Hum Genet (2012) 131:87–97
DOI 10.1007/s00439-011-1053-2

ORIGINAL INVESTIGATION

Identification of a new susceptibility variant for multiple sclerosis in *OAS1* by population genetics analysis

Rachele Cagliani · Matteo Fumagalli · Franca R. Guerini · Stefania Riva · Daniela Galimberti · Giacomo P. Comi · Cristina Agliardi · Elio Scarpini · Uberto Pozzoli · Diego Forni · Domenico Caputo · Rosanna Asselta · Mara Biasin · Elvezia M. Paraboschi · Nereo Bresolin · Mario Clerici · Manuela Sironi

Received: 18 March 2011 / Accepted: 21 June 2011 / Published online: 7 July 2011

Abstract Contrasting results have been reported concerning the association of a splice-site polymorphism (rs10774671) in *OAS1* with multiple sclerosis (MS). We analysed two *OAS1* regions encompassing alternatively spliced exons. While the region carrying the splice-site variant is neutrally evolving, a signature of long-standing balancing selection was observed across an alternative exon 7. Analysis of variants in this exon identified an insertion/deletion polymorphism (rs11352835, A/−) that originates predicted products with distinct C termini. This variant is located along the major branch of the haplotype genealogy, suggesting that it may represent the selection target. A case/control study for MS indicated that rs11352835 is associated with disease susceptibility (for an allelic model with the deleted allele predisposing to MS, OR 1.27, 95% CI 1.072–1.513, $p = 0.010$). No association was found between rs10774671 and MS. As the two SNPs are in linkage disequilibrium in Europeans, the previously reported association between rs10774671 and MS susceptibility might be driven by rs11352835, possibly explaining the contrasting results previously observed for the splice-site polymorphism. Thus, we describe a novel susceptibility variant for MS in *OAS1* and show that population genetic analyses can be instrumental to the identification of selection targets and, consequently, of functional polymorphisms with an effect on phenotypic traits.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-011-1053-2) contains supplementary material, which is available to authorized users.

R. Cagliani · M. Fumagalli · S. Riva · U. Pozzoli · D. Forni · N. Bresolin · M. Sironi (✉)
Bioinformatic Lab, Scientific Institute IRCCS E. Medea, Via don L. Monza 20, 23842 Bosisio Parini, LC, Italy
e-mail: manuela.sironi@bp.lnf.it

F. R. Guerini · C. Agliardi
Laboratory of Molecular Medicine and Biotechnologies, Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

D. Galimberti · G. P. Comi · E. Scarpini · N. Bresolin
Department of Neurological Sciences, Dino Ferrari Centre, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, University of Milan, 20122 Milan, Italy

D. Caputo
Multiple Sclerosis Unit, Don C. Gnocchi Foundation ONLUS, IRCCS, S. Maria Nascente, Milan, Italy

R. Asselta · E. M. Paraboschi
Dipartimento di Biologia e Genetica per le Scienze Mediche, Università degli Studi di Milano, Milan, Italy

M. Biasin
DISC LITA Vialba, University of Milano, Milan, Italy

M. Clerici
Department of Biomedical Sciences and Technologies LITA Segrate, University of Milan, 20090 Milan, Italy

M. Clerici
Fondazione Don C. Gnocchi, IRCCS, 20148 Milan, Italy

Introduction

In humans four 2′,5′-oligoadenylate synthetase genes (*OAS1*, *OAS2*, *OAS3* and *OASL*) are located on the long arm of chromosome 12 and play a central role in the innate immune response against viruses. These enzymes are activated either by the presence of double strand RNA or by single strand RNA with secondary structure, and

catalyze the oligomerization of ATP into 2′,5′-linked oligoadenylates. These, in turn, bind to and activate the latent ribonuclease L (RNase L), which degrades viral and cellular RNA and blocks protein synthesis. In line with this function, human *OAS1* and its murine ortholog, *Oas1b*, modulate the susceptibility to viral infections (Lim et al. 2009; He et al. 2006; Hamano et al. 2005; Mashimo et al. 2002; Perelygin et al. 2002). In particular, variants in human *OAS1* have been associated with West Nile Virus and SARS infection (Lim et al. 2009; He et al. 2006; Hamano et al. 2005), and with liver fibrosis progression in HCV infected individuals (Li et al. 2009).

One of the most intensely studied variants in the human *OAS1* gene is a G>A substitution (rs10774671) at the splice-acceptor site of exon 6. The A allele abrogates the production of OAS1 p46 isoform which displays high antiviral activity (Bonnevie-Nielsen et al. 2005) resulting in the production of the p48 and p52 isoforms (Fig. 1). Additionally, alternative splicing of *OAS1* transcripts originates a p42 isoform which displays an alternative exon 5, and a p44 product deriving from skipping of exon 6 and inclusion of a downstream alternative exon (Fig. 1). A recent study indicated that the different protein products display variable activity in blocking

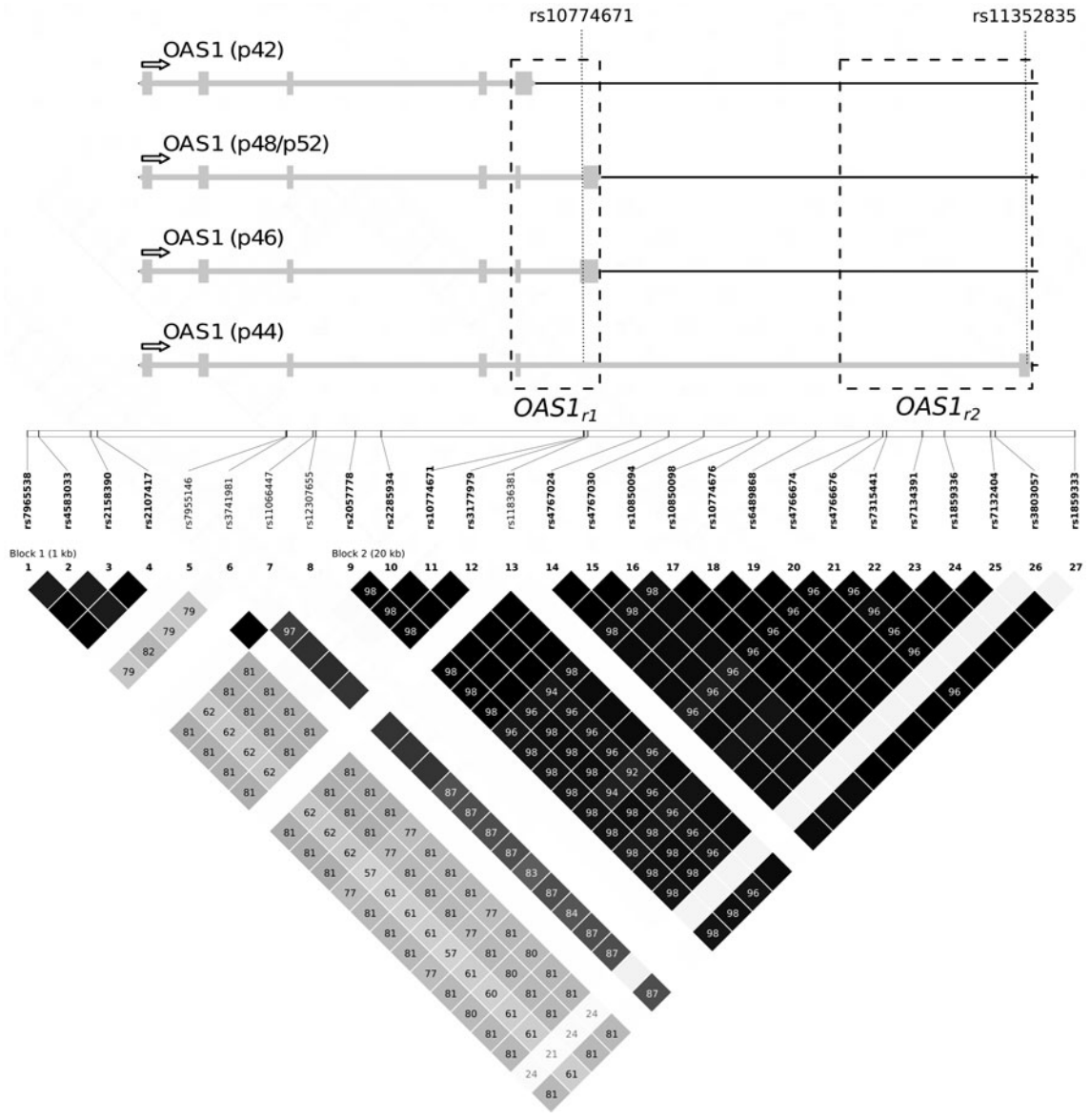

Fig. 1 Schematic representation of the *OAS1* gene region and alternative transcripts. Transcribed regions are shown in *grey* and the corresponding protein products are reported. The direction of transcription is indicated by the *arrows*. Alternatively spliced exons are shown in *darker grey*. The two regions we resequenced ($OAS1_{r1}$ and $OAS1_{r2}$) are indicated by the hatched lines. The *asterisk* shows the position of rs11352835; rs10774671 is *circled*. The LD (r^2) plot refers to CEU and data were derived from HapMap

Dengue virus replication via activation of RNase L (Lin et al. 2009).

In analogy to other genes involved in the antiviral response, polymorphisms and haplotypes in *OAS1* have been identified as potential risk factors for autoimmune conditions, including type 1 diabetes (T1D) and multiple sclerosis (MS). Interestingly, viral agents have been proposed to trigger or exacerbate both diseases. The association between *OAS1* polymorphisms and T1D is still controversial, with discordant results obtained in different studies (Qu et al. 2009; Smyth et al. 2006; Tessier et al. 2006; Field et al. 2005). In the case of MS, *OAS1* SNPs and haplotypes have been associated with both disease susceptibility and severity (Fedetz et al. 2006; O'Brien et al. 2010). In particular, the A allele of rs10774671 (the splice-altering variant in exon 6), which originates protein isoforms with lower antiviral activity, was shown to confer both an increased risk to develop MS and a higher likelihood to have a more severe form of disease in a population from Ireland (O'Brien et al. 2010). Conversely, a previous study in a Spanish cohort found no association between this variant and MS susceptibility, although the authors found a rare haplotype to be more common in patients versus controls (Fedetz et al. 2006).

Antiviral response genes are an obvious target of natural selection as virus-borne infections have represented, and still represent, a major threat to human populations. Our previous data (Fumagalli et al. 2010) suggest that some overlap may exist among genes subjected to virus-driven selective pressure and loci involved in the pathogenesis of MS. Population genetic analyses of candidate gene regions can be regarded as instrumental to the identification of selection targets and, consequently, of functional polymorphisms with an effect on phenotypic traits. Here we analysed two gene regions in *OAS1* and verified that while the region carrying the functional splice-site altering variant is neutrally evolving, long-standing balancing selection has shaped nucleotide diversity in a gene portion encompassing the alternative exon 7. A case/control association analysis indicated that variants in this downstream region, but not the splice-site altering variant, are associated with susceptibility to MS in an Italian cohort.

Materials and methods

HapMap samples and sequencing

Human genomic DNA from HapMap subjects (20 individuals for YRI, CEU and EAS) was obtained from the Coriell Institute for Medical Research. All analysed regions were PCR amplified and directly sequenced; primer sequences are available upon request. PCR products were treated with ExoSAP-IT (USB Corporation, Cleveland, OH, USA), directly sequenced on both strands with a Big Dye Terminator sequencing Kit (v3.1 Applied Biosystems) and run on an Applied Biosystems ABI 3130 XL Genetic Analyzer (Applied Biosystems). Sequences were assembled using AutoAssembler version 1.4.0 (Applied Biosystems), and inspected manually by two distinct operators. The genomic DNA of 3 *Pan troglodytes* was obtained from the Gene Bank of Primates, Primate Genetics, Germany (http://dpz.eu/index.php).

Patients, controls, and genotyping

For the case/control association study, 660 MS subjects (66% females) and 503 age- and sex-matched healthy controls (65% females) were recruited. All subjects were Italian of Caucasian origin. Median age was 42.5 (SD 12.1) and 43.9 (SD 20.1) years for MS and controls, respectively. Patients and controls were recruited at the MS Centre of Don Gnocchi Foundation in Milan and at Department of Neurological Sciences, University of Milan. All subjects gave informed consent according to protocols approved by the local Ethics Committees. MS patients underwent a standard battery of examinations, including medical history, physical and neurological examination, screening laboratory test, brain magnetic resonance imaging (MRI). All patients with MS fulfilled the McDonald's criteria (McDonald et al. 2001).

Genotyping of rs11352835 and rs10774671 was performed by direct resequencing, as described above, using genomic DNA extracted from peripheral blood.

The power to detect genetic association was estimated using the QUANTO software (version 1.2; http://hydra.usc.edu/GxE/) under a multiplicative mode of inheritance, assuming a disease prevalence in the Italian population (non-insular) of 0.0007 (Pugliatti et al. 2006), and a type I error rate of 0.05.

Data retrieval and haplotype construction

Genotype data for 5 kb regions from 238 resequenced human genes were derived from the NIEHS (National Institute of Environmental Health Sciences) SNPs Program web site (http://egp.gs.washington.edu). In particular, we selected genes that had been resequenced in populations of defined ethnicity including CEU, YRI and EAS (NIEHS panel 2).

Haplotypes were inferred using PHASE version 2.1 (Stephens et al. 2001; Stephens and Scheet 2005). Haplotypes for individuals resequenced in this study are available as online resource (Online Table 1).

Linkage disequilibrium analyses were performed using the Haploview (v. 4.1) (Barrett et al. 2005). Data for LD analysis were derived from HapMap.

Statistical analysis

Tajima's *D* (Tajima 1989), Fu and Li's *D** and *F** (Fu and Li 1993) statistics, as well as diversity parameters θ_W (Watterson 1975) and π (Nei and Li 1979) were calculated using *libsequence* (Thornton 2003). Calibrated coalescent simulations were performed using the *cosi* package (Schaffner et al. 2005) and its best-fit parameters for YRI, CEU, and EAS populations with 10,000 iterations. Coalescent simulations were conditioned on mutation rate and recombination rate. The maximum-likelihood-ratio HKA test was performed using the MLHKA software (Wright and Charlesworth 2004), as previously proposed (Fumagalli et al. 2009). Briefly, 16 reference loci were randomly selected among NIEHS loci shorter than 20 kb that have been resequenced in the 3 populations; the only criterion was that Tajima's *D* did not suggest the action of natural selection (i.e. Tajima's *D* is higher than the 5th and lower than the 95th percentiles in the distribution of NIEHS genes). In all analyses, the chimpanzee sequence was used as the out-group. All calculations were carried out in the R environment (Team R Development Core 2008).

Haplotype analysis and TMRCA calculation

Median-joining networks to infer haplotype genealogy were constructed using NETWORK 4.5 (Bandelt et al. 1999). Estimate of the time to the most common ancestor (TMRCA) was obtained using a phylogeny-based approach implemented in NETWORK 4.5 using a mutation rate based on the number of fixed differences between chimpanzee and humans. An additional TMRCA estimate derived from application of a maximum-likelihood coalescent method implemented in GENETREE (Griffiths and Tavare 1995; Griffiths and Tavare 1994). The method assumes an infinite-site model without recombination; therefore, haplotypes and sites that violate these assumptions need to be removed: for the analysis of $OAS1_{r2}$ we removed 1 variant. Again, the mutation rate μ was obtained on the basis of the divergence between human and chimpanzee and under the assumption both that the species separation occurred 6 MY ago (Glazko and Nei 2003) and of a generation time of 25 years. The migration matrix was derived from previous estimated migration rates (Schaffner et al. 2005). Using this μ and θ maximum likelihood (θ_{ML}), we estimated the effective population size parameter (N_e) which resulted equal to 22,838. With these assumptions, the coalescence time, scaled in $2N_e$ units, was converted into years. For the coalescence process, 10^6 simulations were performed.

Real-time PCR for the *OAS1* p44 isoform

Whole blood was collected from 21 Italian healthy controls by venipuncture in Vacutainer tubes containing EDTA (Becton-Dickinson, NJ, USA), and peripheral blood mononuclear cells (PBMC) were separated on lymphocyte separation medium (Organon Teknica, Malvern, PA). Total RNA was extracted PBMCs by using the acid guanidinium thiocyanate–phenol–chloroform method. The RNA was dissolved in RNase-free water, and purified from genomic DNA with RNase-free DNase (RQ1 DNase, Promega, Madison, WI, USA). One microgram of RNA was reverse transcribed into first-strand cDNA in a 20-µl final volume containing 1 µM random hexanucleotide primers, 1 µM oligo dT and 200 U Moloney murine leukemia virus reverse transcriptase (Clontech, Palo Alto, CA, USA). cDNA quantification for *OAS1* p44 and *GAPDH* was performed by real-time PCR (DNA Engine Opticon 2; MJ Research, Ramsey, USA). Reactions were performed using a SYBR Green PCR mix (RealMasterMix SYBR ROX, 5 PRIME) and the results were expressed as $\Delta\Delta C_t$ (where "C_t" is the cycle threshold) and presented as ratios between the target gene and the *GAPDH* housekeeping mRNA. Primers were designed to specifically amplify the p44 isoform of OAS1 and were located on exon 5 and 7 (Fig. 1) (primer sequences available upon request).

Results

Nucleotide diversity and neutrality tests

In order to study the evolutionary pattern of *OAS1* in human populations we resequenced three HapMap population samples (Yoruba, YRI; Europeans, CEU; East Asians, EAS) across two gene regions. Region 1 (thereafter referred to as $OAS1_{r1}$) covers exons 5 and 6, that undergo alternative splicing events to originate p46, p42, p48 and p52 (Fig. 1). Region 2 ($OAS1_{r2}$) is downstream exon 6 and covers the alternative exon 7, whose inclusion originates p44 (Fig. 1). The two regions are in strong linkage disequilibrium (LD) in CEU and EAS (Fig. 1; Online Fig. S1).

A total of 9 and 41 variants were identified in $OAS1_{r1}$ (2.5 kb) and $OAS1_{r2}$ (5.5 kb), respectively. In the second region we analysed, rs11352835 is an A insertion/deletion (A/−) polymorphism falling within the coding sequence of the alternative exon 7. Therefore, due to a reading frame shift, the variant potentially originates two C termini with different length and amino acidic composition. Analysis of genomic alignments and BLAST searches against resequenced primate genomes indicated that the A allele is ancestral and that sequences with homology to exon 7 can

be identified in hominoid primates (*Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus)*, Old World monkeys (*Macaca mulatta*, *Papio hamadryas*) and New World monkeys (*Callithrix jacchus*), but not in lemurs and non-primate mammals.

For the $OAS1_{r1}$ and $OAS1_{r2}$ regions we calculated nucleotide diversity by means of two indexes: θ_W (Watterson 1975), an estimate of the expected per site heterozygosity, and π (Nei and Li 1979), the average number of pairwise sequence nucleotide differences. In order to compare the values we obtained for the two *OAS1* regions, we calculated θ_W and π for 5 kb windows (thereafter referred to as reference windows) deriving from 238 genes resequenced by the NIEHS program in the same population samples; the percentile rank corresponding to $OAS1_{r1}$ and $OAS1_{r2}$ in the distribution of NIEHS gene values is reported in Table 1 and indicates that $OAS1_{r2}$ displays extremely high nucleotide diversity in both CEU and EAS; conversely, no exceptional values are observed for $OAS1_{r1}$, with the exclusion of YRI that shows reduced genetic variability in the region.

High levels of nucleotide diversity are consistent with the action of balancing selection as linked polymorphisms are maintained together with the selected variant(s). Another effect of balancing selection is a distortion of the site frequency spectrum (SFS) towards intermediate frequency alleles.

Common neutrality tests based on the SFS include Tajima's *D* (D_T) (Tajima 1989) and Fu and Li's *D** and *F** (Fu and Li 1993). D_T tests the departure from neutrality by comparing θ_W and π and positive values indicate an excess of intermediate frequency variants. Fu and Li's *F** and *D** are also based on SNP frequency spectra and differ from D_T in that they also take into account whether mutations occur in external or internal branches of a genealogy. Since, population history, in addition to selective processes, is known to affect the SFS, the significance of neutrality tests was evaluated by performing coalescent simulations with population genetics models that incorporate demographic scenarios (see "Materials and methods"). As above, we also applied an empirical comparison by calculating the percentile rank of D_T, *F** and *D** in the *OAS1* regions relative to 5 kb reference windows. Neutrality tests for $OAS1_{r2}$ region indicated departure from neutrality with significantly positive values for most statistics in CEU and EAS (Table 1). In line with these findings, D_T, as well as Fu and Li's *F** and *D** calculated for $OAS1_{r2}$ rank above the 95th percentile of the distribution of 5 kb reference windows in these two populations. Conversely, no departure from neutrality was observed for $OAS1_{r1}$ (Table 1) with the exception of CEU where D_T and *F** were significantly high. Yet these values are calculated over a very small number of segregating sites ($n = 4$) and are likely due to LD with $OAS1_{r2}$.

Overall, these results suggest that nucleotide diversity at $OAS1_{r2}$ has been shaped by balancing selection in European and Asian populations, with the $OAS1_{r1}$ region being neutrally evolving.

As mentioned above, a hallmark of balancing selection is an excess of polymorphism compared to neutral expectations. Indeed, our data (Table 1) indicate that nucleotide diversity indexes are extremely high for $OAS1_{r2}$ in CEU and EAS. Yet, polymorphism level also depends on local mutation rates; therefore, under neutral evolution, the amount of within- and between-species diversity is expected to be similar at all loci in the genome (Kimura 1983). The multi-locus HKA test was developed to verify this expectation (Wright and Charlesworth 2004). We

Table 1 Summary statistics for the $OAS1_{r1}$ and $OAS1_{r2}$ gene regions

Reg.	L[a]	P[b]	S[c]	θ_W[d]		π[e]		Tajima's *D*			Fu and Li's *D**			Fu and Li's *F**'		
				Value	Rank[f]	Value	Rank[f]	Value	Rank[f]	p[g]	Value	Rank[f]	p[g]	Value	Rank[f]	p[g]
$OAS1_{r1}$	2.5	YRI	4	3.72	0.025	2.49	0.020	−0.79	0.32	0.61	−1.10	0.22	0.75	−1.17	0.22	0.74
		CEU	4	3.72	0.28	7.62	0.75	2.49	0.98	<0.01	1.03	0.86	0.10	1.71	0.97	0.023
		EAS	6	5.59	0.59	6.80	0.67	0.58	0.67	0.32	1.19	0.95	0.08	1.17	0.89	0.12
$OAS1_{r2}$	5.5	YRI	11	4.73	0.13	3.74	0.15	−0.63	0.39	0.53	−0.19	0.55	0.38	−0.39	0.51	0.42
		CEU	40	17.20	0.98	31.10	0.99	2.84	0.98	<0.01	0.90	0.83	0.14	1.86	0.97	0.013
		EAS	37	15.91	0.99	21.92	0.98	1.32	0.87	0.13	1.64	>0.99	0.01	1.81	0.99	0.02

[a] Length of analyzed sequenced region (kb)

[b] Population (40 chromosomes were analysed in each population)

[c] Number of segregating sites

[d] Watterson's θ estimation per site ($\times 10^{-4}$)

[e] Nucleotide diversity per site ($\times 10^{-4}$)

[f] Percentile rank relative to a distribution of 238 5 kb segments from NIEHS genes

[g] p value applying demographic coalescent simulations

applied a multi-locus MLHKA (maximum-likelihood HKA) test by comparing polymorphism and divergence levels at the $OAS1_{r1}$ and $OAS1_{r2}$ genomic regions with 16 NIEHS genes resequenced in YRI, CEU and EAS. As shown in Table 2, a significant excess of polymorphism compared to divergence is observed for CEU and EAS in $OAS1_{r2}$, while no deviation from expectations is detectable at $OAS1_{r1}$.

Haplotype analysis

Further insight into the evolutionary history of a gene region can be gained by inferring haplotype genealogies. This has both a descriptive purpose (i.e. showing the relationship among alleles and their distribution in human populations) and can be used to test for selection. In particular, balancing selection is expected to result in two or more major haplotype clades with a deep coalescence time. Here we constructed haplotype genealogies using two approaches: a neighbour-joining network and a maximum-likelihood coalescent method implemented in GENETREE.

Consistently with the extended LD pattern (Fig. 1; Online Fig. S1), the haplotype network for the $OAS1_{r2}$ gene region presented no reticulations or recurrent mutations (Fig. 2). The haplotype genealogy is split into two major clades in CEU and EAS, while all African chromosomes cluster within clade B. In line with these results, calculation of population genetic differentiation (F_{ST}) between CEU and YRI resulted in a value of 0.366, corresponding to a percentile rank of 0.96 in the distribution of F_{ST} values calculated for 5 kb reference windows.

The absence of African chromosomes in clade A is not merely due to the relatively small sample of individuals we analysed, as HapMap data from 120 YRI chromosomes indicated that several SNPs along the major branches are monomorphic in this population.

In order to estimate the time to the most recent common ancestor (TMRCA) of the haplotype genealogy, we applied a phylogeny-based method (Bandelt et al. 1999). For this analysis only single base pair substitutions were included (i.e. rs11352835, A ins/del was excluded). Using a mutation rate based on the number of fixed differences with chimpanzee and a separation time of 6 million years (MY) (Glazko and Nei 2003), we estimated a TMRCA of 7.4 MY (SD 1.36 MY). Consistent with these results, GENETREE analysis (Fig. 3) yielded a TMRCA estimate of 5.61 MY (SD 0.88 MY). These coalesce times are much deeper than estimated under neutrality.

Given the deep coalescence time of $OAS1_{r2}$ haplotypes, we resequenced corresponding region in eight chimpanzees (*Pan troglodytes*) but no variants shared with humans was identified.

Finally, we verified that SNPs defining the two major haplotype clades do not affect (at least overtly) the inclusion in the transcript of the alternative exon 7: a PCR amplification of cDNA derived from lymphoblastoid cell lines of four subject homozygous for A clade haplotypes and four homozygous for clade B revealed similar levels of exon 7 inclusion (not shown). In order to further address this issue, we performed real-time PCR experiments to analyse the level of p44 expression in PBMCs derived from 21 healthy volunteers. No difference in p44 expression levels was observed among individuals with different genotype at rs11352835 (Kruskal–Wallis test, $2df$, $p = 0.54$) (Online Fig. S2).

OAS1 polymorphisms in MS

As mentioned above, *OAS1* has been recently described as a multiple sclerosis susceptibility gene (Fedetz et al. 2006; O'Brien et al. 2010). We therefore analysed a population of 660 subjects suffering from MS and 503 sex-matched healthy controls (HC). All subjects were Italian of Caucasian origin and were genotyped for two SNPs: rs10774671 (located at the splice-acceptor site of exon 6) and rs11352835 (the A ins/del polymorphism in the alternative exon 7) (Fig. 1). This latter was selected because of its putative functional role and as a consequence of its location in the haplotype phylogeny. Indeed, as shown in Fig. 2, rs11352835 is located on the basal branch of the haplotype network, suggesting that it may represent (or be in full LD with) the selection target. The two SNPs we selected for genotype analysis are in relatively tight LD ($r^2 = 0.87$) and both complied to Hardy–Weinberg equilibrium in both cases and controls. We found the allele frequency of both SNPs to be significantly different in MS compared to HC (Table 3), but only rs11352835 withstood Bonferroni correction for multiple tests (Fisher's exact test, Bonferroni corrected $p = 0.010$, OR 1.27, 95% CI 1.072–1.513). Comparison of genotype frequencies in the MS and HC samples indicated no significant difference for

Table 2 MLHKA test for the two gene regions we analysed

Region	Pop.[a]	MLHKA	
		k[b]	p
$OAS1_{r1}$	YRI	0.46	0.48
	CEU	1.16	0.74
	EAS	1.28	0.31
$OAS1_{r2}$	YRI	1.08	0.18
	CEU	6.11	3.73×10^{-6}
	EAS	7.18	5.42×10^{-6}

[a] Population

[b] Selection parameter ($k > 1$ indicates an excess of polymorphism relative to divergence)

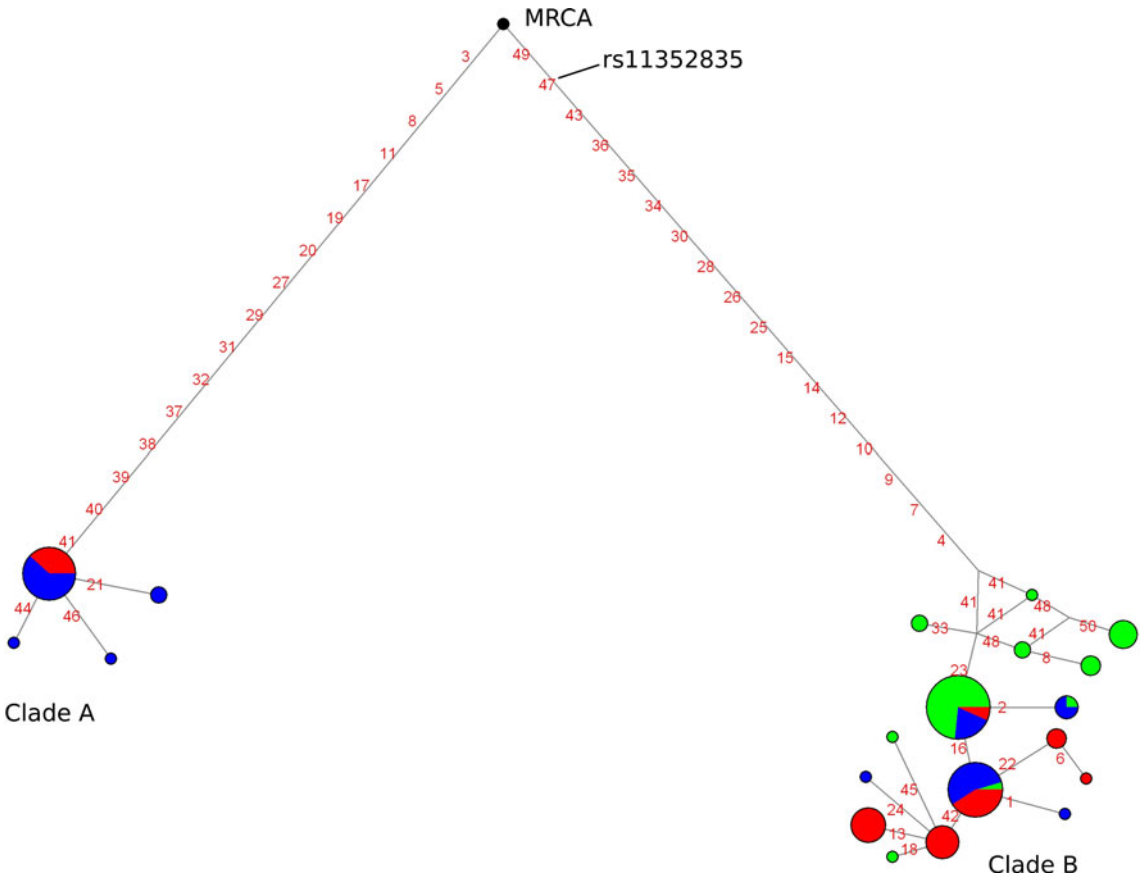

Fig. 2 Network analysis for $OAS1_{r2}$. *Each node* represents a different haplotype, with the size of the circle proportional to frequency. Nucleotide differences between haplotypes are indicated on the *branches of the network*. The position of rs11352835 (A insertion/deletion in the alternative exon 7) is shown. *Circles* are colour-coded according to population (*green* YRI, *blue* CEU, *red* EAS). The most recent common ancestor (MRCA) is also shown (*black circle*). The relative position of mutations along a branch is arbitrary

rs10774671 (Table 3). Conversely, the genotype distribution of rs11352835 was significantly different in the two cohorts (Bonferroni corrected $p = 0.030$, Table 3). The odds ratio for a recessive model with the deleted allele predisposing to MS was 1.41 (95% CI 1.103–1.811, $p = 0.0051$). Haplotype analysis identified no significant association.

Discussion

Most human genes undergo extensive alternative splicing, which is generally regarded as a powerful mechanism to increase proteome diversity (Modrek and Lee 2002). This is likely even more true for immune response genes, which are more frequently engaged in alternative splicing events compared to other functional classes (Modrek et al. 2001). Several studies have addressed the evolutionary forces underlying the appearance and fate of alternatively spliced exons in inter-specific comparisons (Xing and Lee 2005; Lu et al. 2009a, b; Ke et al. 2008; Zhang and Chasin 2006; Plass and Eyras 2006; Chen and Chuang 2006; Ermakova et al. 2006). Conversely, fewer analyses have focused on the selective patterns shaping the intra-specific diversity of alternatively spliced genes. Here we selected for analysis the two gene regions in *OAS1* that encompass alternatively spliced exons. As mentioned above, at least five transcripts originate from *OAS1*, one of these is genetically determined by the presence of a splice-acceptor variant and gives rise to protein products with different antiviral activity (Bonnevie-Nielsen et al. 2005; Lin et al. 2009). The splice-acceptor polymorphism has also been associated with altered susceptibility to viral infections (Lim et al. 2009), indicating this region as a possible selection target. Surprisingly, our resequencing analysis in three human populations showed that $OAS1_{r1}$ is neutrally evolving, as all tests failed to reject the null hypothesis of selective neutrality. Low diversity indexes were observed for this region in YRI, but the MLHKA test revealed that polymorphism levels were not exceptionally low compared to inter-specific diversity; moreover, calculation of Fay and Wu's H (Fay and Wu 2000), which detects an excess of high-frequency derived

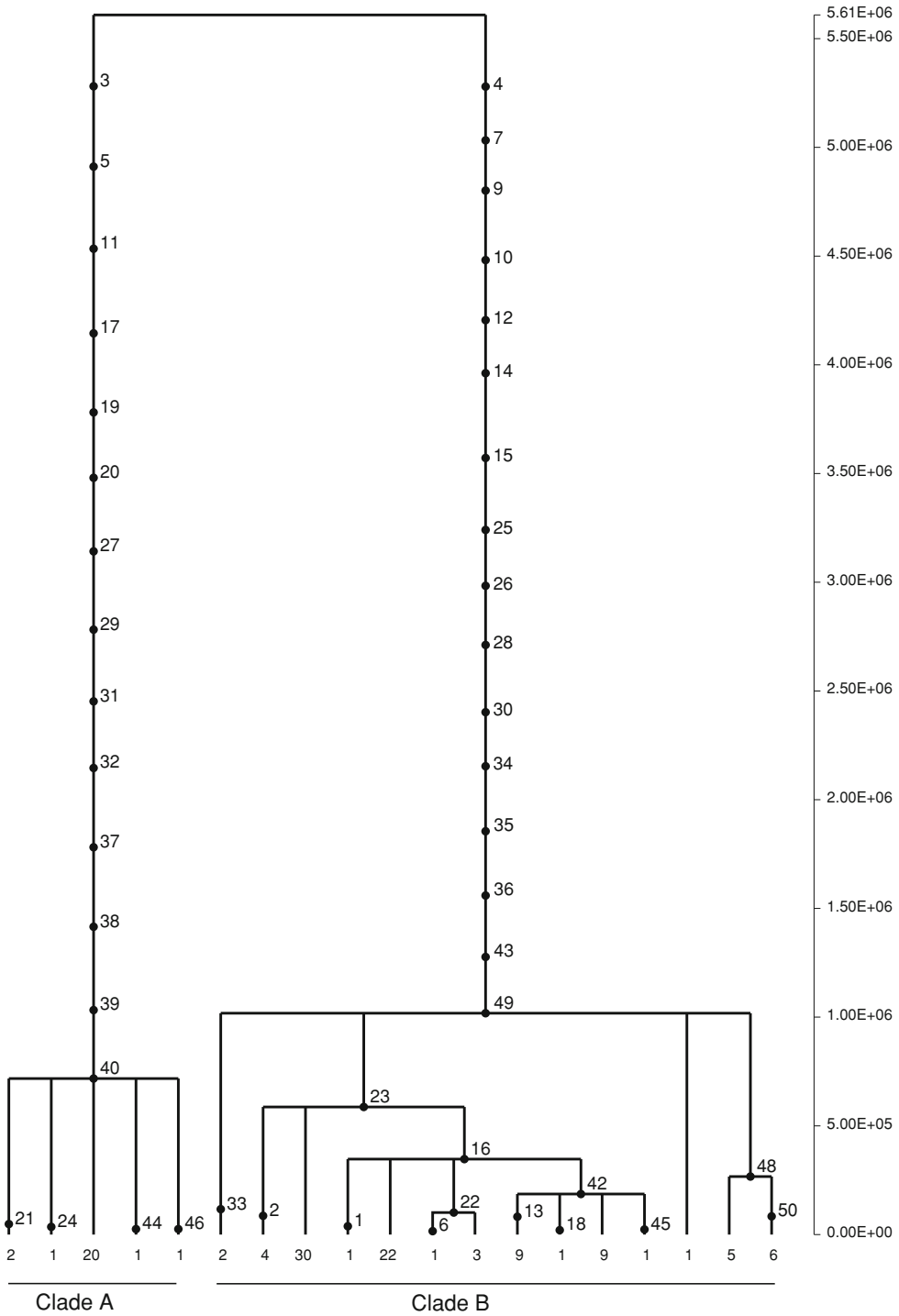

Fig. 3 GENETREE analysis for $OAS1_{r2}$. Mutations are represented as *black dots* and named for their physical position along the region. The absolute frequency of each haplotype is also reported. Note that mutation numbering does not correspond to that reported in Fig. 2

alleles, in both $OAS1_{r1}$ and $OAS1_{r2}$ yielded results consistent with selective neutrality in YRI (not shown), suggesting that weak negative selection (as opposed to positive selection) may be acting on $OAS1_{r1}$ in African populations.

Conversely, data herein indicate that the region encompassing the alternatively spliced exon 7 has been a target of long-standing balancing selection in Europeans and Asians. Inclusion of this exon in the *OAS1* transcript originates a p44 isoform (Fig. 1) whose antiviral activity has been poorly characterized. A recent report indicated that, among all *OAS1* isoforms, only p42 and p46 display strong antiviral activity against Dengue virus (Lin et al. 2009). The reason for the differential activity of OAS1 proteins is presently unknown, as all isoforms display the catalytic OAS domain (which is located at the N terminus). Our analysis indicated that an A insertion/deletion

Table 3 Association of *OAS1* SNPs with susceptibility to MS

SNP	Alleles/genotypes	Counts		p^a (corrected)
		MS ($n = 660$)	HC ($n = 503$)	
Alleles				
rs10774671	G/A	488/832	416/590	n.s.
rs11352835	A/−	469/851	415/591	0.010
Genotypes				
rs10774671	GG/AG/AA	95/298/267	87/242/174	n.s.
rs11352835	AA/A−/−−	87/295/278	83/249/171	0.030

n.s. not significant

[a] Bonferroni corrected *p* values (*p* values were corrected for two tested SNPs)

polymorphism (A/-, rs11352835) is located on the major branch that separates the two haplotype clades, indicating that this variant, or one in linkage to it, may represent the selection target in $OAS1_{r2}$. The derived deleted allele determines a frame shift that is predicted to result in the inclusion of 22 additional amino acid residues compared to p44. Given the differential activity observed against Dengue virus for p42 and p44 (which only differ by 35 terminal amino acids) (Lin et al. 2009), it is tempting to speculate that similarly, such a minor change in sequence composition may alter the antiviral potential of OAS1. One interesting possibility is that the different composition of the C terminus affects the specificity of OAS1 against different viral species. An alternative explanation for our findings is that the selection target is accounted for by a regulatory variant/haplotype with a role in transcription modulation, as several expression QTLs for *OAS1* have been identified in the gene region encompassing exon 7 (http://eqtl.uchicago.edu/).

Although further analyses will be required to identify the selection target in $OAS1_{r2}$, our data provide evidence that a functional polymorphism is located within this region, as signatures of long-standing balancing selection are expected to extend over relatively short genomic regions (Charlesworth 2006). Thus, we reasoned that a susceptibility variant for MS might be located in this region, as well, possibly explaining the previous contrasting results obtained for the splice-acceptor SNP in different MS cohorts. Indeed, while a strong relationship was recently described between the G allele of rs10774671 and protection from MS in a population from Ireland (O'Brien et al. 2010), a previous report in a Spanish sample revealed no such association, despite similar cohort sizes in the two studies (about 400 cases/controls) (Fedetz et al. 2006). Our results on an Italian sample indicated that the A allele of rs11352835, but not the splice-site variant, is significantly associated with protection from MS. Given that the two variants have very similar frequency in the Italian sample, the estimated power to detect association is also comparable, and reaches 78% assuming an OR of 1.27 and a sample size as the one we analysed. The relatively tight LD of the two SNPs ($r^2 = 0.87$) suggests that association between rs10774671 and MS susceptibility is driven by rs11352835. Therefore, minor differences in the level of LD in cohorts with distinct geographic origin might account for the contrasting results previously observed for the splice-site polymorphism. In this respect, it is also worth mentioning that insufficient sample size is regarded as a major explanation for several inconsistent associations reported in the scientific literature. This consideration may apply to previous studies on the role of *OAS1* in MS, which relied on relatively limited patient/control samples (Fedetz et al. 2006; O'Brien et al. 2010). Here we genotyped larger MS/HC cohorts compared to previous studies; our data confirm that *OAS1* represents an MS susceptibility gene, and refine the association signal to a variant different from, but in linkage with, the one previously described to associate with disease.

In summary, our population genetic analysis indicates that the *OAS1* gene region harbouring the functional splice-acceptor polymorphism is neutrally evolving, and the variant itself shows no association with MS in Italians. Conversely, a new MS susceptibility variant was identified within or in proximity to the alternatively spliced exon 7, a region targeted by balancing selection. In line with previous reports (Grossman et al. 2010), our work demonstrates that population genetic approaches can be regarded as instrumental to the identification of selection targets and, consequently, of functional polymorphisms with an effect on phenotypic traits. This observation suggests that the ever-increasing availability of resequencing data, obviously including the 1000 Genomes Project (http://browser.1000genomes.org/), can be exploited to identify selection signatures at the genome-wide level and this information, in turn, may be used to prioritize variants to be typed in association analysis for complex diseases and traits.

Acknowledgments MC is supported by grants from Istituto Superiore di Sanità "Programma Nazionale di Ricerca sull' AIDS", the EMPRO and AVIP EC WP6 Projects, the nGIN EC WP7 Project, the Japan Health Science Foundation, 2008 Ricerca Finalizzata (Italian Ministry of Health), 2008 Ricerca Corrente (Italian Ministry of Health), Progetto FIRB RETI: Rete Italiana Chimica Farmaceutica CHEM-PROFARMA-NET [RBPR05NWWC], and Fondazione CARIPLO. DG is supported by grant Giovani Ricercatori 2007 (D.lgs 502/92) from Italian Ministry of Health. We wish to thank

Dr Christian Roos from the Gene Bank of Primates, Primate Genetics (Germany) for sending chimpanzee genomic DNA.

References

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. doi:10.1093/bioinformatics/bth457

Bonnevie-Nielsen V, Field LL, Lu S, Zheng DJ, Li M, Martensen PM, Nielsen TB, Beck-Nielsen H, Lau YL, Pociot F (2005) Variation in antiviral 2′,5′-oligoadenylate synthetase (2′5′AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene. Am J Hum Genet 76:623–633. doi:10.1086/429391

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2:e64. doi:10.1371/journal.pgen.0020064

Chen FC, Chuang TJ (2006) The effects of multiple features of alternatively spliced exons on the K(A)/K(S) ratio test. BMC Bioinforma 7:259. doi:10.1186/1471-2105-7-259

Ermakova EO, Nurtdinov RN, Gelfand MS (2006) Fast rate of evolution in alternatively spliced coding regions of mammalian genes. BMC Genomics 7:84. doi:10.1186/1471-2164-7-84

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Fedetz M, Matesanz F, Caro-Maldonado A, Fernandez O, Tamayo JA, Guerrero M, Delgado C, Lopez-Guerrero JA, Alcina A (2006) OAS1 gene haplotype confers susceptibility to multiple sclerosis. Tissue Antigens 68:446–449. doi:10.1111/j.1399-0039.2006.00694.x

Field LL, Bonnevie-Nielsen V, Pociot F, Lu S, Nielsen TB, Beck-Nielsen H (2005) OAS1 splice site polymorphism controlling antiviral enzyme activity influences susceptibility to type 1 diabetes. Diabetes 54:1588–1591

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Fumagalli M, Cagliani R, Pozzoli U, Riva S, Comi GP, Menozzi G, Bresolin N, Sironi M (2009) Widespread balancing selection and pathogen-driven selection at blood group antigen genes. Genome Res 19:199–212. doi:10.1101/gr.082768.108

Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Bresolin N, Clerici M, Sironi M (2010) Genome-wide identification of susceptibility alleles for viral infections through a population genetics approach. PLoS Genet 6:e1000849. doi:10.1371/journal.pgen.1000849

Glazko GV, Nei M (2003) Estimation of divergence times for major lineages of primate species. Mol Biol Evol 20:424–434

Griffiths RC, Tavare S (1994) Sampling theory for neutral alleles in a varying environment. Philos Trans R Soc Lond B Biol Sci 344:403–410. doi:10.1098/rstb.1994.0079

Griffiths RC, Tavare S (1995) Unrooted genealogical tree probabilities in the infinitely-many-sites model. Math Biosci 127:77–98

Grossman SR, Shylakhter I, Karlsson EK, Byrne EH, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander ES, Schaffner SF, Sabeti PC (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. Science 327:883–886. doi:10.1126/science.1183863

Hamano E, Hijikata M, Itoyama S, Quy T, Phi NC, Long HT, Ha LD, Ban VV, Matsushita I, Yanai H, Kirikae F, Kirikae T, Kuratsuji T, Sasazuki T, Keicho N (2005) Polymorphisms of interferon-inducible genes OAS-1 and MxA associated with SARS in the Vietnamese population. Biochem Biophys Res Commun 329:1234–1239. doi:10.1016/j.bbrc.2005.02.101

He J, Feng D, de Vlas SJ, Wang H, Fontanet A, Zhang P, Plancoulaine S, Tang F, Zhan L, Yang H, Wang T, Richardus JH, Habbema JD, Cao W (2006) Association of SARS susceptibility with single nucleic acid polymorphisms of OAS1 and MxA genes: a case–control study. BMC Infect Dis 6:106. doi:10.1186/1471-2334-6-106

Ke S, Zhang XH, Chasin LA (2008) Positive selection acting on splicing motifs reflects compensatory evolution. Genome Res 18:533–543. doi:10.1101/gr.070268.107

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Li CZ, Kato N, Chang JH, Muroyama R, Shao RX, Dharel N, Sermsathanasawadi R, Kawabe T, Omata M (2009) Polymorphism of OAS-1 determines liver fibrosis progression in hepatitis C by reduced ability to inhibit viral replication. Liver Int 29:1413–1421. doi:10.1111/j.1478-3231.2009.02061.x

Lim JK, Lisco A, McDermott DH, Huynh L, Ward JM, Johnson B, Johnson H, Pape J, Foster GA, Krysztof D, Follmann D, Stramer SL, Margolis LB, Murphy PM (2009) Genetic variation in OAS1 is a risk factor for initial infection with West Nile virus in man. PLoS Pathog 5:e1000321. doi:10.1371/journal.ppat.1000321

Lin RJ, Yu HP, Chang BL, Tang WC, Liao CL, Lin YL (2009) Distinct antiviral roles for human 2′,5′-oligoadenylate synthetase family members against dengue virus infection. J Immunol 183:8035–8043. doi:10.4049/jimmunol.0902728

Lu H, Lin L, Sato S, Xing Y, Lee CJ (2009a) Predicting functional alternative splicing by measuring RNA selection pressure from multigenome alignments. PLoS Comput Biol 5:e1000608. doi:10.1371/journal.pcbi.1000608

Lu H, Lin L, Sato S, Xing Y, Lee CJ (2009b) Predicting functional alternative splicing by measuring RNA selection pressure from multigenome alignments. PLoS Comput Biol 5:e1000608. doi:10.1371/journal.pcbi.1000608

Mashimo T, Lucas M, Simon-Chazottes D, Frenkiel MP, Montagutelli X, Ceccaldi PE, Deubel V, Guenet JL, Despres P (2002) A nonsense mutation in the gene encoding 2′-5′-oligoadenylate synthetase/L1 isoform is associated with West Nile virus susceptibility in laboratory mice. Proc Natl Acad Sci USA 99:11311–11316. doi:10.1073/pnas.172195399

McDonald WI, Compston A, Edan G, Goodkin D, Hartung HP, Lublin FD, McFarland HF, Paty DW, Polman CH, Reingold SC, Sandberg-Wollheim M, Sibley W, Thompson A, van den Noort S, Weinshenker BY, Wolinsky JS (2001) Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. Ann Neurol 50:121–127

Modrek B, Lee C (2002) A genomic view of alternative splicing. Nat Genet 30:13–19. doi:10.1038/ng0102-13

Modrek B, Resch A, Grasso C, Lee C (2001) Genome-wide detection of alternative splicing in expressed sequences of human genes. Nucleic Acids Res 29:2850–2859

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5269–5273

O'Brien M, Lonergan R, Costelloe L, O'Rourke K, Fletcher JM, Kinsella K, Sweeney C, Antonelli G, Mills KH, O'Farrelly C, Hutchinson M, Tubridy N (2010) OAS1: a multiple sclerosis susceptibility gene that influences disease severity. Neurology 75:411–418. doi:10.1212/WNL.0b013e3181ebdd2b

Perelygin AA, Scherbik SV, Zhulin IB, Stockman BM, Li Y, Brinton MA (2002) Positional cloning of the murine flavivirus resistance gene. Proc Natl Acad Sci USA 99:9322–9327. doi:10.1073/pnas.142287799

Plass M, Eyras E (2006) Differentiated evolutionary rates in alternative exons and the implications for splicing regulation. BMC Evol Biol 6:50. doi:10.1186/1471-2148-6-50

Pugliatti M, Rosati G, Carton H, Riise T, Drulovic J, Vecsei L, Milanov I (2006) The epidemiology of multiple sclerosis in Europe. Eur J Neurol 13:700–722. doi:10.1111/j.1468-1331.2006.01342.x

Qu HQ, Polychronakos C; Type I Diabetes Genetics Consortium (2009) Reassessment of the type I diabetes association of the OAS1 locus. Genes Immun 10(Suppl 1):S69–S73. doi:10.1038/gene.2009.95

Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D (2005) Calibrating a coalescent simulation of human genome sequence variation. Genome Res 15:1576–1583. doi:10.1101/gr.3709305

Smyth DJ, Cooper JD, Lowe CE, Nutland S, Walker NM, Clayton DG, Todd JA (2006) No evidence for association of OAS1 with type 1 diabetes in unaffected siblings or type 1 diabetic cases. Diabetes 55:1525–1528

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462. doi:10.1086/428594

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989. doi:10.1086/319501

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

R Development Core Team (2008) R: a language and environment for statistical computing. Vienna, Austria

Tessier MC, Qu HQ, Frechette R, Bacot F, Grabs R, Taback SP, Lawson ML, Kirsch SE, Hudson TJ, Polychronakos C (2006) Type 1 diabetes and the OAS gene cluster: association with splicing polymorphism or haplotype? J Med Genet 43:129–132. doi:10.1136/jmg.2005.035212

Thornton K (2003) Libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics 19:2325–2327

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wright SI, Charlesworth B (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168:1071–1076. doi:10.1534/genetics.104.026500

Xing Y, Lee C (2005) Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci USA 102:13526–13531. doi:10.1073/pnas.0501213102

Zhang XH, Chasin LA (2006) Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. Proc Natl Acad Sci USA 103:13427–13432. doi:10.1073/pnas.0603042103