ImmunoChip Study Implicates Antigen Presentation to T Cells in Narcolepsy

Juliette Faraco, Ling Lin, [...] and Emmanuel Mignot

Abstract

Recent advances in the identification of susceptibility genes and environmental exposures provide broad support for a post-infectious autoimmune basis for narcolepsy/hypocretin (orexin) deficiency. We genotyped loci associated with other autoimmune and inflammatory diseases in 1,896 individuals with hypocretin-deficient narcolepsy and 10,421 controls, all of European ancestry, using a custom genotyping array (ImmunoChip). Three loci located outside the Human Leukocyte Antigen (HLA) region on chromosome 6 were significantly associated with disease risk. In addition to a strong signal in the T cell receptor alpha (TRA@) locus, variants in two additional narcolepsy loci, Cathepsin H (CTSH) and Tumor necrosis factor (ligand) superfamily member 4 (TNFSF14, also called OX40L), attained genome-wide significance. These findings underline the importance of antigen presentation by HLA Class II to T cells in the pathophysiology of this autoimmune disease.

Author Summary

While there is now broad consensus that narcolepsy-hypocretin deficiency results from a highly specific autoimmune attack on hypocretin cells, little is understood regarding the initiation and progression of the underlying autoimmune process. We have taken advantage of a unique high-density genotyping platform (the ImmunoChip) designed to study variants in genes known to be important to autoimmune and inflammatory diseases. Our study of nearly 20,000 narcolepsy cases compared to 10,000 controls underscores important roles for HLA DQB1*06:02 and the T cell receptor alpha genes and implicated two additional genes, Cathepsin H and TNFSF4/OX40L, in disease pathogenesis. These findings are particularly important, as these encoded proteins have key roles in antigen processing, presentation, and T cell response, and they suggest that specific interactions at the immunological synapse constitute the pathway to the disease. Further studies of these genes and encoded proteins may therefore reveal the mechanism leading to this highly selective and unique autoimmune disease.

Introduction

Narcolepsy is a life-long sleep disorder caused by the autoimmune-mediated loss of 70,000–90,000 hypocretin (orexin)-producing neurons in the hypothalamus. Prevalence is approximately 0.02–0.03% in Caucasian populations, and somewhat higher in Japanese (0.16%). Family and twin studies support the importance of genetic (10–40 fold increased risk in first degree relatives) as well as environmental factors (25% concordance in identical twins)(1). Onset is typically around puberty and displays a seasonal pattern of incidence, with highest rates in spring and summer. Likely triggering factors are influenza A, notably the pandemic H1N1 2009 variant, and Streptococcus Pyogenes infections [2]–[5]. Unique among autoimmune diseases, the condition is almost completely associated with Human HLA DQ0602, a heterodimeric protein encoded by the DQA1*01:02-DQB1*06:02 haplotype (90% versus 25% frequency in European ancestry cases and controls, respectively). The overwhelming effect of this haplotype on risk suggests the importance of antigen presentation by DQ0602. As seen in other autoimmune diseases, additional HLA alleles carried in trans of this haplotype also confer modulatory effects [6], [7]. Most notably, DQA1*01:02-DQB1*06:02 homozygosity increases predisposition by 2–4 fold. Further, DQA1 and DQBI alleles known to heterodimerize with DQA1*01:02 or DQB1*06:02 reduce susceptibility, likely through allelic competition with DQ0602 [8]. However, non-HLA related genes play important roles. In addition to these well-established HLA class II effects, recent genome-wide association studies (GWAS) have identified variants in the T Cell receptor alpha locus (TRA@), on chromosome 14q11.2, and in the region containing P2RY11-DNMT1 on chromosome 19p13.2, as additional susceptibility loci. Finally, exome sequencing in families with a rare autosomal dominant syndrome including cerebellar ataxia, narcolepsy and deafness (ADCA-DN) also indicate an important role for DNMT1 in survival of hypocretin neurons [9].

Based on the recognition that considerable overlap exists in risk loci for various autoimmune diseases, the Immunoship consortium was formed to create a single nucleotide polymorphism (SNP) array for targeted finemapping of these loci [10], [11]. The ImmunoChip was designed for deep replication of signals from large-scale meta-analyses in nine autoimmune diseases, and for finemapping of loci reaching genome-wide significance (2009).

Approximately 200,000 rare and common variants were selected to cover intervals with established genome-wide significant association to autoimmune and seronegative diseases, and at selected loci of known importance in major immune-related diseases, including the major histocompatibility (MHC) and KIR/ILIR loci. For this platform, we conducted a GWAS to identify genetic risk factors for narcolepsy in addition to HLA DQ0602. We analyzed 111,400 high quality SNP markers of minor allele frequency ≥1%, located outside the extended HLA region on chromosome 6, in 1,886 narcolepsy cases and 10,421 controls of European ancestry sampled across global populations including the European Union, Canada and North America (Table 1). To test for potential confounding effects of population stratification in our study cohort, we performed principal component analysis (PCA) of cases and controls (Figure 5). Control samples showed clear separation into distinct European countries in plots of the first and second principal components, with good overlay of case samples, and plots of observed versus expected association results showed no inflation of signal (1.004; Figure 1).

Table 1

| Sample collections. |

Results/Discussion

Figure 1

Manhattan Plot of association statistics.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3573113/
One previously reported, and two novel non-HLA loci surpassed genome wide significance (gws) \( P < 5 \times 10^{-8} \) in this study (Figure 1 and Table 2). The strongest association was with rs11534535 (MAF = 0.15, \( P = 8.87 \times 10^{-8} \) OR = 1.72) in the T cell receptor (TCR) alpha (TRA\()@\) locus, on chromosome 14, replicating signal previously reported using smaller samples \([7],[12],[13]\). The TCR protein is comprised of alpha and beta chains. As for immunoglobulin loci, TCR loci undergo somatic DNA recombination during T cell development, generating a large number of possible proteins specific to individual T-cell clones. T cells bearing specific recombinants are then negatively or positively selected, allowing adaptation of the immune system to past environmental history.

The T cell receptor binds foreign or self-peptides presented by Class II MHC proteins (such as the DQ alpha/beta heterodimer), allowing initiation and regulation of immune responses. It is thus the natural receptor of DQB0602. The TRA\()@\) locus was sparsely covered on the ImmunoChip (15 SNPs within a 1 Mb window of rs11534535, none with \( r^2 > 0.5 \)) precluding fine mapping or haplotype analysis, although providing robust replication of the previously reported findings. SNP rs11534535 is located close to the J10 segment region of the locus, with linkage disequilibrium (LD) data suggesting the involvement of a specific J segment in the narcolepsy pathophysiology. The association with TRA\()@\) is unique to narcolepsy, as no other autoimmune diseases have been associated with this locus.

Two SNPs rs34593439, and rs34843303, located in intron 1 of Cathepsin H (CTSH), a papain-like cysteine protease, reached gws (MAF = 0.11, \( P = 1.78 \times 10^{-8} \) OR = 1.34 and MAF = 0.11, \( P = 2.79 \times 10^{-8} \) OR = 1.35, respectively). Another SNP located in intron 1, rs3829932T has been previously reported to be associated with type 1 diabetes \([14],[15]\). Although in close proximity, this marker is in weak LD with rs34593439 and rs34843303 (\( r^2 = 0.23 \) and 0.29 respectively) and shows no significant association in the present sample (\( P = 0.02 \)). The local region of LD surrounding these markers encompasses exon 1, where a potentially functional polymorphism have been identified. One of which, SNP rs2289702T (p.Gly11Arg, MAF = 0.11), is in tight LD with our markers (pairwise \( r^2 = 0.96 \) and 0.98 respectively, 1000genomes data, phase 1 release V3), and could be the culprit behind this association. Following imputation in a 1 Mb window surrounding CTSH, SNPs rs2289702T and rs34593439 were the two most highly associated variants (respectively) (Figure 2). The Arg allele of rs2289702 also underlies a minor histo compatibility antigen restricted by HLA-A*1101 and HLA-A*3303, causing selective lysis of hematopoietic cells by cytotoxic lymphocytes \([16]\). p.Gly11Arg is located within the signal peptide sequence of CTSH, where the introduction of a highly charged arginine could affect trafficking or cleavage, as predicted by some but not all signal peptide predicting programs (see Methods).

Cathepsins are primarily located within the lysosomal/endosomal compartment and typically activated by low pH. These enzymes play diverse and important roles including cellular recycling of proteins, activation of selected preprohormones, antigen processing, and loading of peptides onto MHC class II proteins. Eleven family members are known. Cystatins and other endogenous inhibitors are known to regulate cathepsin activity, and the balance of these activities has been proposed to be the major selector for the repertoire of surface peptide - MHC II complexes \([17]\). Deficiencies in selected cathepsins impair immune cell development (NK cells in Cathepsin S or L deficient mice, thymocytes and T cell repertoire in cathepsin L deficient mice), and produce defects in immune effector functions (cytotoxic T cell, neutrophil and mast cell defects in cathepsin C deficient mice; see \([18]\). Cathepsin H is somewhat unique in that it can have both exopeptidase and endopeptidase activities, depending on the presence of a bound mini-chain (a remnant of the pro-enzyme) within the active cleft. Although ubiquitously expressed, CTSH is also highly expressed in MHC class II positive immune cells such as B cells, monocytes and dendritic cells, but not T cells, notably in the presence of inflammation. For example, CTSH enzyme activity increases in parallel with proinflammatory cytokines during the development of autoimmune inflammation in a NOD mouse model of Sjogren's syndrome \([20]\). One hypothesis may be that decreased CTSH activity reduces antigen processing resulting in an altered repertoire presented by DQB0602 and resulting in increased risk of narcolepsy.

SNPs located in the Tumor Necrosis Factor (ligand) Superfamily member 4 (TNFSF4; also called OX40L or CD252) are strongly associated with narcolepsy. SNP rs17753771 reached gws (MAF = 0.29, \( P = 4.08 \times 10^{-8} \) and OR = 1.33). No other SNP was more strongly associated with narcolepsy following imputation in a 1 Mb window around this locus, although additional strongly associated variants were identified (Figure 2). TNFSF4 is known to be strongly associated with systemic lupus erythematosus (SLE) \([21]\) and systemic sclerosis \([22],[23]\) and SNPs in this region were densely represented on the ImmunoChip. Two distinct haplotypes composed of SNPs upstream of the gene confer susceptibility or resistance to SLE, whereas our most significantly associated SNP markers in narcolepsy are downstream of the gene in a separate haplotype block. The SNPs associated with SLE and narcolepsy are in weak LD, and rs184448, an established marker of SLE, is not strongly associated with narcolepsy (\( P = 0.036 \)). Interestingly, rs17753771 maps to a potential enhancer site (H3K4Me2 site, UCSC browser, Layered H3K27Tr-Ac)

The association of narcolepsy with SNPs in TNFSF4 is consistent with a primary role of antigen presentation to T cells in narcolepsy. Like CTSH, OX40L is primarily expressed in MHC Class II positive antigen presenting cells (e.g. dendritic and B cells). Optimal activation of T cells following the binding of T cell receptor - MHC class II/antigen complex requires the action of additional costimulatory factors, notably involving receptor/ligand pairs from the tumor necrosis superfamily. The interaction of two of these, OX40 receptor (encoded by TNFRSF4) and OX40L ligand (encoded by TNFRSF4), provides an important costimulatory signal supporting Th0 and Th2 responses, promoting expansion and survival of effector T cells and the generation of T memory cells. Although less understood, OX40p/OX40a interactions also play a role in the activity and homeostasis of T regulatory cells. Signaling of this pair is tightly controlled, as OX40p is not expressed in resting T cells, only appearing approximately one day following initial activation. Similarly, OX40L is found...
only at sites of inflammation, first on the surface of antigen presenting cells, but later on diverse cell types including mast cells, suggesting a role distinct from T cell priming or memory cell generation. OX40-OX40L interactions are known to be involved in autoimmune disease, (e.g. SLE) likely acting through a disruption of tolerance. OX40 signaling within responding T cells renders them resistant to Treg mediated suppression, and acts within the Treg cells to inhibit suppressive functions. In addition, sustained inflammatory response may result from excessive OX40-OX40L signaling and consequent increased survival of effector T-cells (see [24],[25]).

Two other regions showed suggestive associations, including SNPs between MDR537 and GJB5 on chromosome 19p34.3 (rs10915302 MAF = 0.84, P = 5.40×10^{-12}, OR = 1.32), and near ZNF95 on chromosome 19p21.2 (rs10969254 MAF = 0.35, P = 3.24×10^{-12} OR = 1.20). ZNF95 is highly expressed in the brain and has been implicated in susceptibility to breast cancer, Crohn’s Disease, and more recently, atopic dermatitis [26]–[28]. None of these reached genome-wide significance levels after correcting with the EMMAX [29] procedure in the current study, although nearly reaching or surpassing Bonferroni significance (P = 4.5×10^{-11}) (Table S1). Increased sample size and replication will be needed to confirm these loci.

Our study, analyzing 1886 narcolepsy-catecholamine cases of European ancestry, is the largest collaborative cohort study of narcolepsy to date, including samples from across the United States, Canada and Europe, and representing the majority of available case samples of European ancestry. To preserve the statistical power afforded by this sample size, we elected not to split our cases into discovery and replication cohorts, and thus our study is limited by the lack of replication in an ethnically similar population. We identified two novel narcolepsy susceptibility genes, CTSN2 and TNFSF4 (OX40L), and confirmed strong associations with HLA and TRA64. The two new loci identified outline with striking clarity that the key pathology underlying narcolepsy likely resides in the interaction between T cells and antigen presenting cells.

Although a role of antigen presentation to CD4 T cells is likely the primary susceptibility pathway for the disorder, narcolepsy was not associated with any components of this pathway as represented on the array. For example, we found no association at the p<10^{-4} threshold with the class II invariant chain, AEP and cathepsin B (CTSB) genes or, more surprisingly, with genes encoding other co-stimulatory molecules such as CD28, cytosolic T-lymphocyte antigen-4 (CTLA4) and their cognate ligands, CD80 and CD86 (these have been involved in many other autoimmune disorders) (see Table S2). The present results also show limited overlap in susceptibility loci between narcolepsy and loci associated with classical autoimmune disorders, a fact that may be unsurprising based on the lack of readily identifiable autoantibodies, or other clear signs of inflammatory damage in the disease. To date, the TCR locus has only been observed in narcolepsy. Notably, we found no associations with loci widely shared among other autoimmune diseases such as interleukin genes and receptors (IL2, IL6, IL7, IL12, IL2RA, IL23RA, IL32, IL33) acting in differentiation; TTPN2 and 22, SH2B3 and TAP2 involved in immune-cell activation and signaling; and JAK5, TNAIP3 involved in TNF signaling and innate immunity (Table S3). Together with findings implicating pandemic H1N1 influenza as a trigger, narcolepsy may offer a unique opportunity, furthering our understanding of how HLA class II presentation of foreign and self-antigens predispose to autoimmunity.

Methods

Ethics statement

Informed consent in accordance with governing institutions was obtained from all subjects. The research protocol at Stanford was approved by the IRB Panel on Medical Human Subjects.

Samples

Cases included in this study all met criteria for narcolepsy/hypocretin deficiency (clear-cut cataplexy and DQB1*06:02 positive, or low cerebrospinal fluid hypocretin-1). Samples included 1301 patients sourced from the Stanford Center for Narcolepsy database (North America, and worldwide collaborators), and 585 samples contributed by the European narcolepsy network (EU-NN). ImmunoChip typing was performed at centers in the US and in Germany. Informed consent in accordance with governing institutions was obtained. Countries of origin included: United States (657), France (296), Italy (157), Germany (157), the Netherlands (111), Czech Republic (104), Canada (101), Austria (83), Denmark (74), Spain (51) and Norway (32). A further 63 cases came from Argentina, Australia, Finland, Israel, Poland, Portugal, Slovakia, Switzerland and Turkey, each with fewer than 20 samples. Control genotypes were contributed through multiple immunochip consortium collaborators including 4289 samples from the United Kingdom 1958 Birth Cohort, 3609 samples from selected European countries including Italy (1251), Netherlands (1173), Poland (529) and Spain (656), 980 samples from the German KORA cohort; 794 Samples from Cincinnati through CCHMC [30]; and 749 French samples (2 collaborators).

Data analysis and statistics

Genotyping of cases was performed following Illumina’s recommendation at U Virginia, USA, U of Munich, Germany, and Stanford University, Palo Alto, CA USA. NCBI build 36 (hg18) mapping was used as reference. Illumina manifest file Immuno_BeadChip_1349691_B.bpm was used in the majority of cases. In cases where file Immuno_BeadChip_1349691_A was used, map positions were converted to be consistent with 1349691_B, or omitted from the analysis. Genotypes were called using Illumina GeneExpress (Illumina GenomeStudio GenTrain2.0 algorithm), with extensive additional curation. Individuals with call rate under 98% (123 controls, 147 cases), and samples which were related (pi hat<0.2) were excluded from further analysis. Data from all sources were merged in forward-strand format. We identified 142,054 high quality SNPs with call rate above 99% (in both cases and controls separately), and passing HWE filtering in controls (P>1×10^{-5}) using the Plink suite of software [31]. We excluded a broad region around the HLA complex (7,893 markers at Chr 6:24,067–35,474 kb) due to the strong LD effects with DQB1*06:02. This region contained nearly 3000 SNPs associated with narcolepsy at GWA significant levels. We additionally excluded SNPs with minor allele frequency below 1% (22,921 SNPs). Finally 111,240 high quality SNPs of MAF≥0.01 (including 91,804 MAF≥0.05) were selected for the analysis presented here. Principal components analysis (PCA) was performed to identify 162 outliers (133 controls, 29 cases); Golden Hella SYSV, 9), and those were removed. Genome wide association analysis was performed using a variance component model implemented in EMMAX [29]. The EMMAX software does not return odds ratio or adjusted allele frequency data after correction for stratification. We therefore calculated OR and MAF (using Plink) for our tables based on a more homogeneous subsample of 8474 cases/controls based on principal components (EV 11.21<0.004, EV 4.12<0.01, see Figure S1). Linkage disequilibrium (as r^2 values and haplotype analysis were calculated using Plink and Haploview [32] using data from our sample and/or from the 1000 genomes dataset [33]). QQ plots were generated using extlambda (http://www.genabel.org/GenABEL/extlambda.html), and Manhattan and PCA plots were made using SVS software.

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3573113/
Imputation and phasing of ImmunoChip genotypes were performed using Beagle v3.3 [34] against 4 European populations (286 individuals from CEU, TSI, GBR, IBS) in the 1000 genomes integrated data set (phase 1 release v3) within a 1 Mb window of the top hit at the CTSH and TNFSF4 loci. SNPs with an imputation R² value ≥ 0.8 (representing reliability of imputation) were considered in the analysis. Pairwise LD was calculated in Plink. Association P values in Figure 2 were calculated with Plink, as EMMAX would be inappropriate in this context, and therefore P values are slightly different than those presented in Table 2.

http://faculty.washington.edu/browning/beagle/beagle.html
http://bochet.gcc.biosstat.washington.edu/beagle/1000_Genomes.phase1_release_v3/

Cleavage prediction
Sequence used: MWATLPLLCAGAWLL[G/R]VPVCGAAELCVNSLEKFHFKSWTSKHRKTYSTEEYHHRLQTFAS
SignalP: http://www.cbs.dtu.dk/services/SignalP/
Both alleles are predicted to have normal cleavage
SigPred: http://bmbpcu36.leeds.ac.uk/prot_analysis/Signal.html/Predicts cleavage unlikely for Arg variant.

Supporting information
Figure S1
Principal components analysis of the study population. Eigenvectors 1 versus 2 in cases and controls are displayed (a–c). Dashed lines in panel a indicate boundaries of a subset of 8474 samples used to calculate OR and allele frequencies (see Methods).

(TIF)
Click here for additional data file.

Table S1
Top ranking non-HLA narcolepsy risk variant signals to P = 1×10⁻⁴. All non-HLA variants (MAF>1% passing QC measures), and with P values<1×10⁻⁴, are displayed. Chr.: Chromosome; BP: position according to NCBI build 36 (Hg18) coordinates; MAF_N: minor allele frequency in narcolepsy (N) and controls (C); P: P value according to variance component model (EMMAX). EMMAX does not provide OR (Odds Ratio) or adjusted allele frequencies, therefore MAF, OR, and 95% confidence intervals (CI) were calculated with Plink on subset of 8474 samples with the greatest homogeneity (see Figure S1; EV 11.21<0.004, EV 4.12<0.01). Threshold for significance using Bonferroni correction: 4.5×10⁻⁷.

(DOCX)
Click here for additional data file.

Acknowledgments
We are grateful to Wake Up Narcolepsy and individual gifts from patients with narcolepsy for making this work possible. We thank all the participating patients, and members of the EU-NN who contributed additional small sets of samples (<50) and coordinated collection for genotyping: Mehdi Tafti, Johannes Mathis, Raphael Heitner, José Haba-Rubio, Aleksandra Wierzbicka, Eva Feketeova, Joan Santamaria, Christian Baumann, Ramin Khataei, and Per Egl Hesla. Matthew Brown coordinated the use of one cohort of French controls. We acknowledge the use of DNA from the British 1958 Birth Cohort collection provided through the Wellcome Trust. We also thank Jing Zhang, Anna Voros for help in sample preparation, Valentine Olivier for participation in imputation and manuscript preparation, and Hyatt E. Moore and Steve Schoettler for providing software scripts. We thank J. Barrett at the Sanger Institute for coordinating ImmunoChip SNP selection and J. Stone at Illumina for coordinating array production.

Funding statement
The study was largely funded by Wake Up Narcolepsy and individual gifts from patients with narcolepsy, with participation of NS23724 to E Mignot. Collection of European samples was partly supported through EU-NN and UCB Pharma grants to M Tafti. French sample collection was funded by PHRC 07-0318 Narcobank grant, by Assistance Publique-Hôpitaux de Paris to I Arnulf, and by the French Ministry of Research and Higher Education, Project Agence Nationale de la Recherche-07-MRAR (France). Samples from the Czech Republic were funded through grant MSM0021620849 to S Nevsimalova. The genotyping performed at University of Virginia utilized resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation International (JDRF) and supported by NIH grant U10 DK064418. The British 1958 Birth Cohort collection is funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 088533/Z/02/2. The KORA study was initiated and financed by the Helmholtz Zentrum München-German Research Center for Environmental Health, which is funded by the German Federal Ministry of Education and Research (BMBF) and by the State of Bavaria. Furthermore, KORA research was supported within the Munich Center of Health Sciences (MC Health), Ludwig-Maximilians-Universität, as part of LMUinnovativ. We also acknowledged the INSERM, ARSEP, and CRB-REFGENSEP for funding French control cohorts. Cincinnati Children’s Hospital Research Foundation funded collection of the CCHMC control cohort, and NIH RC2AR058587 and U01 AI067159 provided funds for genotyping. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Article information
Published online 2013 February 24. doi: 10.1371/journal.pgen.1003270


