



# A multi-step genomic approach prioritized *TBKBP1* gene as relevant for multiple sclerosis susceptibility

Melissa Sorosina<sup>1</sup> · Nadia Barizzone<sup>2</sup> · Ferdinando Clarelli<sup>1</sup> · Santosh Anand<sup>3</sup> · Sara Lupoli<sup>4</sup> · Erika Salvi<sup>4,5</sup> · Eleonora Mangano<sup>6</sup> · Roberta Bordoni<sup>6</sup> · Tina Roostaei<sup>7</sup> · Elisabetta Mascia<sup>1</sup> · Miriam Zuccalà<sup>2</sup> · Domizia Vecchio<sup>8</sup> · Paola Cavalla<sup>9</sup> · Silvia Santoro<sup>1</sup> · Laura Ferrè<sup>1,10</sup> · Alen Zollo<sup>11</sup> · PROGEMUS · Cristina Barlassina<sup>4</sup> · Daniele Cusi<sup>6,12</sup> · Vittorio Martinelli<sup>10</sup> · Giancarlo Comi<sup>10</sup> · Maurizio Leone<sup>13</sup> · Massimo Filippi<sup>10,14,15,16</sup> · Nikolaos A. Patsopoulos<sup>17,18,19,20</sup> · Philip L. De Jager<sup>7</sup> · Gianluca De Bellis<sup>6</sup> · Federica Esposito<sup>1,10</sup> · Sandra D'Alfonso<sup>2</sup> · Filippo Martinelli Boneschi<sup>11,21</sup>

Received: 3 December 2021 / Revised: 23 March 2022 / Accepted: 24 March 2022 / Published online: 12 May 2022  
© The Author(s) 2022, corrected publication 2022

## Abstract

**Background** Over 200 genetic loci have been associated with multiple sclerosis (MS) explaining ~50% of its heritability, suggesting that additional mechanisms may account for the “missing heritability” phenomenon.

**Objective** To analyze a large cohort of Italian individuals to identify markers associated with MS with potential functional impact in the disease.

**Methods** We studied 2571 MS and 3234 healthy controls (HC) of continental Italian origin. Discovery phase included a genome wide association study (1727 MS, 2258 HC), with SNPs selected according to their association in the Italian cohort only or in a meta-analysis of signals with a cohort of European ancestry (4088 MS, 7144 HC). Top associated loci were then tested in two Italian cohorts through array-based genotyping (903 MS, 884 HC) and pool-based target sequencing (588 MS, 408 HC). Finally, functional prioritization through conditional eQTL and mQTL has been performed.

**Results** Top associated signals overlap with already known MS loci on chromosomes 3 and 17. Three SNPs (rs4267364, rs8070463, rs67919208), all involved in the regulation of *TBKBP1*, were prioritized to be functionally relevant.

**Conclusions** No evidence of novel signal of association with MS specific for the Italian continental population has been found; nevertheless, two MS loci seems to play a relevant role, raising the interest to further investigations for *TBKBP1* gene.

**Keywords** Multiple sclerosis · Genetics · Susceptibility · *TBKBP1* · Genome-wide association study

## Introduction

Multiple Sclerosis (MS) is a complex autoimmune disease caused by the interplay of multiple genetic and environmental factors [1]. Large genome wide association studies (GWAS) provided important clues for the identification of

genetic causes with more than 233 loci being discovered [2–4]. Although they are robustly associated, the identification of causative variants and the functional mechanism influencing MS pathogenesis has been identified only in few cases [5, 6]. Moreover, these studies rely on the inclusion of different populations to reach very large cohorts necessary to detect small effects like those observed for MS; however, different frequencies of disease variants across the different populations may mask the existence of additional associations. This was clearly evidenced in a GWAS in Sardinians, a genetically isolated population, which led to the identification of a novel MS variant in the *TNFSF13B* gene, whose allelic frequency allows its identification only in Sardinians and Southern European populations [6].

In the present project, we aim to analyse a large cohort of more than 5500 Italian individuals to identify markers

---

Melissa Sorosina and Nadia Barizzone contributed equally to this work as first authors. Federica Esposito, Sandra D'Alfonso, and Filippo Martinelli Boneschi contributed equally to this work as last authors.

---

Membership of the consortia are listed in the acknowledgement.

---

✉ Filippo Martinelli Boneschi  
filippo.martinelli@unimi.it

Extended author information available on the last page of the article

associated with MS with potential functional impact in the disease. To boost the power of the analysis, we meta-analysed association results from different cohorts and leveraged quantitative molecular traits on adequately large resources for functional SNPs prioritization.

## Materials and methods

### Subjects

A total of 2571 Italian MS patients and 3234 unrelated HC have been recruited across Italian centres by three consortia (PROGRESSO, PROGEMUS and HYPERGENES). All participants were of Caucasian ancestry.

Based on the genotyping platform used for the genotyping, this entire set of samples was divided in four cohorts:

- a)  $ITA_{GWAS}$ : it was composed of 776 MS and 1315 HC. MS samples were previously genotyped in the context of the International Multiple Sclerosis Consortium (IMSGC) at genome-wide level at the Wellcome Trust Sanger Institute using the Illumina<sup>®</sup> Human660-Quad chip [2]. Healthy individuals (mean ( $\pm$  SD) age of  $53.4 \pm 9.9$  years and no abnormal findings on physical and neurological examination) have been collected within the HYPERGENES project (European Network for Genetic-Epidemiological Studies; [www.hypergenes.eu](http://www.hypergenes.eu)) [7] in Italy and genotyped at the University of Milan, using the Illumina<sup>®</sup> Infinium 1 M-duo BeadChips.
- b)  $ITA_{iChip}$ : it was composed of 964 MS patients and 1054 matched controls non-overlapping with the  $ITA_{GWAS}$  cohort. Samples were previously genotyped in the context of the IMSGC at the Wellcome Trust Sanger Institute using the Illumina<sup>®</sup> ImmunoChip covering 196,524 SNPs and 184 non-HLA genomic regions with potential immunological function [3].
- c)  $ITA_{OA}$ : it was composed of 917 MS patients and 924 controls. Samples were genotyping in the present study for selected SNPs using custom and pre-designed TaqMan assays (Thermo Fisher Scientific, Waltham, USA) assembled on two OpenArray chips run on the TaqMan Open Array Genotyping System (Thermo Fisher Scientific, Waltham, USA) or, in the context of the IMSGC, with the Illumina MS replication chip (Illumina, San Diego, USA), a custom array covering more than 300,000 SNPs localized in MS associated loci [4].
- d)  $ITA_{NGS}$ : subjects were chosen starting from the  $ITA_{GWAS}$  and  $ITA_{iChip}$  cohorts based on their genetic risk burden [8], estimated as a weighted sum of risk alleles in non-HLA known MS-associated loci [9–11]; 600

MS patients with a higher genetic risk (based on the distribution of the entire Italian cohort) were selected together with age- and sex-matched 408 HC, according to sex and Principal Components (PCs) to account for population stratification.  $ITA_{NGS}$  cohort underwent pooled-sequencing as previously described [12] using the Agilent SureSelect target enrichment method (Agilent Technologies, Santa Clara, USA) according to the manufacturer's protocol.

The four cohorts are partially overlapping, as shown in Online Resource 1 fig. S1.

Demographic information for the mentioned cohorts is listed in Table 1, while details on genetic data generation are described in the Online Resource 1.

Additionally to the Italian cohorts, a total of 4088 MS patients and 7,144 HC of European descent (“NonIT-EUR” cohort) have been included in the analyses for the “meta-analysis approach” (see below). This represents a combination of 6 individual sample collections previously described [13].

All patients were diagnosed according to Poser and McDonald criteria and further revisions [14].

### Design of the study

Figure 1 shows the workflow of the study. Briefly, we followed a multi-step approach including a discovery step at genome-wide level followed by replication and fine-mapping through next-generation sequencing of the top-associated loci. Finally, the effect of the identified variants on gene expression and DNA methylation was evaluated to prioritize markers with potential functional role.

### Discovery

We initially investigated the presence of association signals specific for the Italian population by performing a GWAS in the  $ITA_{GWAS}$  cohort alone (here called the “genome-wide Italian approach”), this led to the selection of SNPs ( $P < 5 \times 10^{-7}$ ) to further replicate in the replication step.

An alternative approach (here called the “meta-analysis approach”) was designed to increase the sample size of the discovery phase and detect association signals with a lower impact on MS susceptibility not detectable when studying only the  $ITA_{GWAS}$  cohort. Specifically, two additional cohorts were included in the study design: the  $ITA_{iChip}$ , composed by Italian individuals genotyped at relevant immunological genomic regions, and the NonIT-EUR dataset, which represents a large cohort of non-Italian individuals for which

**Table 1** Baseline demographic and clinical features of Italian individuals included in the study

	ITA <sub>GWAS</sub>		ITA <sub>iChip</sub>		ITA <sub>OA</sub>		ITA <sub>NGS</sub>	
	MS	HC	MS	HC	MS	HC	MS	HC
<i>n</i>	776	1315	964	1054	917	924	600	408
Sex								
Males	277	878	323	610	294	591	238	172
Females	499	437	641	444	623	333	362	236
Disease course								
BOMS	561	–	875	–	709	–	499	–
PPMS	125	–	58	–	86	–	61	–
CIS	7	–	9	–	5	–	6	–
Undefined	83	–	22	–	117	–	34	–
Age at onset (average ± SD)	31.7 ± 10.1	–	30.3 ± 10.0	–	31.0 ± 10.3	–	30.2 ± 10.0	–
Disease duration (average ± SD), years	11.3 ± 8.5	–	11.2 ± 7.5	–	11.0 ± 8.7	–	11.3 ± 8.2	–

*BOMS* Bout-onset MS patients (including relapsing–remitting and secondary-progressive), *PPMS* progressive MS patients, *CIS* Clinically isolated syndrome, *undefined* MS patients with no definition of diagnosis

genome-wide genotyping data are available. According to this approach, we selected SNPs with suggestive evidence of association ( $P < 0.005$ ) in the ITA<sub>GWAS</sub> cohort alone (to detect signals at genome-wide level), as well as performing a meta-analysis between ITA<sub>GWAS</sub> and ITA<sub>iChip</sub> cohorts. A meta-analysis with the NonIT-EUR dataset was then performed for the selected signals and significant SNPs ( $P < 5 \times 10^{-7}$ ) were considered for replication.

## Replication

We defined loci for replication as described in the Online Resource 1. Due to the known strong associations within the MHC region with MS [4], SNPs located in the HLA locus were excluded.

Genotyping was performed using TaqMan assays and the TaqMan Open Array Genotyping System (Thermo Fisher Scientific, Waltham, USA) or the MS Chip [4] (Illumina, San Diego, USA). Single-SNP logistic regression using sex as covariate was conducted under additive model of inheritance. Meta-analyses between the discovery and replication cohorts were performed assuming a fixed-effect model.

## Fine mapping with next generation sequencing (NGS)

We sequenced the Linkage Disequilibrium (LD) blocks containing the replicated associated signals in the ITA<sub>NGS</sub> cohort using the Agilent SureSelect kit (Agilent Technologies, Santa Clara, USA) and a pool-based approach on the GalIx platform (Illumina, San Diego, USA). Bioinformatics

pipeline and variant calling were performed as described [12]. Fisher exact test comparing MS and HC was performed on allelic counts, no correction based on sex or ancestry was included since HC were individually matched with MS patients, as described in the “Subjects” paragraph. For methodological details see Online Resource 1.

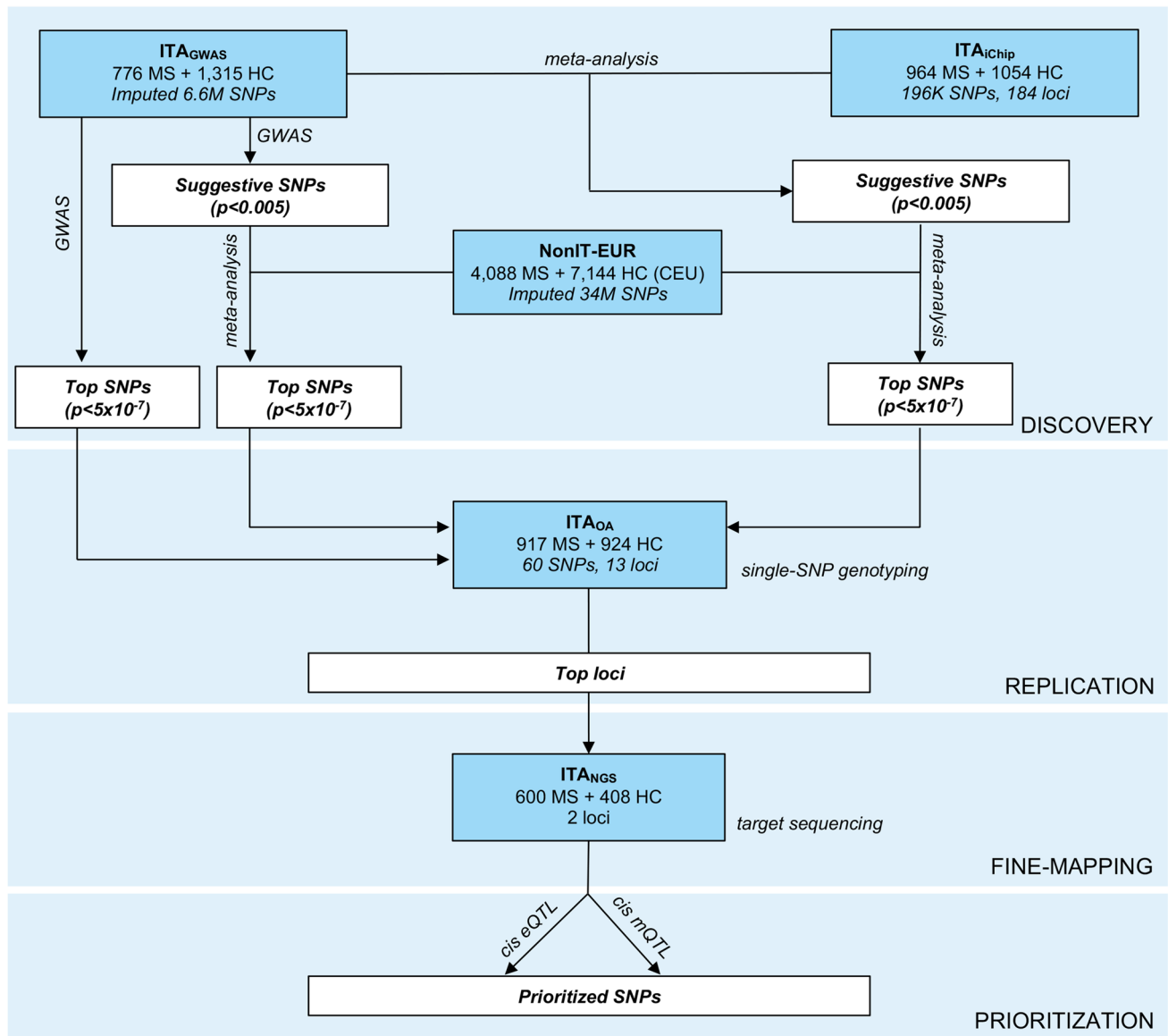
## Prioritization

Identified associated variants were prioritized according to their *cis* effect on expression or CpG methylation by performing conditional *cis*-eQTL and conditional *cis*-mQTL analyses. *Cis*-eQTL analysis was performed taking advantage of the GTEx Portal v6p [15] dataset (release phs000424.v6.p1) for 7 MS-related tissues (whole blood, Brain Frontal\_Cortex (BA9), Brain Cortex, Cerebellar Hemisphere, Cerebellum, Anterior cingulate cortex (BA24), Hypothalamus). *Cis*-mQTL analysis was performed on CD4<sup>+</sup> T-cells collected from 156 MS patients within the Comprehensive Longitudinal Investigation of Multiple Sclerosis at the Brigham and Women's Hospital (CLIMB study [16, 17]). Both analyses were performed using QTLtools [18]. For details see Online Resource 1.

## Results

### Discovery

As described in the method section, the “genome-wide Italian approach” includes the ITA<sub>GWAS</sub> cohort as discovery



**Fig. 1** Workflow of the study. The overall workflow of the study is shown. The discovery phase includes two Italian cohorts (ITAGWAS and ITAChip) from which a set of top MS associated SNPs were identified through a genome-wide approach (in the ITAGWAS cohort only) or a meta-analytic approach including also the NonIT-EUR dataset.

The identified top SNPs ( $P < 5 \times 10^{-7}$ ) were further tested for replication in a third independent Italian cohort (ITAOA) leading to the confirmation of two associated loci. Those loci were sequenced (ITANGS cohort) and the associated SNPs underwent prioritization through conditional *cis* eQTL and mQTL analyses

dataset. After quality control (QC), the association analysis was performed in a final dataset of 737 MS and 1291 HC individuals and ~6.6 million SNPs. Seventy-one variants reached a  $p$  value of association  $< 5 \times 10^{-7}$ , within 26 distinct loci (Online Resource 2, Online Resource 1 Fig. S2).

The “meta-analysis approach” includes both the ITAGWAS and ITAChip cohorts as discovery datasets, with signals with suggestive evidence of association ( $P < 0.005$ ) selected to be meta-analysed with the NonIT-EUR. Regarding the ITAGWAS cohort alone, 32,309 SNPs ( $P < 0.005$ ) were meta-analysed

with the NonIT-EUR leading to a list of 49 SNPs with a  $p$ -value of association  $< 5 \times 10^{-7}$  (Online Resource 3A). The meta-analysis of the two Italian discovery datasets led to the identification of 1373 SNPs suggestive of association ( $P < 0.005$ ). Meta-analysis of these SNPs with the NonIT-EUR dataset resulted in 21 SNPs with a  $p$  value  $< 5 \times 10^{-7}$  (Online Resource 3B) within 5 loci. Overall, and given the fact that some signals overlapped between the different analyses, 127 signals across 35 loci were selected (Online Resource 2 and Online Resource 3).

## Replication

Starting from the 127 signals identified through the discovery analyses (Online Resource 2 and Online Resource 3), we selected 60 SNPs for replication as representative of the identified genomic loci (see Online Resource 1 for details and Online Resource 2 and Online Resource 3 for the final list). After genotyping and QC (see Online Resource 1 for details), 48 SNPs within 30 loci were tested for association in the ITA<sub>OA</sub> dataset (Table 2).

No SNPs identified through the “genome-wide Italian approach” showed evidence of association, while according to the “meta-analysis approach” (thus starting from the two ITA<sub>GWAS</sub> and ITA<sub>iChip</sub> cohorts), the locus 1 on chromosome 3 and locus 8 on chromosome 17 confirmed their association, with rs338603 (OR = 1.27,  $P = 8.01 \times 10^{-4}$ ) and rs669607 (OR = 1.23,  $P = 4.47 \times 10^{-3}$ ) and rs8070463 (OR = 1.24,  $P = 3.43 \times 10^{-3}$ ) as the top associated SNPs. These associations were confirmed also performing a meta-analysis excluding the NonIT-EUR cohort and including only dataset of Italian origin ( $P = 3.73 \times 10^{-7}$ ,  $6.01 \times 10^{-7}$  and  $1.72 \times 10^{-5}$  respectively for rs338603, rs669607 and rs8070463) (Table 2, Online Resource 1 Fig. S3).

## Fine mapping

The replicated locus on chromosome 3 spans from position 28024996 to 28124996 (hg19 assembly). This is an intergenic region including the non-coding genes *LINC01980* (~ 123 kb) and *LINC01981* (~ 150 kb) and the coding genes *EOMES* (~ 160 kb) and *CMC1* (~ 150 kb) as the closest downstream genes. A total of 608 variants were identified through the sequencing approach. Comparing to the post-QC imputed ITA<sub>GWAS</sub> dataset, the inclusion of a sequencing step allowed to investigate additional 74 common variants (MAF  $\geq 0.01$ ), as well as 636 rare variants (Minor allele Frequency (MAF)  $< 0.01$ ), while 109 were excluded and not tested in the ITA<sub>NGS</sub> cohort due to the stringent QC applied. Fisher’s exact test on sequenced data revealed the association of 12 variants (Bonferroni-adjustment  $\alpha = 5.87 \times 10^{-5}$ ). The most associated SNPs were rs669607 and rs427221 ( $P = 5.02 \times 10^{-13}$ ) (Fig. 2a, Online Resource 4).

The replicated locus on chromosome 17 spans a large LD block of ~ 500 kb (chr17:45309693–45824486, hg19 assembly) which includes 8 genes: *ITGB3*, *EFCAB13* (*C17orf57*), *NPEPPS*, *KPNB1*, *TBKBPI*, *TBX21*, *MRPL45P2* (pseudogene) and *LOC102724508* (a non-coding RNA, with alias *THCAT158*). In this region, a total of 2,428 variants were identified through the sequencing approach. Comparing to the post-QC imputed ITA<sub>GWAS</sub> dataset, the inclusion of the fine mapping step allowed to investigate

additional 446 common variants and 2,395 rare variants, while 341 SNPs were not tested in the ITA<sub>NGS</sub> cohort for the sequencing QC. Applying a pipeline which allows an accurate allele frequency estimation from NGS pooled data [12] to perform the association analysis, we found that 202 SNPs were significantly associated after Bonferroni-adjustment ( $\alpha = 1.57 \times 10^{-5}$ ) (Fig. 2b, Online Resource 4). The strongest association signal was rs4239162 ( $P = 1.04 \times 10^{-20}$ ), intronic to *KPNB1*.

## Locus on chromosome 3

The locus on chromosome 3 is a small intergenic genomic region, with apparently no evidence of chromatin annotation (Online Resource 1 Fig. S4), except for the presence of sporadic weak enhancers. *Cis* eQTL analysis did not reveal any association between the 12 SNPs identified in the genetic analyses and the expression of nearby genes.

## Locus on chromosome 17: functional SNPs prioritization

We sought to characterize the identified MS risk variants prioritizing those with a putative functional impact. We then performed a conditional *cis*-eQTL analysis in 7 tissues relevant for MS according to data generated within the GTEx project [15]. As shown in Online Resource 5A, we found 7 SNP-gene pairs associated in the tested tissues. Among them, rs4267364 showed a consistent association with MS in all the steps of the study (Table 3). This SNP is intronic to *TBKBPI* and the MS associated allele (rs4267364<sub>A</sub>) is associated to a higher expression of this gene in blood (Fig. 3a). According to the chromatin state prediction in 14 epigenomes of immune-related cells profiled by the Roadmap Epigenomics [19] and ENCODE [20] projects, this SNP is localized in a putative enhancer, supporting its regulatory function (Online Resource 1 Fig. S5).

To explore additional mechanisms of regulation, we performed a conditional *cis*-mQTL in CD4<sup>+</sup> cells collected from MS patients. The Online Resource 5B lists the significant SNP-CpG pairs resulted from this analysis. Among the identified mQTL associations, one SNP associated to MS according to the discovery strategy and at least one of the replication steps, rs8070463, was found to be associated also with the methylation status of cg08452456 (Table 3, Fig. 3b), exonic to *TBKBPI*. An additional SNP, rs67919208, associated with the methylation status of cg12183861 (Table 3, Fig. 3c), localized in the 5’UTR region of *TBKBPI*, was also found to be associated with MS (discovery step). Both rs8070463 and rs67919208 had the MS risk variant associated with a higher CpG methylation (Table 3).

All the three prioritized variants suggest an involvement of *TBKBPI* gene, thus we then evaluated the expression of

**Table 2** Association results in the replication sample

SNP	LOCUS	A1	Replication				Meta-analysis discovery and replication				
			OR	L95	U95	<i>P</i>	OR	L95	U95	<i>P</i>	<i>I</i> <sup>2</sup>
<b>Table 2A</b>											
rs1595367	locus 27	T	0.80	0.66	0.98	0.03	1.26	1.11	1.44	5.12E-04	97.04
rs1914544	locus 17	C	0.87	0.75	1.02	0.08	1.20	1.08	1.33	7.40E-04	96.94
rs13151412	locus 12	T	0.88	0.75	1.02	0.09	1.15	1.03	1.28	1.29E-02	95.88
rs11492289	locus 18	G	0.89	0.78	1.02	0.10	0.82	0.74	0.91	1.35E-04	97.89
rs6993986	locus 15	C	0.90	0.77	1.04	0.15	0.84	0.75	0.93	1.17E-03	96.58
rs12872924	locus 23	T	0.90	0.78	1.04	0.17	0.85	0.76	0.95	2.75E-03	96.01
rs61500084	locus 11	A	0.91	0.78	1.05	0.19	1.16	1.05	1.28	4.67E-03	95.06
rs35493137	locus 26	T	1.15	0.92	1.44	0.21	1.51	1.31	1.74	1.34E-08	89.3
rs12873058	locus 23	T	0.91	0.78	1.06	0.22	1.18	1.06	1.32	1.87E-03	95.79
rs6011709	locus 31	T	1.11	0.94	1.31	0.22	1.37	1.22	1.53	6.81E-08	91.15
rs12453667	locus 26	G	0.74	0.45	1.21	0.23	1.55	1.33	1.80	1.22E-08	89.61
rs9852162	locus 11	T	0.92	0.79	1.06	0.24	1.20	1.08	1.32	3.77E-04	95.99
rs7316186	locus 22	A	0.93	0.81	1.06	0.28	0.80	0.73	0.89	1.39E-05	97.45
rs3589	locus 31	T	1.09	0.93	1.29	0.28	1.30	1.15	1.47	1.61E-05	89.49
rs2278819	locus 14	G	0.88	0.69	1.11	0.29	0.77	0.65	0.90	9.84E-04	95.01
rs663743	locus 5	G	1.07	0.92	1.24	0.39	0.78	0.70	0.87	5.21E-06	91.18
rs2510066	locus 5	C	1.07	0.92	1.24	0.39	0.78	0.70	0.87	4.72E-06	91.23
rs2508634	locus 21	T	0.92	0.73	1.14	0.44	1.27	1.07	1.50	5.60E-03	94.6
rs6594535	locus 3	T	1.05	0.91	1.22	0.48	1.24	1.12	1.38	3.53E-05	89.31
rs11229374	locus 19	T	0.93	0.76	1.14	0.49	0.79	0.69	0.91	9.04E-04	94.06
rs2933291	locus 10	G	1.06	0.89	1.27	0.50	1.30	1.14	1.49	7.62E-05	90.33
rs528650	locus 16	A	1.18	0.68	2.07	0.56	0.46	0.33	0.64	5.94E-06	86.37
rs12972735	locus 28	T	0.96	0.83	1.11	0.57	1.20	1.09	1.32	2.58E-04	94.17
rs548827	locus 16	T	1.15	0.66	2.02	0.63	0.46	0.33	0.65	7.59E-06	87
rs76053860	locus 24	G	0.97	0.65	1.45	0.88	0.55	0.41	0.72	1.84E-05	94.2
rs6875879	locus 13	A	0.99	0.85	1.16	0.93	1.23	1.10	1.36	1.65E-04	92.37
16:89463366	locus 25	R	0.99	0.86	1.15	0.93	1.21	1.09	1.34	3.64E-04	92.64
rs9927904	locus 25	G	1.00	0.87	1.16	0.96	1.21	1.09	1.34	3.01E-04	91.94
rs6591385	locus 20	T	0.00	0.00	Inf	1.00	2.06	1.57	2.70	1.70E-07	0

SNP	LOCUS	A1	Replication				Meta-analysis discovery and replication					Meta-analysis with only Italians cohorts				
			OR	L95	U95	<i>P</i>	OR	L95	U95	<i>P</i>	<i>I</i> <sup>2</sup>	OR	L95	U95	<i>P</i>	<i>I</i> <sup>2</sup>
<b>Table 2B</b>																
<b>rs338603</b>	<b>locus 1</b>	<b>C</b>	<b>1.27</b>	<b>1.11</b>	<b>1.46</b>	<b>8.01E-04</b>	<b>1.21</b>	<b>1.14</b>	<b>1.28</b>	<b>1.77E-10</b>	<b>0</b>	<b>1.29</b>	<b>1.17</b>	<b>1.42</b>	<b>3.73E-07</b>	<b>0</b>
rs8070463	locus 8	T	1.24	1.08	1.44	3.43E-03	1.21	1.14	1.28	1.79E-11	0	1.25	1.13	1.38	1.72E-05	0
rs438613	locus 1	C	1.23	1.06	1.42	5.74E-03	1.20	1.13	1.27	9.68E-10	0	1.27	1.15	1.40	2.67E-06	0
rs741174	locus 7	C	1.22	1.06	1.41	7.19E-03	1.18	1.11	1.24	5.48E-09	0	1.22	1.11	1.35	7.59E-05	0
rs741175	locus 7	T	1.19	1.03	1.37	1.52E-02	1.16	1.10	1.22	9.70E-09	0	1.20	1.09	1.33	1.80E-04	0
rs4267364	locus 8	A	1.19	1.02	1.39	2.44E-02	1.18	1.11	1.24	1.07E-08	0	1.22	1.10	1.36	2.37E-04	0
rs1883832	locus 9	T	1.16	1.00	1.36	0.051	1.17	1.11	1.24	4.02E-08	0	1.20	1.08	1.34	8.29E-04	0
rs6065926	locus 9	A	1.16	0.99	1.35	0.059	1.17	1.11	1.24	3.18E-08	0	1.21	1.09	1.35	4.67E-04	0
rs2729590	locus 6	T	1.20	0.98	1.48	0.080	1.37	1.23	1.53	1.16E-08	52.8	1.41	1.22	1.62	2.56E-06	75.9
rs56038902	locus 33	T	1.08	0.92	1.27	0.346	1.19	1.11	1.27	3.61E-07	43.9	1.18	1.05	1.32	4.28E-03	60.5
rs55857387	locus 33	T	1.07	0.91	1.26	0.404	1.19	1.11	1.27	2.43E-07	53.6	1.19	1.06	1.33	2.86E-03	71.4
rs529279	locus 3	C	1.03	0.87	1.22	0.722	1.14	1.08	1.20	1.32E-06	55.2	1.19	1.08	1.32	6.59E-04	86.1
rs2729589	locus 6	T	1.12	0.58	2.16	0.726	1.33	1.20	1.48	6.37E-08	0	1.45	1.21	1.73	4.46E-05	0

**Table 2** (continued)

SNP	LOCUS	AI	Replication				Meta-analysis discovery and replication					Meta-analysis with only Italians cohorts				
			OR	L95	U95	<i>P</i>	OR	L95	U95	<i>P</i>	<i>I</i> <sup>2</sup>	OR	L95	U95	<i>P</i>	<i>I</i> <sup>2</sup>
<b>rs669607</b>	<b>locus 1</b>	<b>C</b>	<b>1.23</b>	<b>1.07</b>	<b>1.42</b>	<b>4.47 × 10<sup>-3</sup></b>	<b>1.18</b>	<b>1.12</b>	<b>1.24</b>	<b>1.31 × 10<sup>-10</sup></b>	<b>0</b>	<b>1.23</b>	<b>1.13</b>	<b>1.33</b>	<b>6.01 × 10<sup>-7</sup></b>	<b>0</b>
rs11878602	locus 34	C	1.23	1.03	1.46	2.04 × 10 <sup>-2</sup>	1.16	1.10	1.23	3.18 × 10 <sup>-8</sup>	0	1.20	1.10	1.32	9.54 × 10 <sup>-5</sup>	25.8
rs1870071	locus 34	C	1.22	1.03	1.46	2.24 × 10 <sup>-2</sup>	1.16	1.10	1.23	3.04 × 10 <sup>-8</sup>	0	1.21	1.10	1.33	8.58 × 10 <sup>-5</sup>	32.0
rs1883832	locus 9	T	1.17	1.00	1.36	4.62 × 10 <sup>-2</sup>	1.16	1.10	1.22	1.40 × 10 <sup>-8</sup>	0	1.17	1.08	1.27	2.82 × 10 <sup>-4</sup>	0
rs6065926	locus 9	A	1.16	1.00	1.36	0.053	1.16	1.11	1.23	8.71 × 10 <sup>-9</sup>	0	1.18	1.08	1.28	1.32 × 10 <sup>-4</sup>	0
rs11079784	locus 8	C	1.11	0.96	1.29	0.150	1.16	1.11	1.22	8.21 × 10 <sup>-10</sup>	0	1.15	1.07	1.25	4.66 × 10 <sup>-4</sup>	0
rs11870935	locus 8	A	1.10	0.95	1.27	0.200	1.15	1.10	1.21	1.44 × 10 <sup>-8</sup>	0	1.14	1.06	1.24	1.07 × 10 <sup>-3</sup>	0
rs11256593	locus 35	T	1.06	0.92	1.23	0.409	1.15	1.10	1.20	3.94 × 10 <sup>-9</sup>	11.3	1.12	1.04	1.22	4.61 × 10 <sup>-3</sup>	0

Table 2c

For each SNP, the association data in the replication cohort and results from the meta-analyses between discovery and replication cohorts (including and not including the NonIT-EUR cohort) are reported. Results for the genome-wide Italian approach are reported in the upper panel (A), while results for the approaches including the NonIT-EUR are reported in panel B (discovery sample derived from ITA<sub>GWAS</sub> alone) and C (SNPs identified through the meta-analysis between ITA<sub>GWAS</sub> and ITA<sub>iChip</sub>)

AI: reference allele (the same as defined in Table 1); OR: odds ratio; L95: lower-bound of 95%-confidence interval; U95: upper bound of 95%-confidence interval; *P*: *P* value of association; *I*<sup>2</sup>: heterogeneity index.

The SNPs with an association beyond the multiple testing correction (Bonferroni threshold of *P* value: 1.7 × 10<sup>-3</sup>, 3.6 × 10<sup>-3</sup>, 6.25 × 10<sup>-3</sup> respectively for A, B and C) are highlighted in bold. No significant association was found in the replication cohort for SNPs selected through the Italian genome-wide approach

*TBKBPI* across immune cells according to the DICE dataset [21]. As shown in Online Resource 1 fig. S6, *TBKBPI* is expressed across several immune cell types, with the highest levels in activated T cells (CD4<sup>+</sup> and CD8<sup>+</sup>) and Natural Killer cells.

## Discussion

Large scale international studies showed a highly polygenic architecture for MS genetic susceptibility [4]; nonetheless, the identified variants explain ~ 50% of the genetic predisposition to MS, suggesting that additional signals and disease mechanisms exist. This “missing heritability” could be explained, at least partially, by the presence of population-specific associations that cannot be fully captured by array-based international studies, as already shown studying the Sardinian [6] and German populations [22].

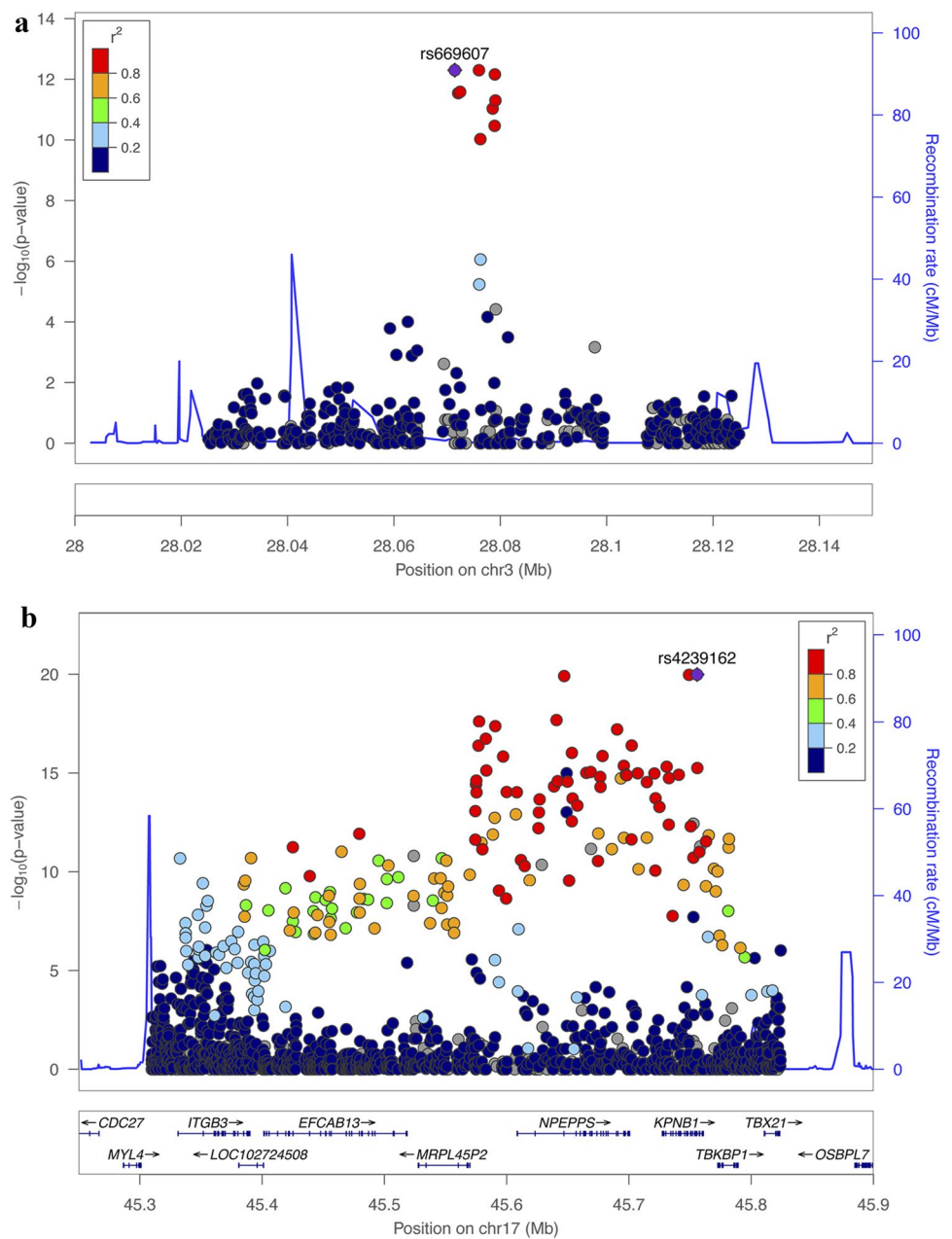
To investigate the presence of strong association signals enriched in the Italian population, we initially performed a genome-wide association study (referred along the paper as the “genome-wide Italian approach”) in a cohort of Italian individuals that led us to the identification of several loci potentially associated with MS, most of them never reported in previous studies (Online Resource 2). However, replication did not confirm any of these associations, supporting the idea that no strong signals exist in the continental Italian population. Nevertheless, we can’t exclude that other events

occurred, including population stratification, genetic heterogeneity, or an insufficient statistical power [23].

We then followed an alternative approach, exploring also signals with suggestive evidence of association and meta-analysing them with a cohort of European descent with larger adequate sample size and statistical power. The genome-wide Italian approach would have the potential to identify signals with a strong association with MS specific for the Italian population, while the second approach combined the advantage of using an ancestrally homogeneous cohort as the driving cohort of loci selection with the statistical power gained using a larger cohort. Specifically, this second approach, referred as “meta-analysis approach”, highlighted the presence of two loci representing the main results of this effort. Both lie in loci already known to be associated with MS [2–4]: an intergenic genomic region on chromosome 3 (~ 100 Kb) and a large locus of ~ 500 kb on chromosome 17.

To better characterize the two selected loci and to prioritize putative causal variants, we then performed a deep target sequencing followed by functional prioritization, allowing to investigate additional associated SNPs, not covered by arrays or imputation. In the case of the locus on chromosome 3, this analysis led to the identification of 12 associated SNPs, with rs669607 and rs427221 as the top associated ones [2]. No clear function is suggested, since this is an intergenic region with no annotation suggestive of specific regulatory function [4]. Nevertheless, very recently, a regulatory effect on the upstream gene *EOMES* in T cells

**Fig. 2** Regional association plots of replicated loci in ITA<sub>NGS</sub> cohort. Regional association plots for locus 1 on chromosome 3 (a) and locus 8 on chromosome 17 (b) designed with LocusZoom (<http://locuszoom.sph.umich.edu>) are shown. For each locus, the negative log-transformed *P* values (left y-axis) and the recombination rate (right y-axis) according to the 1000 Genomes Project (Nov 2014, Europeans only) are plotted over the genomic position (hg19). Each symbol represents one SNP, with the most associated SNP marked in purple and shading of the other points based on the linkage disequilibrium with the top SNP. Positions of genes, if present, are shown below the plot



has been highlighted [24], in line with previous studies supporting a role of this gene in MS [25].

On the contrary, the locus on chromosome 17 is more complex, with several genes and potentially regulatory regions. Sequencing of this region led to the identification of 202 signals of associations with the top associated signals localized in the downstream part of the region, as suggested also by the discovery analysis. This high number of associated signals could be explained by the strong linkage disequilibrium characterizing this locus (Fig. 2b), which further complicates the identification of the causative signal of association. We therefore resorted to exploit regulatory data for prioritization of the identified SNPs. Several

studies reported an enrichment of regulatory regions within MS and other complex diseases GWAS loci [24, 26–30], thus we performed a conditional *cis* eQTL analysis to either overcome the presence of the large LD block and to combine the results with the obtained genetic association. This analysis highlighted rs4267364, which is an intronic polymorphism to *TBKBP1*, localized in a putative enhancer. This SNP is associated with the expression of *TBKBP1* in blood, but not in brain-related tissues, with the MS risk allele rs4267364<sub>A</sub> associated with a higher expression of *TBKBP1* ( $P = 1.71 \times 10^{-23}$ ). We also investigated additional mechanisms of regulation performing a conditional *cis* mQTL analysis to identify SNPs associated with the methylation



**Table 3** MS associated signals showing independent eQTL and mQTL associations

SNP	GENE	SNP POSITION	eQTL			ITA <sub>GWAS</sub>			Meta-analysis between ITA <sub>GWAS</sub> and NonIT-EUR			Replication			ITA <sub>NGS</sub>		
			A1	A2	P	SLOPE	P	OR	P	OR	P	OR	P	OR	P	OR	P
rs4267364	TBKBPI	17:45781264	G	A	1.71 × 10 <sup>-23</sup>	-0.29	3.12 × 10 <sup>-3</sup>	0.81	1.42 × 10 <sup>-7</sup>	0.85	2.44 × 10 <sup>-2</sup>	0.84	9.61 × 10 <sup>-9</sup>	0.58			
			mQTL			ITA <sub>GWAS</sub>			Meta-analysis between ITA <sub>GWAS</sub> and NonIT-EUR			Replication			ITA <sub>NGS</sub>		
			SNP POSITION	A1	A2	CpG POSITION	P	SLOPE	P	OR	P	OR	P	OR	P	OR	P

**Table 3A**

rs67919208	cg12183861	17:45771873	C	T	17:45773468	1.62 × 10 <sup>-12</sup>	-0.68	1.07 × 10 <sup>-3</sup>	0.80	1.36 × 10 <sup>-9</sup>	0.84	<sup>a</sup>	-	<sup>b</sup>	-		
rs8070463	cg08452456	17:45768836	C	T	17:45786439	2.05 × 10 <sup>-17</sup>	-0.79	1.54 × 10 <sup>-3</sup>	0.80	1.34 × 10 <sup>-9</sup>	0.84	3.43 × 10 <sup>-3</sup>	0.81	<sup>b</sup>	-		

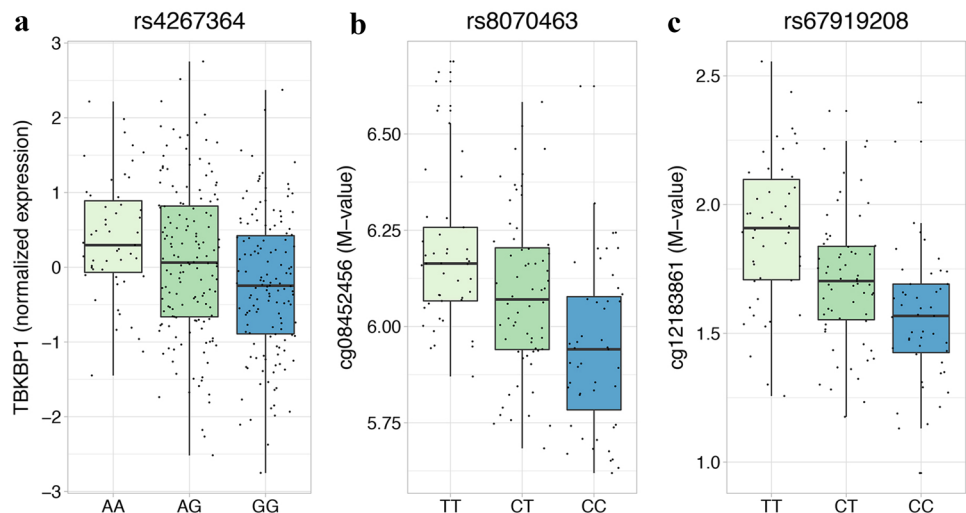
Signals associated with MS according to at least two of the steps performed (discovery: meta-analysis between ITA<sub>GWAS</sub> and NonIT-EUR  $P < 5 \times 10^{-7}$ ; replication:  $P < 3.6 \times 10^{-3}$ , target sequencing:  $P < 1.57 \times 10^{-5}$ ) that show also independent eQTL (A) or mQTL (B) association are shown. For the SNP with eQTL effect, the SNP-associated gene pair is indicated, together, in the eQTL panel, with the backward nominal  $P$  value of association and the backward regression slope referred to A1 in whole blood. For the SNPs with an mQTL association, the SNP-CpG pairs are indicated together, in the mQTL panel, with the backward nominal  $P$  value of association and the backward regression slope referred to A1. When available, genetic association results are reported. No SNP fulfilling the mentioned criteria is present in the ITA<sub>chip</sub> dataset, thus  $P$  values are not reported

A1 reference allele for QTL analyses and ORs, A2 alternative allele,  $P$   $P$  value of association, OR Odds ratio

<sup>a</sup>SNP not selected for the replication phase

<sup>b</sup>SNP failing the QC in the sequencing step

**Fig. 3** Prioritized eQTL and mQTL effects. eQTL and mQTL effects are depicted in the figures. **a** *TBKBP1* normalized expression values in whole blood are plotted against the genotype of rs4267364; **b** cg08452456 *M* values are plotted against the genotype of rs8070463; **c** cg12183861 *M* values are plotted against the genotype of rs67919208. Representation is in box-plot format (Tukey method) with each dot representing one individual



state of CpG in CD4<sup>+</sup> lymphocytes of MS patients. Several associations were found (Online Resource 5B), with two SNPs, rs8070463 and rs67919208, which were also selected by the genetic approach. These two SNPs were found to be associated with the methylation status of respectively cg08452456, exonic to *TBKBP1*, and cg12183861, localized in the 5'UTR of *TBKBP1*. Interestingly enough, rs8070463 SNP has been previously associated with ankylosing spondylitis, a rare autoimmune condition [31]. It is well established that methylation of regulatory regions is an important mechanism [32]: DNA methylation in promoter regions is known to induce gene silencing [33], while it has been reported that DNA methylation in transcribed regions seems to correlate with an induced gene expression [34]. Given their genomic position, we could hypothesize that the influence on the methylation status of the mentioned CpGs could in turn lead to an upregulation of *TBKBP1*, as suggested by eQTL analysis on GTEx dataset, with the MS risk allele being associated with an upregulation of *TBKBP1* in whole blood ( $P = 1.03 \times 10^{-20}$  and  $2.17 \times 10^{-20}$  respectively for rs8070463 and rs67919208), although we could not exclude that this reflects a LD with the eQTL leading SNP rs4267364 ( $r^2 = 0.66$  for both the SNPs). Functional experiments testing methylation and expression profiles in same samples are necessary to clarify the direction of the signal.

All of the three prioritized SNPs seems to be involved in the regulation of the expression of *TBKBP1*, suggesting that an upregulation of this gene could be relevant in MS pathogenesis. *TBKBP1* codes for the TBK1 binding protein 1, an adaptor protein able to bind TBK1 and involved in the TNF $\alpha$ /NF $\kappa$ B pathway [35]. Contrary to *TBKBP1*, for which few reports are available regarding its function, *TBK1* is involved in several immune processes, including T-cell homeostasis and migration cells from lymph nodes to the CNS in the MS murine model [36]. Although the direct involvement of *TBKBP1* in

TBK1-mediated signaling in migration of T cells has never been investigated, we could not exclude that an alteration of the expression of this gene could also influence this mechanism. Moreover, *TBKBP1* was recently found to be an important regulator of development and survival of natural killer (NK) cells [37]. The role of NK cells in MS is still debated, however there is increasing evidence that certain subtypes of NK cells are able to modulate other immune cells and could be involved in autoimmune processes [38, 39]. Supporting these hypotheses, *TBKBP1* is highly expressed in activated CD4<sup>+</sup> and CD8<sup>+</sup> T cells, as well as NK cells (Online Resource 1 Fig. S6). Further tailored experiments are needed to elucidate the involvement of *TBKBP1* in MS.

Concluding, our results suggest that no novel signal associated with MS with a strong effect and specific for the continental Italian population exists. Nevertheless, we prioritized two loci, already identified to be associated with MS, that may have ethnic specificity in the Italian population, one of them pointing to the *TBKBP1* gene as a candidate to be investigated further.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00415-022-11109-8>.

**Acknowledgements** The authors thank the International MS Genetics Consortium for providing part of the genotype data. They also thank individuals with MS and their families for making this study possible and nurses at the MS center at San Raffaele Hospital for help in collection of samples. Members of the PROGEMUS consortium: Lucia Florio<sup>a</sup>, Paolo Ragonese<sup>b</sup>, Alberto Gajofatto<sup>c</sup>, Elio Scarpini<sup>d</sup>, Domenico Caputo<sup>e</sup>, Claudio Gasperini<sup>f</sup>, Franco Granella<sup>g</sup>, Paola Cavalla<sup>h</sup>, Roberto Bergamaschi<sup>i</sup>, Giovanni Ristori<sup>i</sup>, Claudio Solaro<sup>k</sup>, Filippo Martinelli Boneschi<sup>l</sup>, Francesco Passantino<sup>m</sup>, Maura Pugliatti<sup>n</sup>, Antonio Gallo<sup>o</sup>, Laura Brambilla<sup>p</sup>, Marinella Clerico<sup>q</sup>, Fioravante Capone<sup>r</sup>, Maria Trojano<sup>s</sup>, <sup>a</sup>IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy, <sup>b</sup>Policlinico Paolo Giaccone, Palermo, Italy, <sup>c</sup>Policlinico Borgo Roma, Verona, Italy, <sup>d</sup>Foundation IRCCS Cà Granda Ospedale Maggiore Policlinico, Milan, Italy, <sup>e</sup>Foundation Don Carlo Gnocchi,

Milan, Italy, <sup>f</sup>Azienda Ospedaliera S. Camillo Forlanini, Rome, Italy, <sup>g</sup>Azienda Ospedaliera-Universitaria di Parma, Parma, Italy, <sup>h</sup>Azienda Ospedaliera-Universitaria S. Giovanni Battista, Molinette (TO), Italy, <sup>i</sup>Fondazione Istituto Neurologico Nazionale Casimiro Mondino, Pavia, Italy, <sup>j</sup>Sant'Andrea Hospital, Rome, Italy, <sup>k</sup>C.R.R.F. “Mons. Luigi Novarese”, Moncrivello (VC), Italy, <sup>l</sup>IRCCS Policlinico San Donato Milanese (MI), Italy, <sup>m</sup>Ospedale SS Antonio e Biagio, Alessandria, Italy, <sup>n</sup>UO Clinica Neurologica, University of Ferrara, Ferrara, Italy, <sup>o</sup>I Division of Neurology, II university of Naples, Naples, Italy, <sup>p</sup>Fondazione IRCCS Istituto Neurologico “Carlo Besta”, Milan, Italy, <sup>q</sup>Azienda Ospedaliera-Universitaria San Luigi Gonzaga, Orbassano (TO), Italy, <sup>r</sup>Policlinico Universitario Campus Biomedico, Rome, Italy, <sup>s</sup>University of Bari “Aldo Moro”, Bari, Italy.

**Author contributions** Study was conceived and designed by FMB, SDA, FE, MS, NB and FC. Material preparation, data collection and analysis were performed by MS, NB, FC, SA, SL, ES, EM, RB, TR, EM, MZ, DV, PC, SS, LF, AZ, CB, DC, VM, GC, ML, MF, NAP, PLDeJ, GDeB. The first draft of the manuscript was written by MS and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open access funding provided by Università degli Studi di Milano within the CRUI-CARE Agreement. This work was supported by the Italian Foundation of Multiple Sclerosis [grant numbers: 2011/R/14, 2013/R/13, 2015/R/10, 2019/R-Multi/033, 2019/S/4]; the Italian Ministry of Health [grant number RF-2016-02361294]; the AGING Project for Department of Excellence at the Department of Translational Medicine (DIMET) at Università del Piemonte Orientale (Novara, Italy); and HYPERGENES [grant number HEALTH-F4-2007-201550]; Miriam Zuccalà is supported by Consorzio Interuniversitario di Biotecnologie (CIB) and AGING Project; Melissa Sorosina and Nadia Barizzone are partially supported by the Horizon 2020 European Grant [MultipleMS project]; Nikolaos A. Patsopoulos is supported by the National Multiple Sclerosis Society [grant numbers RG-1707-28657, JF-1808-32223 Harry Weaver Award].

## Declarations

**Conflict of interest** M. Sorosina, N. Barizzone, F. Clarelli, S. Anand, S. Lupoli, E. Salvi, E. Mangano, R. Bordoni, T. Roostaei, E. Mascia, M. Zuccalà, D. Vecchio, S. Santoro, L. Ferrè, M. Pozzato, C. Barlassina, D. Cusi, A. Patsopoulos, P.L. De Jager, G. De Bellis, S. D'Alfonso, F. Martinelli Boneschi have no competing interests to declare that are relevant to the content of this article. P. Cavalla received honoraria for speaking and/or for consultancy and support for participation to scientific congresses from Almirall, Biogen, Merck-Serono, Novartis, Sanofi-Genzyme, Roche and Teva. V. Martinelli received honoraria for speaking and/or for consultancy and support for travel expenses and participation in Congresses from Almirall, Biogen, Merck-Serono, Novartis, Genzyme and Teva. G. Comi has received consulting and speaking fees from Novartis, Sanofi Genzyme, Genzyme Corporation, Merck KGaA, Merck Serono SpA, Celgene Group, F. Hoffman-La Roche, Almirall SpA, Janssen-Cilag GmbH, Janssen Global Services LLC, Magenta Therapeutics Inc. Almirall and Merck Serono SpA outside the submitted work. Dr. Leone reports grants from Italian Ministry of Health (RF-2016-02361294 grant), grants from Italian Ministry of Health- Ricerca corrente 2017—RC1701NE08, during the conduct of the study. M. Filippi received compensation for consulting services and/or speaking activities from Almirall, Alexion, Bayer, Biogen, Celgene, Eli Lilly, Genzyme, Merck-Serono, Novartis, Roche, Sanofi, Takeda, and Teva Pharmaceutical Industries and receives research support from Biogen Idec, Merck-Serono, Novartis, Roche, Teva Pharma-

ceutical Industries. F. Esposito received consulting and speaking fees from Novartis, Sanofi Genzyme.

**Ethics approval** This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by each study site from the local ethics committee or institutional research board. Written informed consent was obtained for all individuals at every site by the principal investigator.


**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Compston A, Coles A (2008) Multiple sclerosis. *Lancet* 372:1502–1517
2. Sawcer S, International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2 et al (2011) Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476:214–219
3. Beecham AH, Patsopoulos NA, International Multiple Sclerosis Genetics Consortium (IMSGC) et al (2013) Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat Genet* 45:1353–1360. <https://doi.org/10.1038/ng.2770>
4. International Multiple Sclerosis Genetics Consortium (IMSGC) (2019) Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* 365:eaav7188. <https://doi.org/10.1126/science.aav7188>
5. Gregory AP, Dendrou CA, Attfield KE et al (2012) TNF receptor 1 genetic risk mirrors outcome of anti-TNF therapy in multiple sclerosis. *Nature* 488:508–511. <https://doi.org/10.1038/nature11307>
6. Steri M, Orrù V, Idda ML et al (2017) Overexpression of the cytokine BAFF and autoimmunity risk. *N Engl J Med* 376:1615–1626. <https://doi.org/10.1056/NEJMoa1610528>
7. Salvi E, Kutalik Z, Glorioso N et al (2012) Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. *Hypertension* 59:248–255. <https://doi.org/10.1161/HYPERTENSIONAHA.111.181990>
8. De Jager PL, Chibnik LB, Cui J et al (2009) Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: a weighted genetic risk score. *Lancet Neurol* 8:1111–1119. [https://doi.org/10.1016/S1474-4422\(09\)70275-3](https://doi.org/10.1016/S1474-4422(09)70275-3)
9. Sorosina M, Brambilla P, Clarelli F et al (2014) Genetic burden of common variants in progressive and bout-onset multiple sclerosis. *Mult Scler J* 20:802–811. <https://doi.org/10.1177/1352458513512707>
10. Sorosina M, Esposito F, Guaschino C et al (2015) Inverse correlation of genetic risk score with age at onset in bout-onset and

- progressiveonset multiple sclerosis. *Mult Scler J* 21:1463–1467. <https://doi.org/10.1177/1352458514561910>
11. Barizzzone N, Zara I, Sorosina M et al (2015) The burden of multiple sclerosis variants in continental Italians and Sardinians. *Mult Scler* 21:1385–1395. <https://doi.org/10.1177/1352458515596599>
  12. Anand S, Mangano E, Barizzzone N et al (2016) Next generation sequencing of pooled samples: guideline for variants' filtering. *Sci Rep* 6:33735. <https://doi.org/10.1038/srep33735>
  13. Patsopoulos NA, Genetics Working Group, Steering Committees of Studies Evaluating IFN $\beta$ -1b and a CCR1-Antagonist et al (2011) Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. *Ann Neurol* 70:897–912. <https://doi.org/10.1002/ana.22609>
  14. Polman CH, Reingold SC, Banwell B et al (2011) Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. *Ann Neurol* 69:292–302. <https://doi.org/10.1002/ana.22366>
  15. GTEx Consortium (2013) The genotype-tissue expression (GTEx) project. *Nat Genet* 45:580–585. <https://doi.org/10.1038/ng.2653>
  16. Gauthier SA, Glanz BI, Mandel M, Weiner HL (2006) A model for the comprehensive investigation of a chronic autoimmune disease: the multiple sclerosis CLIMB study. *Autoimmun Rev* 5:532–536
  17. Roostaei T, Klein H-U, Felsky D, et al (2020) Impact of genetic susceptibility to multiple sclerosis on the T cell epigenome: proximal and distal effects. *bioRxiv* 2020.07.11.198721. <https://doi.org/10.1101/2020.07.11.198721>
  18. Delaneau O, Ongen H, Brown AA et al (2017) A complete tool set for molecular QTL discovery and analysis. *Nat Commun* 8:15452. <https://doi.org/10.1038/ncomms15452>
  19. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W et al (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–329. <https://doi.org/10.1038/nature14248>
  20. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489:57–74. <https://doi.org/10.1038/nature11247>
  21. Schmiedel BJ, Singh D, Madrigal A et al (2018) Impact of genetic polymorphisms on human immune cell gene expression. *Cell* 175:1701–1715.e16. <https://doi.org/10.1016/j.cell.2018.10.022>
  22. Andlauer TFM, Buck D, Antony G et al (2016) Novel multiple sclerosis susceptibility loci implicated in epigenetic regulation. *Sci Adv* 2:e1501678. <https://doi.org/10.1126/sciadv.1501678>
  23. Greene CS, Penrod NM, Williams SM, Moore JH (2009) Failure to replicate a genetic association may provide important clues about genetic architecture. *PLoS ONE* 4:e5639. <https://doi.org/10.1371/journal.pone.0005639>
  24. International Multiple Sclerosis Genetics Consortium (2019) A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis. *Nat Commun* 10:2236. <https://doi.org/10.1038/s41467-019-09773-y>
  25. Parnell GP, Gatt PN, Krupa M et al (2014) The autoimmune disease-associated transcription factors EOMES and TBX21 are dysregulated in multiple sclerosis and define a molecular subtype of disease. *Clin Immunol* 151:16–24. <https://doi.org/10.1016/j.clim.2014.01.003>
  26. Disanto G, Sandve GK, Berlanga-Taylor AJ et al (2012) Genomic regions associated with multiple sclerosis are active in B cells. *PLoS ONE* 7:e32281. <https://doi.org/10.1371/journal.pone.0032281>
  27. Maurano MT, Humbert R, Rynes E et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337:1190–1195. <https://doi.org/10.1126/science.1222794>
  28. Corradin O, Saiakhova A, Akhtar-Zaidi B et al (2014) Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res* 24:1–13. <https://doi.org/10.1101/gr.164079.113>
  29. Farh KKH, Marson A, Zhu J et al (2015) Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518:337–343. <https://doi.org/10.1038/nature13835>
  30. Cotsapas C, Mitrovic M (2018) Genome-wide association studies of multiple sclerosis. *Clin Trans Immunol* 7:e1018. <https://doi.org/10.1002/cti2.1018>
  31. Evans DM, Spencer CCA, Pointon JJ et al (2011) Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. *Nat Genet* 43:761–767. <https://doi.org/10.1038/ng.873>
  32. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet* 14:204–220
  33. Schultz MD, He Y, Whitaker JW et al (2015) Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* 523:212–216. <https://doi.org/10.1038/nature14465>
  34. Yang X, Han H, DeCarvalho DD et al (2014) Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* 26:577–590. <https://doi.org/10.1016/j.ccr.2014.07.028>
  35. Ryzhakov G, Randow F (2007) SINTBAD, a novel component of innate antiviral immunity, shares a TBK1-binding domain with NAP1 and TANK. *EMBO J* 26:3180–3190. <https://doi.org/10.1038/sj.emboj.7601743>
  36. Yu J, Zhou X, Chang M et al (2015) Regulation of T-cell activation and migration by the kinase TBK1 during neuroinflammation. *Nat Commun* 6:6074. <https://doi.org/10.1038/ncomms7074>
  37. Zhu L, Xie X, Zhang L et al (2018) TBK-binding protein 1 regulates IL-15-induced autophagy and NKT cell survival. *Nat Commun* 9:2812. <https://doi.org/10.1038/s41467-018-05097-5>
  38. Gross CC, Schulte-Mecklenbeck A, Wiendl H et al (2016) Regulatory functions of natural killer cells in multiple sclerosis. *Front Immunol* 7:606. <https://doi.org/10.3389/fimmu.2016.00606>
  39. Nagy N (2017) Establishment of EBV-infected lymphoblastoid cell lines. *Methods in Molecular Biology*. Frontiers Media S.A., Humana Press, New York, pp 57–64

## Authors and Affiliations

Melissa Sorosina<sup>1</sup> · Nadia Barizzone<sup>2</sup> · Ferdinando Clarelli<sup>1</sup> · Santosh Anand<sup>3</sup> · Sara Lupoli<sup>4</sup> · Erika Salvi<sup>4,5</sup> · Eleonora Mangano<sup>6</sup> · Roberta Bordoni<sup>6</sup> · Tina Roostaei<sup>7</sup> · Elisabetta Mascia<sup>1</sup> · Miriam Zuccalà<sup>2</sup> · Domizia Vecchio<sup>8</sup> · Paola Cavalla<sup>9</sup> · Silvia Santoro<sup>1</sup> · Laura Ferrè<sup>1,10</sup> · Alen Zollo<sup>11</sup> · PROGEMUS · Cristina Barlassina<sup>4</sup> · Daniele Cusi<sup>6,12</sup> · Vittorio Martinelli<sup>10</sup> · Giancarlo Comi<sup>10</sup> · Maurizio Leone<sup>13</sup> · Massimo Filippi<sup>10,14,15,16</sup> · Nikolaos A. Patsopoulos<sup>17,18,19,20</sup> · Philip L. De Jager<sup>7</sup> · Gianluca De Bellis<sup>6</sup> · Federica Esposito<sup>1,10</sup> · Sandra D'Alfonso<sup>2</sup> · Filippo Martinelli Boneschi<sup>11,21</sup> 

<sup>1</sup> Laboratory of Human Genetics of Neurological Disorders, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

<sup>2</sup> Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Avogadro University, 28100 Novara, Italy

<sup>3</sup> Department of Informatics, Systems and Communications (DISCO), University of Milano-Bicocca, Milan, Italy

<sup>4</sup> Department of Health Sciences, University of Milan, 20139 Milan, Italy

<sup>5</sup> Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico “Carlo Besta”, 20133 Milan, Italy

<sup>6</sup> National Research Council of Italy, Institute for Biomedical Technologies, Segrate, 20090 Milan, Italy

<sup>7</sup> Center for Translational and Computational Neuroimmunology, Department of Neurology and the Taub Institute for Research On Alzheimer’s Disease and the Aging Brain, Columbia University Irving Medical Center, New York, NY 10032, USA

<sup>8</sup> MS Centre, SCU Neurology, AOU Maggiore Della Carità, Department of Translational Medicine, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), University of Eastern Piedmont Avogadro, 28100 Novara, Italy

<sup>9</sup> MS Center, Department of Neuroscience and Mental Health, City of Health and Science University Hospital of Torino, 10126 Turin, Italy

<sup>10</sup> Neurology Unit, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

<sup>11</sup> Department of Pathophysiology and Transplantation (DEPT), Dino Ferrari Centre, Neuroscience Section, University of Milan, 20122 Milan, Italy

<sup>12</sup> Bio4Dreams, Business Nursery for Life Sciences, Piazzale Principessa Clotilde 4/A, 20121 Milan, Italy

<sup>13</sup> SC Neurologia, Dipartimento Di Scienze Mediche, IRCCS Casa Sollievo Della Sofferenza, San Giovanni Rotondo, Italy

<sup>14</sup> Vita-Salute San Raffaele University, 20132 Milan, Italy

<sup>15</sup> Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

<sup>16</sup> Neurophysiology Unit, IRCCS San Raffaele Scientific Institute, IRCCS San Raffaele Scientific Institute, 20132 Milan, Italy

<sup>17</sup> Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham and Women’s Hospital, Boston, MA 02115, USA

<sup>18</sup> Division of Genetics, Department of Medicine, Harvard Medical School, Brigham and Women’s Hospital, Boston, MA 02115, USA

<sup>19</sup> Harvard Medical School, Boston, MA 02115, USA

<sup>20</sup> Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>21</sup> Neurology Unit and MS Centre, Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico, Via Francesco Sforza 35, 20122 Milan, Italy