nature genetics

Variants within the immunoregulatory *CBLB* gene are associated with multiple sclerosis

Serena Sanna[1,18], Maristella Pitzalis[2,18], Magdalena Zoledziewska[2,18], Ilenia Zara[3], Carlo Sidore[1,4], 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[7–9], Mariano Dei[1], Sandra Lai[1], Antonella Mulas[1], Gianmauro Cuccuru[1], Eleonora Porcu[1], Liming Liang[7,10,11], Patrizia Zavattari[12], Loredana Moi[5], Elisa Deriu[2], M Francesca Urru[4], Michele Bajorek[13], Maria Anna Satta[14], Eleonora Cocco[5], Paola Ferrigno[15], Stefano Sotgiu[6], Maura Pugliatti[6], Sebastiano Traccis[16], Andrea Angius[4], Maurizio Melis[15], Giulio Rosati[6], Gonçalo R Abecasis[7], Manuela Uda[1], Maria Giovanna Marrosu[5], David Schlessinger[17] & Francesco Cucca[1,2]

A genome-wide association scan of ~6.6 million genotyped or imputed variants in 882 Sardinian individuals with multiple sclerosis (cases) and 872 controls suggested association of *CBLB* gene variants with disease, which was confirmed in 1,775 cases and 2,005 controls (rs9657904, overall $P = 1.60 \times 10^{-10}$, OR = 1.40). *CBLB* encodes a negative regulator of adaptive immune responses, and mice lacking the ortholog are prone to experimental autoimmune encephalomyelitis, the animal model of multiple sclerosis.

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 studies (GWAS)[1–3]. Thus far, the major GWAS findings have come from analyses of northern European populations and derived populations in which MS is particularly common. Although MS has a low incidence in geographically neighboring populations, it is common in the founder, isolated population of Sardinia; its prevalence of 0.16% in this population is among the highest worldwide[4], despite the rarity of the major MS-susceptibility *HLA* haplotype *HLA-DRB1**1501-*DQB1**0602 (frequency = 0.014 compared to ~0.25 in Europeans)[5]. Therefore, it is of particular interest to dissect the genetic basis of MS in Sardinia. Here we report analyses that show a strong association between MS and a specific immunoregulatory gene in Sardinians.

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 genotypes at 555,335 autosomal SNPs that passed strict quality checks (see **Supplementary Table 1**) to perform three rounds of imputation[6] using the HapMap phase II European CEU, HapMap phase III CEU and Italian TSI populations and the 1000 Genomes Project samples as references (**Supplementary Methods**). Although imputation using HapMap II as a reference is now a standard analysis in most GWAS[6], using HapMap phase III and 1000 Genomes Project data here allowed us to examine evidence for association at additional sites. Overall, we were able to test for MS association 6,607,266 markers that had been either directly genotyped or successfully imputed. Population stratification was evaluated and corrected for using principal component analysis, though we observed no large-scale substructure in our samples; the genomic control parameter was <1.057 in all three imputed datasets (**Supplementary Methods** and **Supplementary Fig. 1**).

Several signals in the *HLA* region satisfied the genome-wide significance threshold of $P = 5 \times 10^{-8}$. 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]. Notably, 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 the most associated marker in the region to be rs9267955 ($P = 4.5 \times 10^{-7}$, frequency of allele A in cases = 0.05 and frequency in controls = 0.03), which partially correlates ($r^2 = 0.41$; **Supplementary Methods**) with the *HLA-DRB1**1501 haplotype. Altogether, these data reflect the fact that a different spectrum of *HLA* variants associated with MS may exist in this population and that

[1]Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche (CNR), Monserrato, Italy. [2]Dipartimento di Scienze Biomediche, Università di Sassari, Sassari, Italy. [3]Center for Advanced Studies, Research and Development in Sardinia (CRS4), Laboratorio di Bioinformatica, Parco tecnologico della Sardegna, Pula, Italy. [4]CRS4, Laboratorio di Genomica, Parco tecnologico della Sardegna, Pula, Italy. [5]Centro Sclerosi Multipla, Dipartimento di Scienze Neurologiche e Cardiovascolari, Università di Cagliari, Cagliari, Italy. [6]Istituto di Neurologia Clinica, Università di Sassari, Sassari, Italy. [7]Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan, USA. [8]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [9]Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina, USA. [10]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [11]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [12]Dipartimento di Scienze Biomediche e biotecnologie, Università di Cagliari, Cagliari, Italy. [13]Azienda Ospedaliera Brotzu, Centro Trasfusionale, Cagliari, Italy. [14]Azienda Sanitaria Locale 1, Sassari, Italy. [15]Azienda Ospedaliera Brotzu, Divisione di Neurologia, Cagliari, Italy. [16]Presidio Ospedaliero, Divisione Neurologia, Ozieri, Italy. [17]Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. [18]These authors contributed equally to this work. Correspondence should be addressed to F.C. (francesco.cucca@inn.cnr.it).

Received 21 December 2009; accepted 13 April 2010; published online 9 May 2010; doi:10.1038/ng.584

Table 1 Association results for rs9657904

Study type	*n* (cases/controls)	Freq (cases/controls)	OR (95% CI)	*P* value
GWA samples	882/872	0.875/0.815	1.58 (1.31–1.90)	1.20×10^{-6}
Independent samples	1,775/2,005	0.867/0.830	1.34 (1.17–1.52)	9.34×10^{-6}
Joint analysis	2,657/2,877	0.870/0.826	1.40 (1.27–1.57)	1.60×10^{-10}

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

further work is required to fine map these effects. Some previously described markers at known non-*HLA* loci were also confirmed, although they did not show genome-wide significant *P* values (see **Supplementary Table 2**).

We then ranked non-*HLA* SNPs based on their level of significance, the evidence for association at nearby SNPs, their proximity to functional candidate genes and their *P* values in previously published GWA scans, prioritizing variants that were not typed or imputed in previous GWAS or that were 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 replication sample set; this SNP was then genotyped in a further 511 MS cases and 700 controls, which included 128 AFBAC individuals (**Table 1** and **Supplementary Table 3**). In particular, rs9657904 (containing a T > C alteration), located in intron 1 of *CBLB* on chromosome 3q13.11, was genotyped with the Affymetrix chip (call rate = 93.5%) and showed an initial $P = 7.95 \times 10^{-5}$ in the first stage of the GWAS that was decreased to $P = 3.19 \times 10^{-6}$ by filling in undetermined genotypes by imputation using HapMap III. This SNP was then fully validated ($P = 1.20 \times 10^{-6}$) by direct TaqMan genotyping and was confirmed in an additional 1,775 MS cases and 2,005 controls ($P = 9.34 \times 10^{-6}$; **Table 1** and **Supplementary Table 3**).

Jointly analyzing all the available data from the GWAS together with the follow-up dataset (which contained 2,657 cases and 2,877 controls), we observed convincing evidence for association, with $P = 1.60 \times 10^{-10}$ (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; **Fig. 1** and **Table 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 (**Fig. 1**). One other variant, rs12487066, 325 kb upstream of *CBLB* 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 on our GWAS data, rs9657904 entirely explained the association at the region (that is, no other nearby SNPs showed 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 the haplotypes from the HapMap III or 1000 Genomes samples as references. 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 that of rs9657904 (rs9846534 $P = 3.93 \times 10^{-6}$, rs10511246 $P = 3.90 \times 10^{-6}$, rs6804152 $P = 3.87 \times 10^{-6}$, rs7649466 $P = 3.66 \times 10^{-6}$ and rs9631436 $P = 3.72 \times 10^{-6}$; **Fig. 1**). However, these polymorphisms could not be distinguished statistically and therefore any of them, or other variants yet to be identified, could be the causal variant(s). The SNPs with the smallest *P* values, 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. As assessed in $CD4^+$ T-cell lines, this region contains a CpG island, along with DNase I hypersensitivity sites, a combinatorial pattern of 17 histone acetylation and methylation sites and a DNA polymerase II binding site (**Supplementary Fig. 2**). All these features indicate the presence of a regulatory core with transcriptional activity. Furthermore, rs7649466 is predicted to affect an exonic splicing enhancer for SF2 (allele G score = 2.81 and allele C score 3.42 from PupaSuite 2.0). Overall, these data suggest that rs9657904 and rs7649466, or other variants in close linkage disequilibrium 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 and B-cell receptor activation[7,8]. Furthermore, this molecule is also critical for the induction of anergy in NKT cells, a specific class of

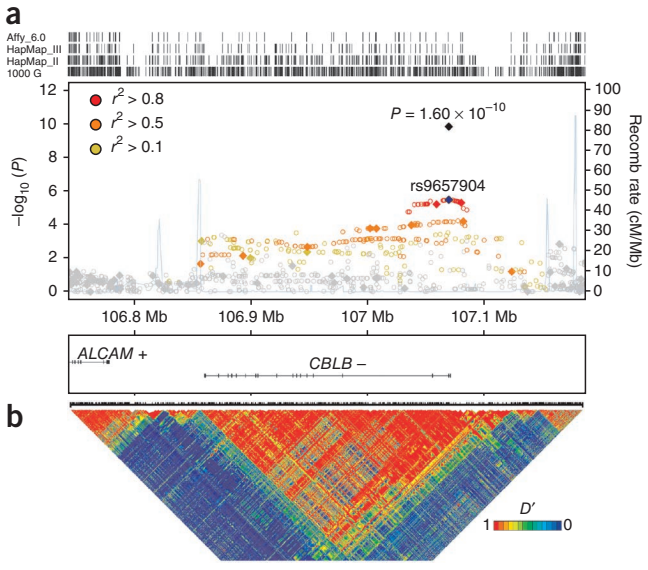

Figure 1 Association and linkage disequilibrium patterns at the *CBLB* locus. (**a**) Association of genotyped and imputed SNPs ($-\log_{10} P$) around *CBLB*. Comb diagrams indicate the location of successfully genotyped SNPs in Affymetrix 6.0 (filled diamonds) and of SNPs imputed (open circles) using haplotypes from 1000 Genomes as the reference panel. The top SNP (rs9657904) is highlighted in blue, and other SNPs are colored according to their degree of disequilibrium with this variant; for color coding, see legend at the upper left corner. Recombination rate (cM/Mb, from the CEU HapMap population) is depicted by a blue line. The *P* value resulting from a joint analysis of samples used in the GWA scan and replication is annotated and indicated with a black diamond. The *CBLB* transcript is indicated in the lower box, with '–' indicating the transcript's direction. (**b**) The linkage disequilibrium pattern in the Sardinian population.

T cells activated by glycolipid antigens and characterized by the cell-surface expression of a single invariant T-cell receptor[9].

Mice deficient in the ortholog *Cblb* are highly susceptible to experimental autoimmune encephalomyelitis[10]. *Cblb* may also have a broader role in general autoimmunity, because its disruption causes lymphocytic infiltration into many organs in another mouse genetic background[11], and *Cblb* was also discovered to be a major susceptibility gene in different rodent models of autoimmune disease[12,13]. Furthermore, other studies have shown that Cblb-deficient mice spontaneously reject a variety of cancers[14]. Hence, *CBLB* variation appears to be critical in maintaining the delicate balance between immunological activation and tolerance, which are anticipated to have opposite effects in autoimmunity and cancer.

Given the versatility of *CBLB* as a gatekeeper in the adaptive immune response[7,14], 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 the role of *CBLB* will highlight key features of disease pathogenesis and help us to understand the overall functioning and malfunctioning of the immune system in the central nervous system.

URLs. MACH software, http://www.sph.umich.edu/csg/abecasis/mach/; R project, http://www.r-project.org/; HapMap project, http://www.hapmap.org/; 1000 Genomes Project, http://www.1000genomes.org/; 1000 Genomes Haplotypes for imputation, http://www.sph.umich.edu/csg/yli/mach/download/1000G-Sanger-0908.html; PupaSuite, http://pupasuite.bioinfo.cipf.es/; Ensemble, http://www.ensemble.org/; UCSC, http://genome.ucsc.edu/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This study was supported by the Fondazione Italiana Sclerosi Multipla (FISM) Cod. 2008/R/7 to F.C., by the Italian Ministry of Scientific Research (MIUR grant 2007KXNKNP) and by US National Institutes of Health contract NO1-AG-1-2109 from National Institute of Aging (NIA) to the SardiNIA ('ProgeNIA') team. The contributions of G.R.A., Y.L. and L.L. were supported in part by NIH grants HG002651, HG005214 and MH084698.

We thank all the cases and controls who made this research possible; J. Todd for a critical revision of the manuscript and helpful suggestions; A. Cao for his continuous help and support; D. Longo, D. Taub, S. Naitza, L. Crisponi, M. Congia, C. Jones, P. Zanella, R. Tirler, L. Leoni, R. Cusano, R. Lampis, V. Orrù, F. Coraddu, J. Foster, M. Luiu, L. Davolio Marani, S. Solveig Fois and M. Arru for help and advice; J. Frau, L. Lorefice, G. Coghe, G. Fenu, L. Mannu, P. Cossu, W. Satta, I. Pirastru, S. Leoni and C. Buscarinu for collaboration in the collection of MS cases; and W. Garau and F. Virdis, along with the personnel of the 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).

AUTHOR CONTIBUTIONS

Designed the study: S. Sanna, D.S., M.G.M. and F.C. **Recruited cases and controls:** M.P., M.Z., R.M., F.D., G.F., L. Morelli, M.A.S., E.C., M.B., P.F., S. Sotgiu, S.T. and F.C. **Supervised collection of samples and clinical data from subjects:** M.M, G.R. and M.G.M. **Extracted and prepared DNA samples:** M.P., M.Z., R.M., G.C., M.C.M, F.D., F.P., L. Morelli and L. Moi **Performed gene-chip genotyping:** M.Z., R.M., F.B., A. Maschio, F.D., M.D., S.L., A. Mulas, P.Z., L. Moi, E.D., M.F.U. **Organized facilities for microarray genotyping:** A.A. and M.U. **Performed SNP calling:** I.Z. and G.C. **Performed imputation and GWAS analyses:** S. Sanna, I.Z., C.S. and E.P. **Selected variants for follow-up**: S. Sanna, M.P. and M.Z. **Interpreted results:** S. Sanna, M.P., M.Z., M.B.W. and F.C. **Developed and implemented methods for imputation with 1000 Genomes haplotypes**: Y.L., L.L. and. G.R.A. **Performed genotyping with TaqMan**: M.P., M.Z. and R.M. **Wrote the first draft of the manuscript:** S. Sanna and F.C. **Provided important contribution to manuscript revision:** S. Sanna, M.B.W., G.R.A., D.S. and F.C.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. The Australia and New Zealand Multiple Sclerosis Genetics Consortium. *Nat. Genet.* **41**, 824–828 (2009).
2. De Jager, P.L. *et al. Nat. Genet.* **41**, 776–782 (2009).
3. Hafler, D.A. *et al. N. Engl. J. Med.* **357**, 851–862 (2007).
4. Pugliatti, M. *et al. Eur. J. Neurol.* **13**, 700–722 (2006).
5. Marrosu, M.G. *et al. Hum. Mol. Genet.* **10**, 2907–2916 (2001).
6. Li, Y., Willer, C., Sanna, S. & Abecasis, G. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
7. Schmitz, M.L. *Sci. Signal.* **2**, pe38 (2009).
8. Qiao, G. *et al. J. Immunol.* **179**, 4473–4479 (2007).
9. Kojo, S. *et al. Proc. Natl. Acad. Sci. USA* **106**, 17847–17851 (2009).
10. Chiang, Y.J. *et al. Nature* **403**, 216–220 (2000).
11. Bachmaier, K. *et al. Nature* **403**, 211–216 (2000).
12. Yokoi, N. *et al. Nat. Genet.* **31**, 391–394 (2002).
13. Jeon, M.S. *et al. Immunity* **21**, 167–177 (2004).
14. Loeser, S. & Penninger, J.M. *Semin. Immunol.* **19**, 206–214 (2007).