.285	0.244	0.0630	0.289	0.250	0.0348	0.308	0.278	2.61E-05
TGG	0.134	0.126	0.0989	0.099	0.093	0.620	0.079	0.077	0.527
CGA	0.065	0.038	0.117	0.041	0.036	0.497	0.066	0.078	0.556

Results of the three SNP haplotype analysis (rs6503695, rs744166 and rs957970).
[a]The Meta subgroup includes the IMSGC US and UK sets[7], the Gene MSA Switzerland, The Netherlands and US sets[46], and the BWH[16] sets combined using CMH.
[b]The putative protective haplotype (TAA) is less common in cases compared to controls.
[c]The putative predisposing haplotype (CGG) is enriched in MS cases in all populations.