Identification and characterization of Multiple Sclerosis associated loci in the Continental Italian population

M. Sorosina[1], N. Barizzone[2,3], F. Clarelli[1], S. Anand[4], S. Lupoli[5], F. Esposito[1,6], E. Mangano[4], R. Bordoni[4], PROGEMUS, PROGRESSO, V. Martinelli[6], G. Comi[6], M.Leone[2], D. Cusi[5], N.A. Patsopoulos[7,8,9], P. De Jager[7,8,9], G. De Bellis[4], S. D'Alfonso[2,3], F. Martinelli Boneschi[1,6]**

1) Laboratory of Human Genetics of Neurological Disorders, CNS Inflammatory Unit & INSPE, San Raffaele Scientific Institute, Milan, 2) Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), 3) Department of Health Sciences, University of Eastern Piedmont Amedeo Avogadro, Novara, 4) National Research Council of Italy/Institute of Biomedical Technology, Segrate, 5) Department of Health Sciences, University of Milan, 6) Department of Neurology and Neurorehabilitation, San Raffaele Scientific Institute, Milan, Italy, 7) Ann Romney Center for Neurologic Diseases & Program in Translational Neuropsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, 8) Harvard Medical School, Boston, 9) Program for Medical and Population Genetics, Broad Institute, Cambridge, MA, United States. * equal contribution

Introduction and methods

Multiple sclerosis (MS) is a complex disorder characterized by inflammatory and neurodegenerative processes. Until now, more than 100 susceptibility genetic loci have been identified though genome-wide association studies (GWAS) [1,2], however they were not focused on specific populations.

AIM

To identify and characterise the most associated loci with MS in the Italian population.

METHODS

- Genotyping ITA_{GWAS}: Illumina® Human660-Quad chip *[1]* + Illumina® Infinium II 1M duo BeadChips *[3]*
- Genotyping ITA_{ICHIP}: Illumina® ImmunoChip *[2]*
- Imputation ITA_{GWAS}: 1000Genome dataset (1Kg_2012-03-14_v3), 6.6M SNPs with Mac MiniMac
- Genotyping of Replication cohort: pre-designed or custom TaqMan assays (Openarray technology), MS replication chip
- Sequencing: target regions were captured using Agilent SureSelect target enrichment kit from 84 pools (12 individuals/pool). Paired-end multiplexed sequencing was performed on the Illumina GaIIx platform. Average depth: 351.9x/pool. Subjects were chosen based on the genetic risk burden (based on the known MS-associated SNPs [4]), comparing high-risk MS patients and low-risk matched healthy controls (HC).
- eQTL analyses: Braineac, GTEx Portal and SNPExpress

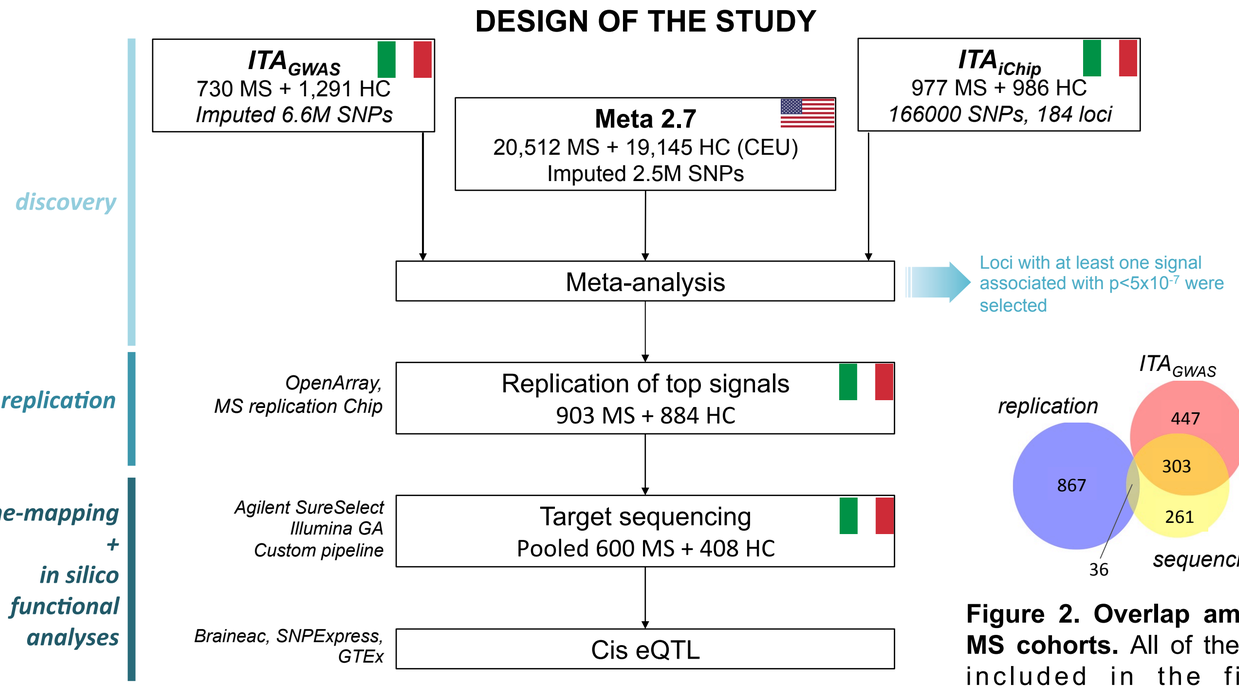


Figure 1. Design of the study.

Figure 2. Overlap among MS cohorts. All of the HC included in the fine-mapping phase belong to ITA_{iChip}.

Results

In the initial design we tried to identify high-impact signals present in the Italian population by analyzing two independent cohorts of Italian subjects. Signals didn't show a full replication suggesting that, although the population is homogeneous, in Italy there is no a high-impact signal. Thus, to bust the statistical power, the top associated signals (p<0.005) in the two Italian cohorts were meta-analysed with an imputation-based cohort from USA, following two strategies:

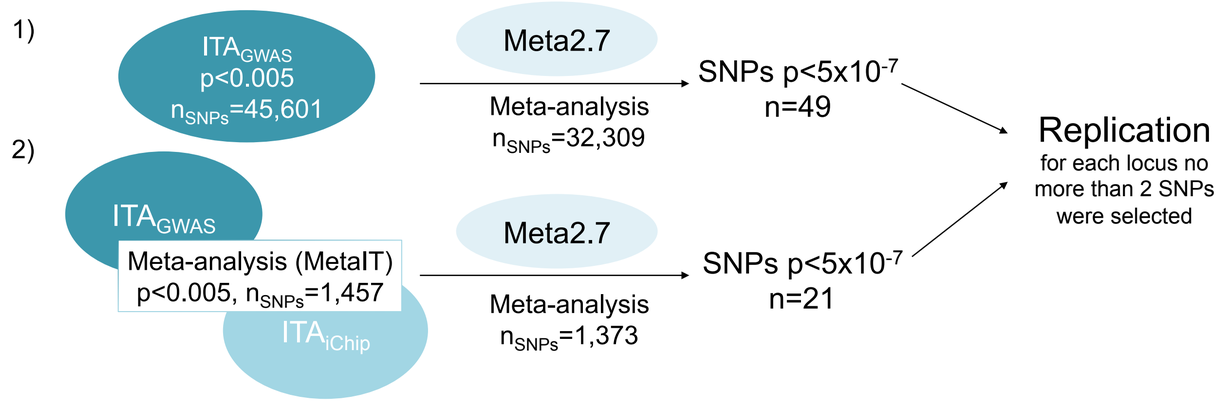


Figure 3. Discovery phase and selection of SNPs to test in the replication phase

1) ITA_{GWAS}-Meta2.7			*Discovery*		Replication		Meta-analysis			Meta-analysis Only Italians		
CHR	**SNP**	**A1**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**I**	**OR**	**P**	**I**
3	***rs338603***	C	***1.20***	***4.40E-08***	***1.27***	***8.01E-04***	***1.21***	***1.77E-10***	***0***	***1.29***	***3.73E-07***	***0***
17	***rs8070463***	T	***1.20***	***1.30E-09***	***1.24***	***3.43E-03***	***1.21***	***1.79E-11***	***0***	***1.25***	***1.72E-05***	***0***
3	rs438613	C	1.20	4.70E-08	1.23	5.74E-03	1.20	9.68E-10	0	1.27	2.67E-06	0
16	rs741174	T	0.86	1.95E-07	0.82	7.19E-03	0.85	5.48E-09	0	0.82	7.59E-05	0
16	rs741175	C	0.86	1.95E-07	0.84	0.015	0.86	9.70E-09	0	0.83	1.80E-04	0
17	rs4267364	G	0.85	1.42E-07	0.84	0.024	0.85	1.07E-08	0	0.82	2.37E-04	0
20	rs1883832	T	1.17	2.85E-07	1.16	0.051	1.17	4.02E-08	0	1.20	8.29E-04	0
20	rs6065926	A	1.17	2.01E-07	1.16	0.059	1.17	3.18E-08	0	1.21	4.67E-04	0
6	rs651971	A	0.80	2.46E-11	0.89	0.063*	0.82	1.18E-11	50.6	0.65	8.08E-19	98.23
16	rs2729590	G	0.70	1.90E-08	0.83	0.080	0.73	1.16E-08	52.8	0.71	2.56E-06	75.85
16	rs56038902	T	1.22	2.26E-07	1.08	0.346	1.19	3.61E-07	43.9	1.18	4.28E-03	60.45
16	rs55857387	T	1.22	1.14E-07	1.07	0.404	1.19	2.43E-07	53.6	1.19	2.86E-03	71.44
5	rs529279	T	0.87	4.42E-07	0.97	0.722	0.88	1.32E-06	55.2	0.84	6.59E-04	86.07
16	rs2729589	C	0.75	5.82E-08	0.89	0.726	0.75	6.37E-08	0	0.69	4.46E-05	0

2) MetaIT-Meta2.7			*Discovery*		Replication		Meta-analysis			Meta-analysis Only Italians		
CHR	**SNP**	**A1**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**I**	**OR**	**P**	**I**
3	***rs669607***	***C***	***1.18***	***6.88E-09***	***1.23***	***5.02E-03***	***1.18***	***1.43E-10***	***0***	***1.23***	***6.72E-07***	***0***
19	rs11878602	C	1.16	4.18E-07	1.22	2.40E-02	1.16	3.55E-08	0	1.20	1.10E-04	25.09
19	rs1870071	C	1.16	3.76E-07	1.22	2.63E-02	1.16	3.40E-08	0	1.21	9.88E-05	31.6
20	rs1883832	T	1.16	1.09E-07	1.16	0.051	1.16	1.55E-08	0	1.17	3.14E-04	0
20	rs6065926	A	1.16	5.99E-08	1.16	0.06	1.16	9.44E-09	0	1.18	1.45E-04	0
17	rs11079784	C	1.17	1.91E-09	1.12	0.13	1.16	7.02E-10	0	1.16	3.98E-04	0
17	rs11870935	A	1.16	2.69E-08	1.10	0.18	1.15	1.27E-08	0	1.14	9.45E-04	0
10	rs11256593	T	1.16	3.12E-09	1.06	0.42	1.15	4.04E-09	13.8	1.12	4.71E-03	0

Table 1. Association results across the discovery and replication cohorts. For each SNPs are shown: position (hg19), association in the discovery cohort, association in the replication cohort, association after meta-analysis of the two cohorts, association after meta-analysis taking account of only Italian cohorts. The SNPs are listed based on the p-value of association in the replication. SNPs that passed the Bonferroni correction threshold ($3.6x10^{-3}$ and $6.25x10^{-3}$ respectively for the analysis 1 and 2) in the replication phase are highlighted. * do not pass the HWE test p<0.0001.

We sequenced the entire LD blocks containing the top SNPs selected in the replication phase and performed cis-eQTL analyses on the significantly associated signals in blood- and brain-related tissues:

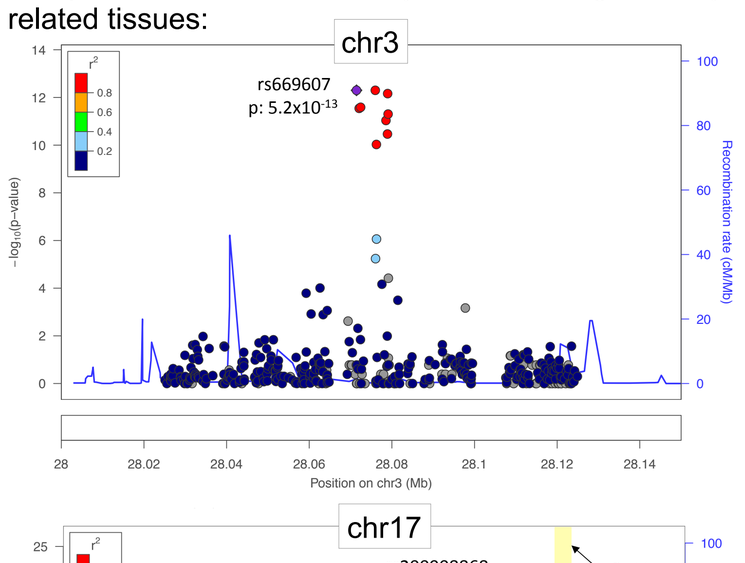
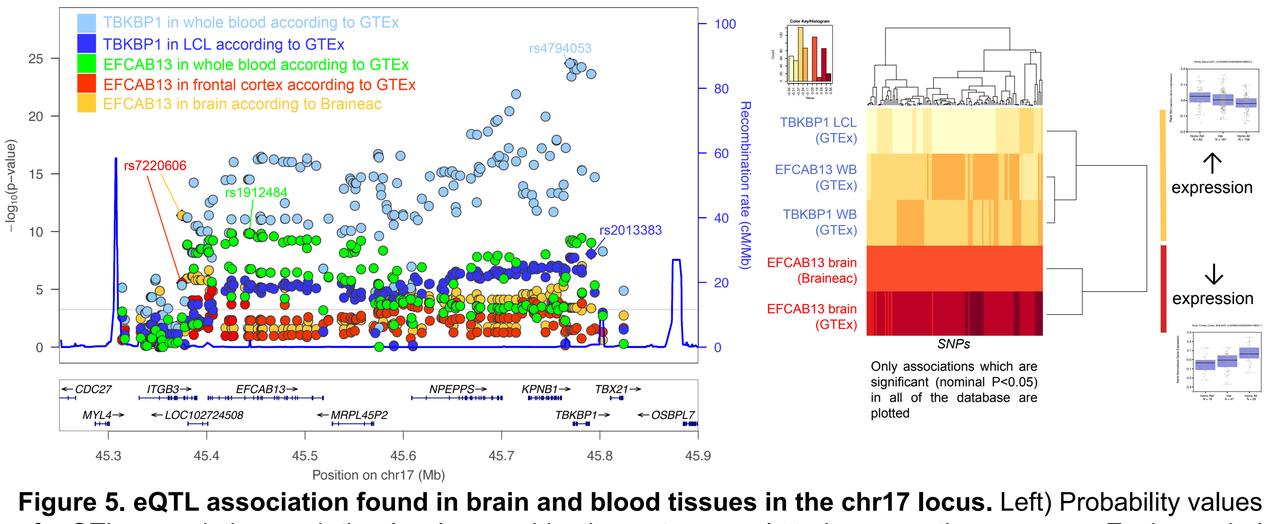


Figure 4. Association results in the sequenced cohort. The two figures show the regional association plots of the two sequenced loci. Probability values and the local recombination rates are plotted across the genome. Each symbol represents one SNP found in each locus. The most associated SNP with disease is marked by a purple symbol and the p-value is indicated. The color of each symbol is defined by the extent of linkage disequilibrium with the top SNP derived from 1000Genomes project (Nov 2014) including only European ancestry subjects. Physical positions are based on hg19 of the human genome. Top) 12 SNPs show an association ($p<5.87x10^{-5}$) in locus on chr 3, while Bottom) 203 signals were found on chr 17 ($p<1.57x10^{-5}$).



Only associations which are significant (nominal P<0.05) in all of the database are plotted

Figure 5. eQTL association found in brain and blood tissues in the chr17 locus. Left) Probability values of eQTL association and the local recombination rate are plotted across the genome. Each symbol represents one SNP tested and is colored according to the reference database and the gene. The top associated SNP for each analysis is indicated. The horizontal grey line represents the Bonferroni threshold of significance ($2.5x10^{-4}$). Right) Clustering of the eQTL effects of each SNP of the locus (p eQTL<0.05) is shown. The direction of the effect was consistent across all of the databases: EFCAB13 in brain is regulated in an opposite way compared to TBKBP1 and EFCAB itself in blood tissues. Representative plots for eQTL association of rs7220606 with the expression of EFCAB13 in whole blood and frontal cortex according to GTEx Portal database are shown.

Conclusions

- This is the larger association study performed in MS field on the continental Italian population, with ~3000 MS and ~3000 HC included.
- Although we didn't identify a novel signal of association with a strong impact, we confirmed also in Italians the association of loci previously identified by IMSGC on chr 3 and chr 17.
- Using the sequencing approach we were able to fine-map the loci, although a conditional analysis needs to be performed to identify the real signals of association.
- A strong tissue-depending eQTL association was found in locus on chr 17 with *EFCAB13* and TBKBP1 that need further investigations. Conditional eQTL analyses are ongoing to better characterize the origin of the signal.
- The locus on chr 3 is intergenic (EOMES: 300kb, CMC1: 200kb) and did not show a clear function.

All these data are confidential. You are bound not to communicate or disclose these information and results to any third party.

References

Bioinformatics tools

http://www.braineac.org
http://www.gtexportal.org
http://igm.cumc.columbia.edu/SNPExpress
http://www.hiv.lanl.gov

Literature

[1] IMSGC Nature 2011
[2] IMSGC Nature Genetics 2013
[3] Salvi E, et al. Hypertension 2012
[4] Sorosina M , et al. Mult Scler 2015

Disclosures

M Sorosina, N. Barizzone, F. Clarelli, S. Anand, S. Lupoli, E. Mangano, R. Bordoni, M. Leone, D. Cusi, G. De Bellis, S. D'Alfonso: nothing to disclose; F. Esposito: received honoraria from TEVA and Merck; V. Martinelli has received honoraria for consulting and speaking activities from Biogen-Idec, Merck, Bayer, TEVA, Novartis and Genzyme; G. Comi has received compensation for consulting services with the following companies: Novartis, Teva, Sanofi, Genzyme, Merck, Biogen, Excemed, Roche, Almirall, Chugai, Receptos, Forward Pharma and compensation for speaking activities from Novartis, Teva, Sanofi, Genzyme, Merk, Biogen, Excemed, Roche. N. Patsopoulos: consultancy, advisory boards, Merck. P. De Jager has received research support from Biogen Idec and Genzyme; F. Martinelli Boneschi: has received personal compensation for activities with Teva CNS as a speaker and/or an advisor.