Published in final edited form as:
Nat Genet. 2010 June ; 42(6): 495–497. doi:10.1038/ng.584.

Variants within the immunoregulatory CBLB gene are associated with Multiple Sclerosis

Serena Sanna[1,*], **Maristella Pitzalis**[2,*], **Magdalena Zoledziewska**[2,*], **Ilenia Zara**[3], **Carlo Sidore**[4,1], **Raffaele Murru**[5], **Michael B. Whalen**[4], **Fabio Busonero**[1], **Andrea Maschio**[1], **Gianna Costa**[5], **Maria Cristina Melis**[5], **Francesca Deidda**[2], **Fausto Poddie**[2], **Laura Morelli**[2], **Gabriele Farina**[6], **Yun Li**[7,8], **Mariano Dei**[1], **Sandra Lai**[1], **Antonella Mulas**[1], **Gianmauro Cuccuru**[1], **Eleonora Porcu**[1], **Liming Liang**[7,9], **Patrizia Zavattari**[10], **Loredana Moi**[5], **Elisa Deriu**[2], **M. Francesca Urru**[4], **Michele Bajorek**[11], **Maria Anna Satta**[12], **Eleonora Cocco**[5], **Paola Ferrigno**[13], **Stefano Sotgiu**[6], **Maura Pugliatti**[6], **Sebastiano Traccis**[14], **Andrea Angius**[4], **Maurizio Melis**[13], **Giulio Rosati**[6], **Gonçalo R. Abecasis**[7], **Manuela Uda**[1], **Maria Giovanna Marrosu**[5], **David Schlessinger**[15], and **Francesco Cucca.**[1,2]

[1]Istituto di Neurogenetica e Neurofarmacologia, CNR, Monserrato, 09042 (CA), Italy

[2]Dipartimento di Scienze Biomediche, Università di Sassari, 07100 SS, Italy

[3]CRS4, Laboratorio di Bioinformatica, Parco tecnologico della Sardegna, 09010 Pula (CA), Italy

[4]CRS4, Laboratorio di Genomica, Parco tecnologico della Sardegna, 09010 Pula (CA), Italy

[5]Centro Sclerosi Multipla, Dipartimento di Scienze Neurologiche e Cardiovascolari, Università di Cagliari, 09124 Cagliari, Italy

[6]Istituto di Neurologia Clinica, Università di Sassari, 07100 SS, Italy

[7]Department of Biostatistics, Center for Statistical Genetics, Ann Arbor 48109, University of Michigan, MI, USA

To whom correspondence should be addressed: Francesco Cucca, Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, c/o Cittadella Universitaria di Monserrato SS 554 Bivio per Sestu, 09042 Monserrato (CA), Italy, Tel: (+39) 070 675 4653, fcucca@uniss.it.

*These authors contributed equally to this work

AUTHOR CONTIBUTIONS

F.C., D.S., S.San. and M.G.M. designed the study;
F.C. and D.S. obtained fundings;
M.P., M.Z., R.M., F.D., L.M. and F.C. recruited cases and controls;
G.F., E.C., P.F., S.Sot. and S.T. collected blood samples and clinical data from MS patients, M.G.M, G.R. and M.M. supervised collection of samples and clinical data from patients;
M.B. and M.A..S. collected donors blood samples;
MF.P., L.M., F.D. and M.C.M. extracted DNA from peripheral blood;
M.P., M.Z., R.M., G.D., F.D. and L.M. prepared DNA samples;
M.Z., R.M., F.B., A.Mas., F.D., M.D., S.L., A.M., P.Z., L.M., E.D., M.F.U. performed gene-chip genotyping;
A.A. and M.U. organized facilities for micro-array genotyping;
S.San., I.Z., C.S., and E.P. performed imputation and GWAS analyses;
S.San., M.P. and M.Z., selected variants for follow up;
F.C. S.San., M.P., M.Z., I.Z., C.S. and M.B.W. interpreted results;
Y.L. and G.A. developed and implemented methods for imputation with 1000 Genomes haplotypes;
L.L. and G.A. performed gene-expression analyses;
M.P., M.Z. and R.M. performed replication and extension phases with TaqMan;
F.C. and S.San. wrote the first draft of the manuscript;
S.San., M.B.W., D.S., G.A. and F.C. provided important contribution to manuscript revision.
All authors read and approved the final manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

[8]Department of Genetics, Department of Biostatistics, University of North Carolina, Chapel Hill 27599, NC, USA

[9]Department of Epidemiology and Department of Biostatistics, Harvard School of Public Health, Boston 02115, Massachusetts, USA

[10]Dipartimento di Scienze Biomediche e biotecnologie, Università di Cagliari, 09126 CA, Italy

[11]AOBrotzu, Centro Trasfusionale, 09134 Cagliari, Italy

[12]ASL n°1, Centro Immunotrasfusionale, 07100 SS, Italy

[13]AOBrotzu, S.C. Neurologia, 09134 Cagliari, Italy

[14]Presidio Ospedaliero, Divisione Neurologia, 07014 Ozieri SS, Italy

[15]Laboratory of Genetics, National Institute on Aging, Baltimore, MD, USA

Abstract

A genome wide association scan of ~6.6 million genotyped or imputed variants in 882 Sardinian Multiple Sclerosis (MS) cases and 872 controls suggested association of *CBLB* gene variants with disease, which was confirmed in 1,775 cases and 2,005 controls (overall $P = 1.60 \times 10^{-10}$). *CBLB* encodes a negative regulator of adaptive immune responses and mice lacking the orthologue are prone to experimental autoimmune encephalomyelitis, the animal model of MS.

Multiple sclerosis (MS) is a multifactorial neuroinflammatory and autoimmune disorder. A primary cause of disability in young adults, it results from interactions between unknown environmental factors and alleles of many susceptibility loci across the genome. Recent investigations of the genetics of MS have resulted in important advances, driven largely by completion of the first genome-wide association scans (GWAS)[1-3]. Thus far, the major GWAS findings have come from analysis of northern European and derived populations where MS is particularly common. However, MS is notably and given the low incidence in neighboring populations, anomalously common in the founder, isolated population of Sardinia; the prevalence of 0.16% is among the highest world-wide[4], despite the rarity of the major HLA susceptibility haplotype HLA-DRB1*1501-DQB1*0602 (freq=0.014 vs ~ 0.25 in Europeans)[5]. Therefore, it is of particular interest to dissect the genetic bases of MS in Sardinia. Here we report analyses that show a strong association between MS and a specific immunoregulatory gene.

We first analyzed 882 MS cases and 872 controls genotyped with the Affymetrix 6.0 chip (Supplementary Methods). To increase the spectrum of variants tested for association, we used imputation methods[6] and genotypes at 555,335 autosomal SNPs that passed strict quality checks (see Supplementary Table 1) to perform three rounds of imputation using the HapMap phase II CEU, HapMap phase III CEU and TSI and 1000 Genomes Project samples as references (Supplementary Methods, Supplementary Table 1). While imputation using HapMap II as reference is now a standard analysis in most GWAS[6], using HapMap phase III and 1000 Genomes Project data allowed us to examine evidence for association at additional sites. Overall we were able to test for MS association using 6,607,266 markers that had been either directly genotyped or successfully imputed. Population stratification was evaluated and corrected for using principal component analyses, though we observed no large scale substructure in our samples; the genomic control parameter was < 1.057 in all three imputed data sets (Supplementary Methods).

Several signals in the HLA region satisfied the genome-wide significance threshold of 5×10^{-8}. The top associated variant, rs2040406 ($P = 1.45 \times 10^{-20}$) is located ~45 Kb from the

DRB1 locus and showed a strong correlation (r^2= 0.85) with the common DRB1*0301-DQB1*0201 haplotype in a subset of 423 MS cases who were also fully typed for HLA-DRB1 and DQB1 loci. This haplotype has been previously found to be associated with MS in Sardinia[5]. Interestingly, the best tag for the canonical HLA-DRB1*1501 allele, rs3135388 (see Supplementary Methods), did not show evidence for association either in the GWAS or after conditioning for the top SNP, rs2040406 (*P*=0.69 and *P*=0.045, respectively). Instead, the conditional analysis showed as most associated marker in the region, SNP rs9267955 (*P*=4.5 ×10^{-07}, allele A freq. cases= 0.05, freq. controls =0.03), which partially correlates (r^2=0.41, Supplementary Methods) with the HLA-DRB1*1501. Furthermore, no association was observed at SNP rs2523393, the best proxy in Europeans[2] for the HLA-B44 allele (*P*=0.930 in the GWAS, and *P*= 0.177 in the conditional analysis), recently described as associated with MS in other populations. Altogether, these data reflect the fact that a different spectrum of HLA variants associated with MS may exists in this population and that further work is required to fine map these effects. Previously described markers at known non-HLA loci were also confirmed, although they did not show genome-wide significant pvalues (see Supplementary Table 2).

We then ranked non-HLA SNPs based on their level of significance, evidence for association at nearby SNPs, proximity to functional candidate genes and *P*-values in previously published GWA scans, prioritizing variants that were not typed or imputed in previous GWAS or were located in loci with inconclusive evidence of disease association (Supplementary Methods). We selected nine SNPs to be validated with an independent genotyping method (Supplementary Methods and Supplementary Table 3). *De novo* genotyping with Taqman yielded an average concordance rate of 97.5% between inferred and directly typed genotypes. All nine of the chosen variants were then assessed in an additional 1,264 MS cases and 1,305 unrelated controls from the sample population; the controls included 403 Affected Family Based pseudo Controls (AFBAC) from 403 MS trios (Supplementary Methods). Only one SNP showed evidence of association at $P < 0.05$ in the extension sample set and was then genotyped in further 511 MS cases and 700 controls (including 128 AFBAC, Table 1 and Supplementary Table 3). In particular, SNP rs9657904 (T>C), located in intron 1 of the *CBLB* gene on chr3q13.11, was genotyped with the Affymetrix chip (call rate = 93.5%), and showed in the first stage of the GWAS an initial *P* of 7.95 × 10^{-5} that increased to $P = 3.19 \times 10^{-6}$ by filling in undetermined genotypes by imputation using HapMap III. It was then fully validated ($P = 1.20 \times 10^{-6}$) by direct Taqman genotyping and confirmed in an additional 1,775 MS cases and 2,005 controls ($P = 9.34 \times 10^{-6}$, see Table 1 and Supplementary Table 3) Jointly analyzing all the available data from the GWAS together with the follow-up data set (2,657 cases and 2,877 controls) we observed convincing evidence for association: $P = 1.60 \times 10^{-10}$ (OR = 1.40, 95%CI = 1.27 - 1.57 for the common allele T, and OR = 2.5, 95% CI = 1.71 - 3.52, for the homozygous TT genotype when compared to the baseline risk conferred by the CC genotype, with no significant evidence for departure from the multiplicative model; Table 1 and Figure 1). Likewise, no significant interaction (P =0.31) was detected in a case-only analysis with the DRB1*0301-DQB1*0201 haplotype, using rs2040406 genotypes as a proxy.

CBLB was the only gene in the associated region (Figure 1). Only one other variant, rs12487066, 325Kb upstream of the *CBLB* gene and independent of rs9657904 (r^2 = 0.2), showed some nominal evidence of association in the previous GWAS[3], but it was not associated with MS (P = 0.74) in our sample.

In a conditional regression analysis performed in our GWAS data, SNP rs9657904 entirely explained the association of the region (that is, no other nearby SNPs exhibited significant evidence for association after adjusting for rs9657904). This marker was not included in HapMap II; thus, we were only able to fully analyze it after imputation with HapMap III or

1000 Genomes haplotypes as reference. Other nearby SNPs that were included in HapMap II did show strong evidence for association in all analyses. Indeed, several other markers spanning the region from the gene promoter to intron 3 showed a degree of association similar to SNP rs9657904 (rs9846534 $P = 3.93 \times 10^{-6}$, rs10511246 $P = 3.90 \times 10^{-6}$, rs6804152 $P = 3.87 \times 10^{-6}$, rs7649466 $P = 3.66 \times 10^{-6}$, rs9631436 $P = 3.72 \times 10^{-6}$) (Figure 1). These polymorphisms could not be distinguished statistically and therefore any of these, or other variants yet to be identified, could be the causal variant(s). The SNPs with highest pvalues, rs9657904 and rs7649466 (r^2 =1 with each other), both fall in the 5' region of *CBLB*, a gene characterized by at least 11 alternative isoforms. This region contains a CpG island along with DNAseI hypersensitivity sites, a combinatorial pattern of 17 histone acetylation and methylation sites, and a PolII binding site, as assessed in CD4+ T cell lines (Supplementary Figure 2). All these features indicate the presence of a regulatory core with transcriptional activity. Furthermore, SNP rs7649466 is predicted to affect an exonic splicing enhancer for SF2 (score allele G 2.81; score allele C 3.42 - Pupasuite 2.0 - "http://pupasuite.bioinfo.cipf.es/"). To evaluate in more detail the impact of rs9657904 and neighbouring SNPs on gene expression, we used 1000 Genomes haplotypes to impute rs9657904 in individuals previously included in a genome-wide association study of global gene expression[7]. In Epstein-Barr virus–transformed lymphoblastoid cells, allele T (which is associated with increased MS risk) was strongly associated ($P = 8 \times 10^{-9}$) with decreased expression of 234112_at, a probe that overlaps *CBLB* but is not part of any of its annotated isoforms. After removing non-genetic effects (see Supplementary Methods), the marker explained 11.4% of the variation in transcript levels for 234112_at. No association was observed with any of the other three *CBLB* probesets in this data set (208348_s_at, 209682_at, 227900_at) (see Supplementary Figure 3). Overall, these data suggest that rs9657904, rs7649466, or other variants in close LD with them, might affect *CBLB* splicing and/or transcription regulation, and could provide mechanistic clues for the observed disease association.

CBLB encodes a multifunctional adaptor protein that works as a RING-family E3 ubiquitin ligase to negatively regulate T cell receptor (TCR) and B cell receptor (BCR) activation[8,9]. Furthermore, this molecule is also critical for the induction of anergy in NKT cells, a specific class of T cells activated by glycolipid antigens and characterized by the cell surface expression of a single invariant TCR[10].

Mice deficient in the orthologue Cblb are highly susceptible to experimental autoimmune encephalomyelitis[11]. The gene may also have a broader role in general autoimmunity, because Cblb disruption causes lymphocytic infiltration into many organs in another mouse genetic background[12] and was also discovered to be a major susceptibility gene in different rodent models of autoimmune disease[13,14]. Furthermore, other studies have shown that Cblb deficient mice spontaneously reject a variety of cancers[15]. Hence, *CBLB* variation appears to be critical in maintaining the delicate balance between immunological activation and tolerance with anticipated opposite effects in autoimmunity and cancer.

Given the versatility of *CBLB* as a gatekeeper in adaptive immune-response[15,8], its involvement in MS might suggest that both T cells and B cells cooperate in the disease process, most likely with lowered activity leading to an exacerbated immune response in the etiology of MS. Still, unraveling exactly how *CBLB* contributes to MS risk will require additional fine mapping, expression and functional experiments. Clarifying this will highlight key features of disease pathogenesis and help to understand the overall functioning and malfunctioning of the immune system in the Central Nervous System.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by FISM (Fondazione Italiana Sclerosi Multipla) Cod. 2008/R/7 to FC, by the Italian Ministry of Scientific Research (MIUR grant 2007KXNKNP) and by NIH contract NO1-AG-1-2109 from National Institute of Aging (NIA) to the SardiNIA ("ProgeNIA") team.

We thank all the patients and controls who made this research possible, John Todd for a critical revision of the manuscript and helpful suggestions, Antonio Cao for his continuous help and support, Dan Longo, Dennis Taub, Silvia Naitza, Laura Crisponi, Mauro Congia, Chris Jones, Paolo Zanella, Roman Tirler, Lidia Leoni, Roberto Cusano, Rosanna Lampis, Valeria Orrù, Francesca Coraddu, Jamie Foster, Monica Luiu, Luca Davolio Marani, Sara Solveig Fois and Marco Arru for help and advice, the physicians Jessica Frau, Lorena Lorefice, Giancarlo Coghe, Giuseppe Fenu, Luisella Mannu, Paola Cossu, Wanda Satta, Immacolata Pirastru, Stefania Leoni, Chiara Buscarinu for collaboration in the collection of MS patients, and Wanda Garau and Francesca Virdis along with the personnel of blood donor centers of Cagliari and Sassari for their help in the collection of the controls. We are grateful for the important computing resources made available for the imputation and analysis by the CRS4 Computing Cluster in Pula (Cagliari, Italy).

References

1. ANZgene. Nat Genet. 2009; 41:824–8. [PubMed: 19525955]
2. De Jager PL, et al. Nat Genet. 2009; 41:776–82. [PubMed: 19525953]
3. Hafler DA, et al. N Engl J Med. 2007; 357:851–62. [PubMed: 17660530]
4. Pugliatti M, et al. Eur J Neurol. 2006; 13:700–22. [PubMed: 16834700]
5. Marrosu MG, et al. Hum Mol Genet. 2001; 10:2907–16. [PubMed: 11741834]
6. Li Y, Willer C, Sanna S, Abecasis G. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]
7. Dixon AL, et al. Nat Genet. 2007; 39:1202–1207. [PubMed: 17873877]
8. Schmitz ML. Sci Signal. 2009; 2:pe38. [PubMed: 19549983]
9. Qiao G, et al. J Immunol. 2007; 179:4473–9. [PubMed: 17878343]
10. Kojo S, et al. Proc Natl Acad Sci U S A. 2009; 106:17847–51. [PubMed: 19815501]
11. Chiang YJ, et al. Nature. 2000; 403:216–20. [PubMed: 10646609]
12. Bachmaier K, et al. Nature. 2000; 403:211–6. [PubMed: 10646608]
13. Yokoi N, et al. Nat Genet. 2002; 31:391–4. [PubMed: 12118252]
14. Jeon MS, et al. Immunity. 2004; 21:167–77. [PubMed: 15308098]
15. Loeser S, Penninger JM. Semin Immunol. 2007; 19:206–214. [PubMed: 17391982]

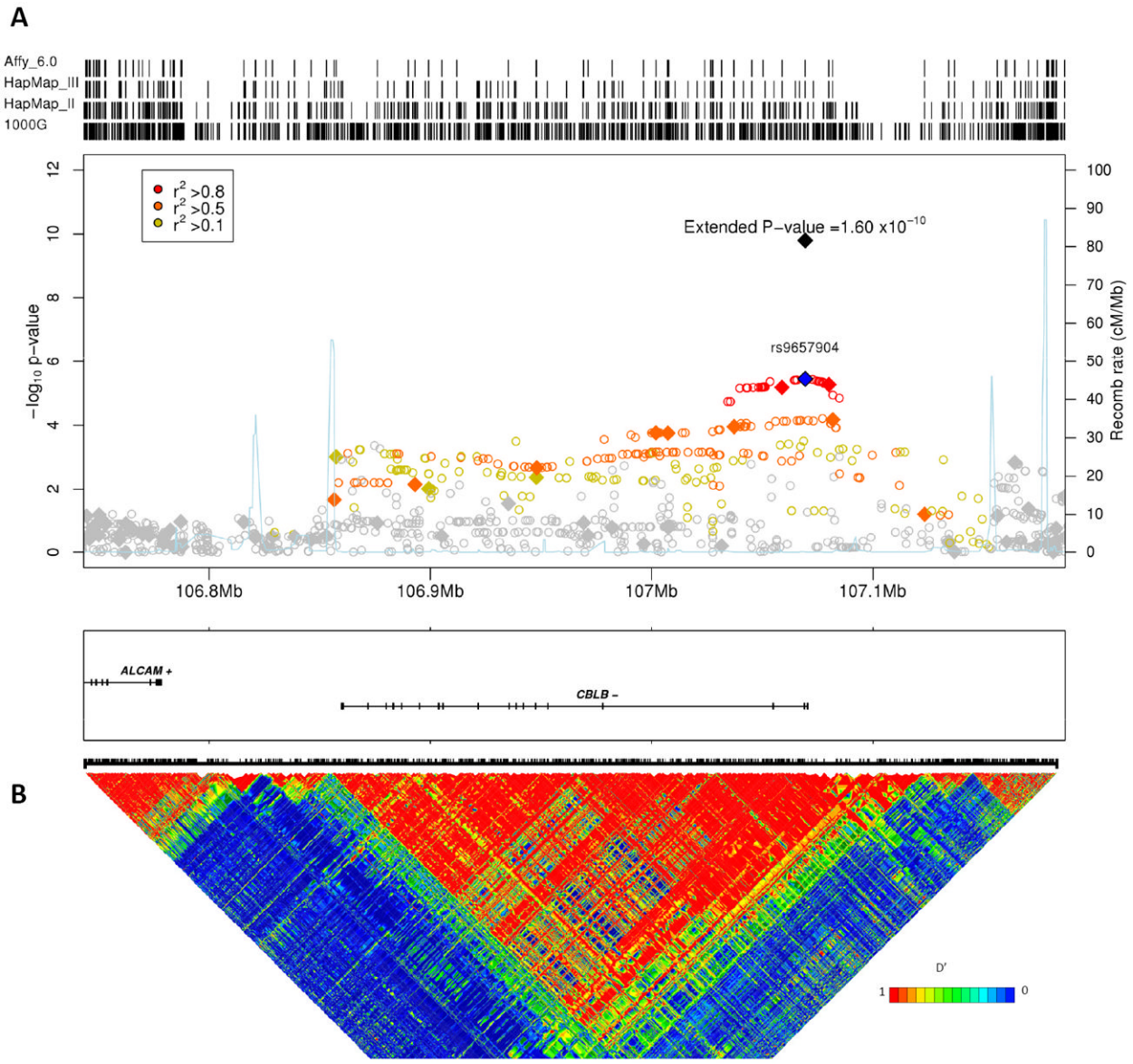

Figure 1. Association and Linkage Disequilibrium patterns at the *CBLB* locus

A) Association of genotyped and imputed SNPs (-$\log_{10}$ of the pvalue) around the *CBLB* gene. Comb diagrams indicate the location of successfully genotyped SNPs in Affymetrix 6.0 (filled diamonds) and of SNPs imputed (open circles) using haplotypes from 1000 Genomes as reference panel. The top SNP (rs9657904) is highlighted, and other SNPs are colored according to their degree of disequilibrium with this variant. The pvalue resulting from a joint analyses of samples used in the GWA scan and replication is annotated. *CBLB* transcript is indicated in the lower box, with '-' indicating transcript's direction. B) LD pattern in the Sardinian population.

Table 1

Association results for SNP rs9657904

The table summarizes results at SNP rs9657904 in the *CBLB* locus using a 1 d.f. test. The independent samples consist of 1,775 and 1,474 controls and 531 AFBAC controls. Frequency and odds ratio are given respect to allele T.

Study type	N (Cases/Controls)	Freq (Cases/Controls)	OR (95% CI)	pvalue
GWA samples	882 / 872	0.875 / 0.815	1.58 (1.31 - 1.90)	1.20×10^{-06}
Independent samples	1775 / 2005	0.867 / 0.830	1.34 (1.17 - 1.52)	9.34×10^{-06}
Joint analysis	2657 / 2877	0.870 / 0.826	1.40 (1.27 - 1.57)	1.60×10^{-10}