nature genetics

Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy

Hyun Hor[1,2,28], Zoltán Kutalik[3,4,28], Yves Dauvilliers[2,5], Armand Valsesia[3,4,6], Gert J Lammers[7], Claire E H M Donjacour[7], Alex Iranzo[8], Joan Santamaria[8], Rosa Peraita Adrados[9], José L Vicario[10], Sebastiaan Overeem[11,12], Isabelle Arnulf[13], Ioannis Theodorou[14], Poul Jennum[15], Stine Knudsen[15], Claudio Bassetti[16], Johannes Mathis[17], Michel Lecendreux[18], Geert Mayer[19], Peter Geisler[20], Antonio Benető[21], Brice Petit[1], Corinne Pfister[1], Julie Vienne Bürki[1], Gérard Didelot[1], Michel Billiard[5], Guadalupe Ercilla[22], Willem Verduijn[23], Frans H J Claas[23], Peter Vollenwider[24], Gerard Waeber[24], Dawn M Waterworth[25], Vincent Mooser[25], Raphaël Heinzer[26], Jacques S Beckmann[3,27], Sven Bergmann[3,4] & Mehdi Tafti[1,26]

Narcolepsy is a rare sleep disorder with the strongest human leukocyte antigen (HLA) association ever reported. Since the associated *HLA-DRB1*1501-DQB1*0602* haplotype is common in healthy populations (15–25%), it has been suggested that it is almost necessary but not sufficient for developing narcolepsy. To further define the genetic basis of narcolepsy risk, we performed a genome-wide association study (GWAS) in 562 European individuals with narcolepsy (cases) and 702 ethnically matched controls, with independent replication in 370 cases and 495 controls, all heterozygous for *DRB1*1501-DQB1*0602*. We found association with a protective variant near *HLA-DQA2* (rs2858884; $P < 3 \times 10^{-8}$). Further analysis revealed that rs2858884 is strongly linked to *DRB1*03-DQB1*02* ($P < 4 \times 10^{-43}$) and *DRB1*1301-DQB1*0603* ($P < 3 \times 10^{-7}$). Cases almost never carried a *trans DRB1*1301-DQB1*0603* haplotype (odds ratio = 0.02; $P < 6 \times 10^{-14}$). This unexpected protective HLA haplotype suggests a virtually causal involvement of the HLA region in narcolepsy susceptibility.

Narcolepsy-cataplexy is a sleep disorder characterized by invalidating excessive daytime sleepiness and cataplexy (loss of muscle tone triggered by strong emotions)[1]. Epidemiologic surveys estimate a disease prevalence ranging from 0.03% to 0.16% depending on ethnicity. However, narcolepsy is still an underdiagnosed condition.

Extensive efforts over the past 30 years to gain a better understanding of the genetic basis of the disease have identified a striking association with the HLA region[2]. Nearly 100% of individuals of European ancestry affected by narcolepsy with cataplexy carry the HLA haplotype *DRB5*0101-DRB1*1501-DQA1*0102-DQB1*0602* (ref. 2). This association is thought to represent a virtually necessary but not sufficient risk factor, because 15–25% of healthy individuals in the general population carry the associated HLA haplotype. More recently, it has been shown that the best biological marker for narcolepsy is a deficiency in the hypothalamic neuropeptide hypocretin-1 (orexin-A), which has significantly reduced, if not undetectable, levels in the cerebrospinal fluid of almost all *HLA-DRB1*1501*-positive individuals with narcolepsy and cataplexy[3]. This, together with the selective loss of hypocretin neurons in the lateral hypothalamus[4,5], has led to the hypothesis that narcolepsy is caused by an autoimmune attack targeting hypocretin-producing neurons. Recently, we reported the identification of circulating TRIB2-specific antibodies reactive with hypocretin neurons, confirming the autoimmune nature of narcolepsy[6] for the first time.

[1]Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland. [2]Institut National de la Santé et de la Recherche Médicale U888, Montpellier, France. [3]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [4]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [5]National Reference Network for Orphan Diseases (Narcolepsy and Idiopathic Hypersomnia), Department of Neurology, Gui-de-Chauliac Hospital, Montpellier, France. [6]Ludwig Institute for Cancer Research, Lausanne, Switzerland. [7]Department of Neurology, Leiden University Medical Centre, Leiden, The Netherlands. [8]Neurology Service, Hospital Clinic, Barcelona, Spain. [9]Sleep and Epilepsy Unit, Clinical Neurophysiology Department, Gregorio Marañón University Hospital, Madrid, Spain. [10]Histocompatibility, Blood Center of the Community of Madrid, Madrid, Spain. [11]Sleep Medicine Center 'Kempenhaeghe', Heeze, The Netherlands. [12]Department of Neurology, Donders Institute for Neuroscience, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. [13]National Reference Network for Orphan Diseases (Narcolepsy and Idiopathic Hypersomnia), Sleep Disorders Unit, Pitié-Salpêtrière Hospital, Paris, France. [14]Institut National de la Santé et de la Recherche Médicale U543, Pitié-Salpêtrière Hospital, Paris, France. [15]Department of Clinical Neurophysiology, University of Copenhagen, Glostrup, Denmark. [16]Neurocentro (EOC) della Svizzera Italiana, Ospedale Civico, Lugano, Switzerland. [17]Abteilung Klinische Neurophysiologie, Inselspital, Bern, Switzerland. [18]Pediatric Sleep Center, National Reference Network for Orphan Diseases (Narcolepsy and Idiopathic Hypersomnia), Department of Child and Adolescent Psychopathology, Robert Debré Hospital, Paris VII University, Paris, France. [19]Hephata-Clinic for Neurology, Schwalmstadt-Treysa, Germany. [20]Sleep Disorders and Research Center, Department of Psychiatry and Psychotherapy, University Hospital Regensburg, Regensburg, Germany. [21]Unidad de Sueño, Servicio Neurofisiología Clínica, Hospital Universitario La Fe, Valencia, Spain. [22]Immunology Service, CDB, Hospital Clinic, Barcelona, Spain. [23]Department of Immunohaematology and Blood Transfusion, Leiden University Medical Centre, Leiden, The Netherlands. [24]Department of Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [25]Division of Genetics, Research and Development, GlaxoSmithKline, King of Prussia, Pennsylvania, USA. [26]Center for Investigation and Research in Sleep, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [27]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [28]These authors contributed equally to this work. Correspondence should be addressed to M.T. (mehdi.tafti@unil.ch).

Received 28 April; accepted 21 July; published online XX XX 2010; doi:10.1038/ng.XXXX

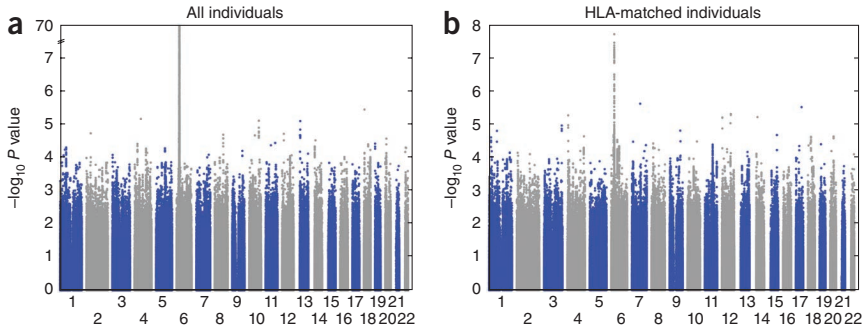

Figure 1 Manhattan plots of association results. Association *P* values ($-\log_{10}$ transformed) are plotted by genomic position. (**a**) Plot of association *P* values using all individuals independent of their HLA genotype. (**b**) Plot of association *P* values using *HLA-DRB1*1501*-matched individuals. Genomic positions are indicated by chromosome with alternating colors.

Although hypocretin deficiency causes the disease, it is not due to mutations in either hypocretin ligand or hypocretin receptor genes, suggesting a complex genetic etiology. A GWAS in a Japanese population[7] has reported evidence for two new susceptibility genes (*CPT1B* and *CHKB*), but these were not confirmed in populations of European ancestry[8]. A more recent GWAS of mainly individuals of European ancestry has identified an association with the T-cell receptor-α (*TRA@*, also known as *TCRA*) locus on chromosome 14, strengthening the autoimmune hypothesis[8]. The discoveries of both the autoantigen (TRIB2) and associations with HLA and *TRA@* suggest that other immune-related gene variants may be involved and remain to be discovered, similar to recent GWAS in a variety of autoimmune disorders[9].

We assessed SNP associations by logistic regression analysis. The first analysis included 562 European cases and 954 control subjects (irrespective of their HLA genotype). Because we included only the typical cases with cataplexy and *HLA-DRB1*1501* positivity, the strongest association ($P < 10^{-68}$) was found within the HLA region on chromosome 6p21 (**Fig. 1a**), as also reported in other HLA-associated disorders[10]. Besides the HLA region, no genetic variant showed genome-wide significant association ($P < 5 \times 10^{-8}$). With our discovery panel, we had 5% to 80% power to detect associations at $P < 1 \times 10^{-5}$ for variants with 15% to 20% minor allele frequency and odds ratios (ORs) in the range of 1.5 to 2, respectively. Variants with $P < 10^{-5}$ are reported in **Supplementary Table 1**. We tested two top candidate SNPs (rs9511411 and rs625106) in an independent sample of cases and controls (see below), but neither could be replicated (**Supplementary Table 2**). Of note, none of the SNPs reported as significant in the Japanese study[7] reached nominal $P < 0.02$. However, given the allele frequency and effect size in Europeans compared to Japanese, our study had only 1.5% power to detect the association (at nominal $P < 0.001$) with SNP rs5770917. Furthermore, in the same study[7], the association could not be replicated in the sample of European descent ($P = 0.12$). The association *P* value for the previously reported[8] *TCA@* SNP (rs1154155) was 4.90×10^{-5}.

In a subsequent HLA-matched analysis using the rs3135388 tag SNP, we included 702 healthy control subjects carrying at least one copy of the *HLA-DRB1*1501* haplotype. Other genomic loci not identified in the first analysis showed potential associations ($P < 10^{-5}$; **Fig. 1b**). These possibly associated variants (none of which overlap with those shown in **Supplementary Table 1**) are listed in **Supplementary Table 3**. In addition to these, rs1154155 within the *TCA@* locus (14q11.2), previously reported to contribute to narcolepsy susceptibility, reached $P < 2 \times 10^{-4}$ (OR = 1.54, 95% confidence interval (CI_{95}) = 1.23–1.93). SNPs in the HLA class II region on chromosome 6, especially near the *HLA-DQA2* locus, were the only ones showing genome-wide significant associations (rs2858884-C, OR = 0.48, CI_{95} = 0.37–0.62, $P = 5 \times 10^{-8}$). However, the specific contribution of any single variant in this region is difficult to determine because of very high polymorphism and tight linkage disequilibrium.

In addition to rs1154155 (*TCA@*), we selected rs2858884, located 8.8 kb upstream of *HLA-DQA2*, for replication. The replication study included only *HLA-DRB1*1501* subjects, comparing 370 narcolepsy cases to 495 controls. Both SNPs were replicated successfully (**Table 1**). rs1154155 in *TCA@* reached $P < 8 \times 10^{-4}$, resulting in a combined $P = 5 \times 10^{-7}$ (OR = 1.55, CI_{95} = 1.31–1.84; **Table 1**). rs2858884 also showed direction-consistent association with $P = 0.038$ (OR = 0.71, CI_{95} = 0.51–0.98) and was replicated with a combined value of $P < 3 \times 10^{-8}$ (OR = 0.56, CI_{95} = 0.46–0.69; **Table 1**). Replication of rs1154155 confirms the *TCA@* association with narcolepsy. Association of rs2858884, located in the HLA class II region, has not been reported before, to our knowledge. Among cases, 14% were found to be homozygous for rs3135388, tagging *DRB1*1501-DQB1*0602*, whereas 8% of controls were homozygous (OR = 1.87, CI_{95} = 1.3–2.67, $P < 6 \times 10^{-4}$), confirming that this homozygous haplotype significantly increases the risk for narcolepsy. In the combined (discovery and replication) sample, the minor allele frequency of SNP rs2858884 (C) was higher (17%) in the control population carrying the *HLA-DRB1*1501* haplotype than in individuals with narcolepsy (10%), demonstrating a strong protective effect of rs2858884.

To test whether rs2858884 is associated independently of any *trans* HLA haplotype, we compared a subset of cases ($n = 426$) and controls

Table 1 Discovery and replication of SNP association data in *HLA-DRB1*1501*-positive individuals

Chr.	Position	SNP	Gene	Distance to gene (kb)	Risk allele	Other allele	*r*-sq-hat	Stage	Effect allele freq.		OR	CI_{95}	*P*
									Controls	Cases			
14	22,072,524	rs1154155	*TCA@*	0	G	T	1	Discovery	0.14	0.20	1.54	[1.23–1.92]	1.92×10^{-4}
								Replication	0.16	0.23	1.56	[1.20–2.04]	7.65×10^{-4}
								Combined	0.15	0.21	1.54	[1.30–1.95]	5.30×10^{-7}
6	32,808,061	rs2858884	*HLA-DQA2*	9	A	C	1	Discovery	0.78	0.90	2.08	[1.61–2.70]	4.65×10^{-8}
								Replication	0.84	0.89	1.41	[1.02–1.96]	3.77×10^{-2}
								Combined	0.81	0.90	1.79	[1.45–2.17]	2.94×10^{-8}

Chr., chromosome; position, physical map position; *r*-sq-hat, imputation quality. In cases where multiple hits were below the threshold and within 10 kb of each other, only the one with the lowest *P* value is reported. Note that ORs and confidence intervals were calculated using logistic regression with covariates. In the discovery phase, the first ten ancestry principal components and rs31135388 genotype were used as covariates, and *P* values and standard errors were genomic control–adjusted. In the replication phase, only the rs31135388 genotype was used as a covariate.

Q15 **Table 2** ***Trans*** **HLA haplotype and rs2858884-C allele frequency in** ***HLA-DRB1*1501-DQB1*0602*** **heterozygous cases and controls**

	Sample size		Case-control association			Allele frequency (rs2858884-C)		rs2858884 association	
DRB1-DQB1 haplotype	Cases	Controls	OR	CI_{95}	*P*	Cases	Controls	Effect size	*P*
DRB1*01-DQB1*05	52	66	0.79	[0.53–1.16]	0.236	0.14	0.10	–0.04	0.437
DRB1*03-DQB1*02	54	80	0.68	[0.46–0.98]	0.0394	0.32	0.41	0.56	3.35×10^{-43}
DRB1*04-DQB1*03	81	81	1.00	[0.71–1.41]	1.00	0.03	0.06	–0.21	1.66×10^{-7}
DRB1*0701-DQB1*0202	40	40	1.00	[0.63–1.58]	1.00	0.11	0.09	–0.08	0.174
DRB1*0701-DQB1*0303	7	15	0.47	[0.19–1.16]	0.129	0.00	0.03	–0.23	0.0258
DRB1*08-DQB1*04	22	14	1.57	[0.79–3.11]	0.234	0.02	0.11	–0.16	0.0423
DRB1*09-DQB1*03	2	8	0.25	[0.05–1.18]	0.107	0.00	0.06	–0.17	0.259
DRB1*10-DQB1*05	1	8	0.12	[0.02–1.00]	0.0381	0.00	0.25	0.18	0.268
DRB1*11-DQB1*03	65	55	1.18	[0.80–1.74]	0.431	0.08	0.05	–0.17	3.64×10^{-4}
DRB1*12-DQB1*03	4	7	0.57	[0.17–1.97]	0.546	0.00	0.00	–0.27	0.0594
DRB1*1301-DQB1*0603	**1**	**51**	**0.02**	**[0.00–0.14]**	**5.28×10^{-14}**	**0.00**	**0.30**	**0.35**	**2.93×10^{-7}**
DRB1*1302-DQB1*0604	34	20	1.70	[0.96–3.01]	0.0687	0.09	0.12	–0.07	0.298
DRB1*1302-DQB1*0609	3	5	0.60	[0.14–2.53]	0.725	0.33	0.40	0.49	4.10×10^{-3}
DRB1*1303-DQB1*0301	15	7	2.14	[0.86–5.31]	0.129	0.03	0.07	–0.18	0.0757
DRB1*14-DQB1*05	17	22	0.77	[0.40–1.48]	0.513	0.12	0.05	–0.12	0.123
DRB1*16-DQB1*05	**28**	**9**	**3.11**	**[1.45–6.68]**	**2.26×10^{-3}**	**0.11**	**0.00**	**–0.11**	**0.163**
	426	488				0.11	0.16		

Q16 Shown in bold are haplotypes that were significant after Bonferroni correction for the number of haplotypes ($n = 16$) tested.

($n = 488$) in whom high-resolution (four-digit) *DRB1-DQB1* typing was available (**Table 2**). The frequency of the protective rs2858884-C allele was found to be increased among both *DRB1*03-DQB1*02* (37%) and *DRB1*1301-DQB1*0603* (30%) individuals independent of case-control status. Even when overall rs2858884-C frequency was again significantly increased in controls (15.6% compared with11.0%, $P < 5 \times 10^{-3}$), analysis conditional on these two haplotypes revealed no association, indicating that rs2858884 preferentially tags *DRB1*03-DQB1*02* and *DRB1*1301-DQB1*0603* haplotypes. Notably, only a single case out of 426 (0.2%) carried the *DRB1*1301-DQB1*0603* haplotype, compared to 51 out of 488 (10.4%) controls (Fisher exact test, Bonferroni corrected $P < 8.5 \times 10^{-13}$). The only *DRB1*1301-DQB1*0603* case is a typical narcolepsy-cataplexy case who does not differ on the *DRB1*1501-DQB1*0602* chromosome compared to other narcolepsy cases nor on the *DRB1*1301-DQB1*0603* chromosome compared to control subjects carrying this haplotype. We also checked for markers that could differentiate this single case from the 51 controls with exactly the same *DRB1-DQB1* haplotype, but we could not identify any difference, indicating that this individual does not carry an unusual recombinant haplotype. A potential protective effect of this haplotype has been suggested in a previous study but is masked by a major *DQB1*0601* protective effect present in the Japanese population[2] (*DQB1*0601* is extremely rare in individuals of European ancestry). The only other haplotype associated with case-control status was *DRB1*16-DQB1*05* (OR = 3.11; CI_{95} = 1.45–6.68, Bonferroni corrected $P < 0.04$).

To elucidate whether the rs2858884 signal is primarily due to tagging of *DRB1*1301-DQB1*0603* or rather *DRB1*03-DQB1*02*, we conducted conditional analysis. When the *DRB1*03-DQB1*02* haplotype was included as a covariate in the logistic regression, rs2858884 was significantly associated ($P = 0.006$). However, when *DRB1*1301-DQB1*0603* was included as a covariate, rs2858884 was not significantly increased ($P = 0.06$). Finally, when both haplotypes were included as covariates, the *P* value for rs2858884 was $P = 0.07$. This suggests that the signal of rs2858884 is more reliant on its linkage disequilibrium with *DRB1*1301-DQB1*0603*.

We also estimated the extent of the discovered protective haplotype (*DRB1*03-DQB1*02*) by phasing our genotype data to reveal all HLA 8.1 ancestral haplotypes present in our population (**Supplementary Note**). We then selected individuals that carried exactly one copy of the *DRB1*03-DQB1*02* haplotype and found extensive variations in the extended *HLA-A*0101-HLA-B*0801* haplotype (**Supplementary Fig. 1**). This indicates that the signal reported here most probably does not extend beyond the HLA class II region.

In addition, we extracted information on copy-number variations (CNVs) from Affymetrix 5.0 SNP arrays used for genotyping (**Supplementary Note**). A genome-wide search for association between narcolepsy and CNVs revealed only three genomic regions, all within the HLA class II region (**Supplementary Fig. 2**). These CNV regions are part of previously identified CNVs[11–16] and are due to the presence or absence of different *DRB1* genes and several pseudogenes within the HLA region. This finding indicates that these CNVs are common genomic reorganizations that can be reliably tagged by SNPs and HLA haplotypes, as has recently been found in several other HLA-associated disorders[17].

Recent SNP-based GWAS in major HLA-associated disorders have concentrated on genetic variants outside the HLA region because of the well-established HLA risk factor. However, in most cases, the identified variants have revealed only a minor risk increment (OR < 2). This is also the case in the two available GWAS of narcolepsy. Our discovery of new variants within the HLA class II region provides important new insights into the enigmatic HLA association with narcolepsy. First, our finding that heterozygous *DRB1-DQB1* cases almost never carry *DQB1*0603*, while *DQB1*0602* and *DQB1*0603* differ only by two amino acids, suggests a major role for the *DQB1* locus. Also, previous comparisons between ethnic groups have indicated that, within the *DRB1-DQB1* region, the susceptibility locus is closer to *DQB1* than *DRB1*; for instance, almost all African-Americans with narcolepsy carry *DQB1*0602* but not *DRB1*1501* (ref. 18), and 15% of the rare individuals of European ancestry who are DRB1*1501-negative with narcolepsy also carry DQB1*0602 (ref. 19). Second, Q13 even if roughly 20% of the general population carry the same *HLA-DRB1*1501-DQB1*0602* haplotype, more individuals with narcolepsy than controls are homozygous for this haplotype, and nearly 30% of heterozygous controls carry a protective *trans* HLA haplotype (*DRB1*03-DQB1*02* and *DRB1*1301-DQB1*0603*). It also worth noting that *DRB1*1301-DQB1*0603* has a much larger effect than

the *TCA@* variant, suggesting that only minor risk factors can be uncovered by GWAS with increased sample sizes. *DQB1*0603*, which is almost never found in people with narcolepsy, is known to be one of the most protective alleles against autoimmune disorders (for example, type 1 diabetes and rheumatoid arthritis). These findings suggest that the HLA region recapitulates most of the disease susceptibility in narcolepsy. As narcolepsy presents the strongest HLA association, the mechanism of its contribution, which remains to be discovered, might help in understanding other HLA-associated disorders.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

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

ACKNOWLEDGMENTS

We thank all the subjects for their generous participation in this study; our colleagues K. Harshman and O. Hagenbuchle and the staff of the Lausanne DNA array facility for their expert technical assistance; and F. Canellas from the University Hospital Son Dureta, Palma-Spain, and P. Young from the Department of Neurology, University Hospital Muenster-Germany for providing additional data on narcolepsy patients. Support for this research was provided by the French Ministry of Research and Higher Education, Project Agence Nationale de la Recherche-07-MRAR (France), PHRC 2007-P070138 (France), European Narcolepsy Network, an unrestricted grant from UCB Pharma S.A. (Belgium), the University and the State of Vaud (Switzerland), the Swiss National Science Foundation (grants 3100AO-108478/2, 3100AO-116323/1 and 310000-112552/1), the European Framework Project 6 (EuroDia and the AnEuploidy projects) and GlaxoSmithKline (which sponsored in part the CoLaus study). S.B. is funded by the Giorgi-Cavaglieri Foundation and the Leenaards Foundation. Special thanks to T. Johnson for his valuable advice on data analysis. The computations were performed in part at the Vital-IT center for high-performance computing of the Swiss Institute of Bioinformatics. This study was conducted on behalf of the European Narcolepsy Network.

AUTHOR CONTRIBUTIONS

Participant recruitment and clinical assessment: Y.D., G.J.L., C.E.H.M.D., A.I., J.S., R.P.A., J.L.V., S.O., I.A., I.T., P.J., S.K., C.B., J.M., M.L., G.M., P.G., A.B., M.B., G.E., F.H.J.C., P.V., G.W., D.M.W., V.M. and R.H. Sample processing and genotyping: H.H., J.L.V., I.T., B.P., C.P., J.V.B., G.D., G.E., W.V. and F.H.J.C. Data analysis: Z.K., H.H., A.V. and M.T. Study design and management: Y.D., G.J.L., J.S.B., S.B. and M.T. Manuscript preparation: H.H., Z.K., Y.D., J.S.B., S.B. and M.T. CoLaus control data: P.V., G.W., D.M.W. and V.M. All authors contributed to the final version of the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

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

1. Tafti, M., Dauvilliers, Y. & Overeem, S. Narcolepsy and familial advanced sleep-phase syndrome: molecular genetics of sleep disorders. *Curr. Opin. Genet. Dev.* **17**, 222–227 (2007).
2. Mignot, E. *et al.* Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. *Am. J. Hum. Genet.* **68**, 686–699 (2001).
3. Nishino, S., Ripley, B., Overeem, S., Lammers, G.J. & Mignot, E. Hypocretin (orexin) deficiency in human narcolepsy. *Lancet* **355**, 39–40 (2000).
4. Peyron, C. *et al.* A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat. Med.* **6**, 991–997 (2000).
5. Thannickal, T.C. *et al.* Reduced number of hypocretin neurons in human narcolepsy. *Neuron* **27**, 469–474 (2000).
6. Cvetkovic-Lopes, V. *et al.* Elevated Tribbles homolog 2 specific antibody levels in narcolepsy patients. *J. Clin. Invest.* **120**, 713–719 (2010).
7. Miyagawa, T. *et al.* Variant between *CPT1B* and *CHKB* associated with susceptibility to narcolepsy. *Nat. Genet.* **40**, 1324–1328 (2008).
8. Hallmayer, J. *et al.* Narcolepsy is strongly associated with the T-cell receptor alpha locus. *Nat. Genet.* **41**, 708–711 (2009).
9. Lettre, G. & Rioux, J.D. Autoimmune diseases: insights from genome-wide association studies. *Hum. Mol. Genet.* **17**, R116–R121 (2008).
10. Klein, J. & Sato, A. The HLA system. Second of two parts. *N. Engl. J. Med.* **343**, 782–786 (2000).
11. Korbel, J.O. *et al.* Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. *Proc. Natl. Acad. Sci. USA* **104**, 10110–10115 (2007).
12. McCarroll, S.A. *et al.* Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
13. Perry, G.H. *et al.* Copy number variation and evolution in humans and chimpanzees. *Genome Res.* **18**, 1698–1710 (2008).
14. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
15. Wheeler, D.A. *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).
16. Wong, K.K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80**, 91–104 (2007).
17. Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720 (2010).
18. Mignot, E. *et al.* DQB1*0602 and DQA1*0102 (DQ1) are better markers than DR2 for narcolepsy in Caucasian and black Americans. *Sleep* **17**, S60–S67 (1994).
19. Mignot, E. *et al.* Extensive HLA class II studies in 58 non-DRB1*15 (DR2) narcoleptic patients with cataplexy. *Tissue Antigens* **49**, 329–341 (1997).

ONLINE METHODS

Genome-wide association study subjects. For our genome-wide association study, we recruited 649 individuals with narcolepsy and clear-cut cataplexy and 280 control subjects from independent European populations (France, Spain, Netherlands, Denmark and Italy). An additional 683 origin-matched and 626 *HLA-DRB1*1501*-positive individuals of European ancestry from the CoLaus project[20] were used as additional control subjects in two different analyses. The design allowed us to perform two association analyses: (i) cases vs. general control population; (ii) cases vs. *DRB1*1501*-positive controls. Owing to the reported high linkage disequilibrium between the *HLA-DRB1*1501* haplotype and rs3135388 (ref. 10), we used this SNP to distinguish between *HLA-DRB1*1501*-positive and *HLA-DRB1*1501*-negative subjects (that is, individuals with at least one T allele were classified as *HLA-DRB1*1501*-positive). We tested rs3135388 in 426 cases and 488 controls with high-resolution HLA typing available who were known to be heterozygous for *DRB1*1501*, and it was found to correctly detect this haplotype in 98% of subjects. Cases were diagnosed according to the guidelines of the International Classification of Sleep Disorders[21]. We included only *HLA-DRB1*1501*-positive cases to ensure a highly homogenous population.

Genotyping and quality control. All cases and non-CoLaus controls were genotyped for 443,816 probes on the GenomeWide Affymetrix 5.0 Array. We used the Affymetrix Power Tool software (apt version 1.10.1), which implements the BRLMM-P algorithm. All calls with a confidence score above 0.05 (default) were considered as unknown/uncalled. SNPs with a call rate below 90% were not considered for further analysis, yielding 435,749 SNPs. Individuals with a call rate below 97% were excluded from the analysis, resulting in 860 called individuals (580 cases, 280 controls).

The additional 683 origin-matched and 626 *DRB1*1501*-positive controls were previously genotyped as part of the CoLaus project[20], with Affymetrix 500K SNP arrays. Genotype calling was done using the BRLMM algorithm with similar quality-control procedures as above.

On both platforms, 405,104 SNPs were called. Allele frequencies in the control samples of the Affy 5.0 and of the Affy 500K samples were compared to that of HapMap. SNPs with Fisher-test *P* values higher than 0.5 (after Benjamini-Hochberg multiple testing corrections) were excluded from further analysis. The allele frequencies of the remaining SNPs among the Affy 5.0 controls were compared to those among the Affy 500K controls using the Fisher test. The *P* values were ranked, and for different thresholds, SNPs were selected whose *P* values were above that threshold. For each threshold (and hence SNP selection), the genomic control coefficient λ was calculated. A *P*-value threshold was selected that resulted in perfectly calibrated *P* values ($\lambda = 1$). Of the 405,104 SNPs, 392,949 passed this filter.

Individuals with suspected relatedness (coancestry coefficient > 1/8; ref. 22) were excluded from the analysis, and only one representative was kept (the one with the highest call rate). Next, ancestry principal components were calculated[23], and individuals whose first two principal components were considered as outliers (more than 5 s.d. away from the mean) were removed. With this, we arrived at our final sample size of 1,264 (562 cases, 702 *DRB1*1501*-positive controls) for the HLA-matched analysis and 1,531 (562 cases, 954 controls) for the nonmatched analysis.

Cases and controls were matched for gender (cases, 298 males and 260 females; controls, 317 males and 371 females), but not for age (cases, 46.6 ± 18.5; controls 53.0 ± 11.6 years), although all control individuals were beyond the major latest peak of age at onset for narcolepsy (40 years old).

Imputation. SNPs were imputed using the MACH 1.0.16 software[24]. We only included in the imputation those measured SNPs whose call rate was above 95%, whose Hardy-Weinberg–equilibrium *P* value was above 10^{-5} and whose minor allele frequency was above 1%.

Statistical analyses. For both the total and HLA-matched samples, genome-wide analysis was carried out by logistic regression, with the first ten ancestry principal components included as covariates. For the HLA-matched analysis, the rs3135388 genotype was also included as covariate to disentangle fully *HLA-DRB1*1501*-independent associations only. Genomic control was applied to the resulting *P* values ($\lambda = 1.02$ and 1.03 for HLA-matched and nonmatched analysis, respectively). We assumed an additive model. False-discovery rate was controlled using the Benjamini-Hochberg step-up procedure[25].

Replication study. For the replication study, we recruited 370 additional narcolepsy cases (from France, Spain, Germany, Switzerland and Italy) and 495 control subjects. 91 German cases were included in an earlier GWAS[8]. All control subjects were positive for *HLA-DRB1*1501*. The replication samples were genotyped by KBiosciences using fluorescence-based competitive allele-specific PCR.

20. Firmann, M. *et al.* The CoLaus study: a population-based study to investigate the epidemiology and genetic determinants of cardiovascular risk factors and metabolic syndrome. *BMC Cardiovasc. Disord.* **8**, 6 (2008).
21. Billiard, M. Diagnosis of narcolepsy and idiopathic hypersomnia. An update based on the International classification of sleep disorders, 2nd edition. *Sleep Med. Rev.* **11**, 377–388 (2007).
22. Weir, B.S., Anderson, A.D. & Hepler, A.B. Genetic relatedness analysis: modern data and new challenges. *Nat. Rev. Genet.* **7**, 771–780 (2006).
23. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
24. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
25. Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. & Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav. Brain Res.* **125**, 279–284 (2001).

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 647]
Author	
Editor	
Publisher	

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections directly on the galley proof. It is also imperative that you include a typewritten list of all corrections and comments, as handwritten corrections sometimes cannot be read or are easily missed. Please verify receipt of proofs via e-mail

Query No.	*Nature of Query*
Q1	Please carefully check the spellings of author names and affiliations.
Q2	Please spell out EOC.
Q3	Please spell out CDB.
Q4	OK?
Q5	OK?
Q6	Note: HGNC official symbol should be used.
Q7	Not clear what respectively means. Should it be moved next to “15% to 20%”? Do you mean the MAFs are about 1.5 and the odds ratios are about 2?
Q8	OK?
Q9	Ref. 7 correct?
Q10	As meant?
Q11	CI95 defined correctly?
Q12	OK? Or “even though”?
Q13	OK?
Q14	OK?
Q15	Should haplotypes in first column be in italics?

QUERY FORM

Nature Genetics	
Manuscript ID	[Art. Id: 647]
Author	
Editor	
Publisher	

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections directly on the galley proof. It is also imperative that you include a typewritten list of all corrections and comments, as handwritten corrections sometimes cannot be read or are easily missed. Please verify receipt of proofs via e-mail

Query No.	*Nature of Query*
Q16	Note: *P* values reformatted to scientific notation; please check they are correct.
Q17	Note: Year corrected.