Observational Study

Medicine®

OPEN

Constructing gene network for type 1 narcolepsy based on genome-wide association study and differential gene expression analysis (STROBE)

Hui Ouyang, MD[a], Shiying Wang, BS[b], Qiwen Zheng, MD[b], Jun Zhang, PhD[a,*]

Abstract

Although many genes that affect narcolepsy risk have been identified, the interactions among these genes are still unclear. Moreover, there is a lack of research on the construction of the genetic network of narcolepsy. To screen candidate genes related to the onset of narcolepsy type 1, the function and distribution of important genes related to narcolepsy type 1 were studied and a gene network was constructed to study the pathogenesis of narcolepsy type 1.

A case-control study (observational study) of 1075 Chinese narcoleptic patients and 1997 controls was conducted. The gene-sequencing data was analyzed using genome-wide association analysis. The candidate genes related to narcolepsy were identified by differential gene expression analysis and literature research. Then, the 28 candidate genes were input into the KEGG database and 32 pathway data related to candidate genes were obtained. A gene network, with the pathways as links and the genes as nodes, was constructed. According to our results, *TNF*, *MHC II*, *NFATC2,* and *CXCL8* were the top genes in the gene network.

TNF, *MHC II, NFATC2,* and *CXCL8* are closely related to narcolepsy type I and require further study. By analyzing the pathways of disease-related genes and the network of gene interaction, we can provide an outlinefor the study of specific mechanisms of and treatments for narcolepsy.

Abbreviations: EDS = excessive daytime sleepiness, eQTL = expression Quantitative Trait Loci, FDR = false discovery rate, GEO = gene expression omnibus, GIF = genomic inflation factor, GWAS = Genome-wide association study, HCRT = hypocretin, HWE = Hardy-Weinberg equilibrium, KEGG = Kyoto Encyclopedia of Genes and Genomes, MAF = Minor Allele Frequency, MHC II = major histocompatibility complex II, PCA = principle component analysis, Q-Q plot = quantile-quantile plot, REM = rapid eyes movement, SNP = single nucleotide polymorphism, TCR = T cell receptor alpha.

Keywords: gene, gene network, neurodegeneration, pathway, single nucleotide polymorphism

1. Introduction

1.1. The epidemiology and pathogenesis of narcolepsy

Narcolepsy is a life-long neurological disorder characterized by excessive daytime sleepiness (EDS), cataplexy, sleep paralysis, hallucinations, and disrupted nocturnal sleep.[1] Epidemiological data show that the global incidence of narcolepsy is 0.03%.[2] Narcolepsy displays a strong genetic predisposition: the incidence of narcolepsy among first-degree relatives of narcolepsy patients is 1% to 2%, 10 to 40 times that of the normal population,[3] and 25% to 31% of identical twins are co-infected.[4]

Several studies have found that the cerebrospinal fluid hypothalamic secretion, which plays an important role in promoting awakening,[5] is lower in narcolepsy patients.[6] Autopsy results have shown that 90% to 95% of hypothalamic-secretin-producing neurons were lost in some patients with narcolepsy, but the mechanism of neuron loss is still unclear. Recently, the genetic risk of narcolepsy is evaluated based on the carriage of *HLA-DQB1*06:02*, an important but imperfect predictor of narcolepsy,[7,8] since 10% to 40% of individuals in the unaffected population carry it as well.[9] The correlation between influenza A (H1NI) epidemics and increased incidence of narcolepsy suggest that influenza-virus-induced autoimmunity is the possible route of pathogenesis for narcolepsy.[10] However, no specific autoantibodies or T cells have been found to cross-react with hypocretin neurons,[11] hence, the mechanisms of disease occurrence and development still need further study.

1.2. The Genes associated with narcolepsy

Gene-related studies of narcolepsy mainly use genome-wide association analysis and other pathological analyses. Genome-

Editor: Undurti N. Das.

This study has been supported by Research project of central health care special fund (Grant No. W2017BJ52).

The datasets generated during and/or analyzed during the current study are not publicly available, but are available from the corresponding author on reasonable request.

The authors have declared that they have no competing interests.

[a] Department of Clinical Neurology, Peking University, People's Hospital,
[b] Department of Epidemiology and Biostatistics, Peking University Health Science Center, Beijing, China.

[*] Correspondence: Jun Zhang, Department of Clinical Neurology, Peking University, People's Hospital, 11 Xizhimen South Street, Beijing 10044, China (e-mail: zhangjun@pkuph.edu.cn).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc.
This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Ouyang H, Wang S, Zheng Q, Zhang J. Constructing gene network for type 1 narcolepsy based on genome-wide association study and differential gene expression analysis (STROBE). Medicine 2020;99:18 (e19985).

Received: 30 October 2019 / Received in final form: 22 March 2020 / Accepted: 23 March 2020

http://dx.doi.org/10.1097/MD.0000000000019985

wide association study (GWAS) is a method to identify disease or trait-related loci through genetic variation, mainly single nucleotide polymorphism (SNP), using the linkage disequilibrium principle. Many genes have been identified as narcolepsy risk factors in international SNP-based GWASs of narcolepsy type I across different ethnic groups. A Japanese team found that the *CPT1B* gene (carnitine palmitoyltransferase 1B) and the *CHKB* gene (choline kinase B), both involved in REM (rapid eye movement) regulation, are associated with narcolepsy type I.[12] Subsequent genome-wide association analysis revealed that narcolepsy is highly associated with the *TCR alpha* (T cell receptor alpha chain) loci,[13] which has been repeated in subsequent studies.[14] The *TCR alpha* locus produces a unique protein in T lymphocytes and plays an important role in identifying antigens bound to HLA. A GWAS study in a Chinese population showed that *PR2Y11, PPAN, TRB, IL10RB,* and *ZN365* may predict susceptibility to narcolepsy.[15] Among them, *P2RY11* is highly expressed in cytotoxic T lymphocytes and plays a role in regulating cell migration, cytokine release, and apoptosis.[16] Potential susceptibility genes identified by genome-wide association analysis include *CTSH, TNFSF4,*[17]*CCR1, CCR3,*[18]*CLOCK,*[19]*TEAD4, UBXN2-B,*[20]*EIF3G.*[21] In addition, studies combining GWAS with gene pathway information suggest that the genes *CACNA1C, NFATC2, POLE, FAM3D,* and *SCP2* may be associated with narcolepsy.[22]

Mutant gene research found that *MOG*[23] and *DNMT 1*[24] are associated with narcolepsy. *DNMT1* is expressed in immune cells and plays a role in the differentiation of CD4 + T cells into regulatory T cells. Pathological studies have compared the concentration of molecules in cerebrospinal fluid and blood. In addition to the level of hypocretin (HCRT) mentioned above, the concentrations of *BDNF,*[25] GFAP,[26] *TNF alpha,* and *CXCL8*[27] were also significantly different. Differential gene expression studies have found that *IGFBP3* is associated with narcolepsy.[28] However, although so many genetic risk factors for narcolepsy have been identified, the interactions among these genes are still unclear. Moreover, there is a lack of research on the construction of the genetic network of narcolepsy.

2. Materials and methods

2.1. Ethical approval

The research protocols were approved by the Institutional Review Board Panels on Medical Human Subjects at the Peking University People's Hospital. The control group consisted of 1997 staff members and students from several universities.

2.1.1. Genome-wide association study

2.1.1.1. Study population. The study was conducted on 1,075 patients with narcolepsy type 1. The patients were recruited from the sleep laboratory at Peking University People's Hospital. The narcolepsy patients were diagnosed according to the ICSD-3 diagnostic criteria. Cases were all Chinese, and most of them were of Han descent (95%). Clinical data included the presence or absence of cataplexy, sleepiness, sleep paralysis, hypnogogic hallucination, and disturbed nocturnal sleep. Trained interviewers used structured questionnaires to collect information on demographic variables, medical history, and medications. Informed consent (in accordance with governing institutions) was obtained from all subjects.

2.1.1.2. Quality control of the sample and data filtering. Quality control was conducted according to information from the 1075 narcolepsy patients. We excluded patients with narcolepsy type 2 and patients that were suspected to have errors in their input information. In total, 992 patients passed quality control and continued to the next phase of the study.

2.1.1.3. Genotyping and quality control. DNA samples were genotyped on the Affymetrix Axiom CHB array. Genotypes were called using the Affymetrix Genotyping Console. Imputation of 1000 Genomes project SNPs from this array using IMPUTE2 and the Phase I v2, cosmopolitan (integrated) reference panel, build 37.[29] Individuals with call rate <99%, MAF (Minor Allele Frequency) <1%, HWE (Hardy-Weinberg equilibrium), *P* value < .001, or related were removed, leaving 903 cases and 1997 controls.

2.1.1.4. Population stratification analysis. Principal component analysis of the SNPs within linkage equilibrium ($r^2 < 0.25$) was conducted using the PLINK software. Before obtaining the first 20 principal components, people in the cohort who did not cluster with their specified ancestral group (±6SD from the cluster mean on the first 2 principal components) were excluded. To further eliminate the possible impact of population stratification on the results of the association analysis, illness was used as a binary dependent variable to carry out logistic regression and find the principal components with significant differences between the disease group and the control group according to the methods described in the study by Price et al.[30] In subsequent association studies, these principal components were included in the regression analysis as covariates to eliminate them, in order to eliminate the impact of population stratification.

2.1.1.5. Statistical analysis. We conducted the GWAS using the case-control study. Logistic regression was performed using PLINK software. Presence of narcolepsy was the dependent variable. Covariates included the sex of the individual and the first 3 principal components of the principal component analysis. The independent variable was the genotype of the SNP locus. In this study, the error false discovery rate method was used,[31] hence the genome-wide significant *P* value threshold is $P = 4.2^{*}10^{-5}$. Genome-wide association studies were performed using regression models using the R software. A Manhattan map and quantile-quantile plot were drawn, and the genomic inflation factor (GIF) was found.[32] For the region containing multiple positive loci, a conditional regression analysis was conducted using the locus with the lowest *P* value to determine the number of loci associated with disease in the region. A map of the region showing genes and degrees of recombination was drawn using the LocusZoom.[33]

2.1.2. Expression analysis of differential genes

2.1.2.1. Data source. The data were downloaded from the gene expression omnibus (GEO), and the gene expression data were collected from the genome-wide gene expression profile of the human narcolepsy project. The project used chips to collect gene expression data from the circulating lymphocyte mononuclear cells of 10 white narcolepsy patients matched by age and sex and 10 white, healthy controls.

2.1.2.2. Data processing and Statistical Analysis. The background correction, standardization, and logarithmic transforma-

Table 1

The basic and clinical information.

Sex	Male 616 (68.2%)	Female 287 (31.8%)
Ethnicity	Han 860 (95.2%)	Other ethnicities 43 (5.8%)
Native place	Northern 787 (87.2%)	Southern 116 (12.8%)
	Mean (SD)	
Age of onset	10.3 (7.52)	
The age cataplexy	12.5 (8.78)	
Average sleep latency (min)	3.22 (2.19)	
REM sleep latency (min)	2.19 (1.66)	
Apnea and hypopnea index	2.48 (5.45)	

tion of gene expression data were completed according to the original research of Akintomide et al.[34] The gene expression data from the circulating lymphocyte monocytes in the case group and control group were analyzed by ANOVA, and significant differences in gene expression between the 2 groups were compared with the threshold set at $P < .01$. The R software was used to cluster the samples and the significant difference genes, and the thermal map was drawn.

2.1.3. Construction of gene network

2.1.3.1. KEGG database. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database was established in 1995 by Kanehisa Laboratory of the Bioinformatics Center of Kyoto University, Japan.[35] The database provides basic information on genes, DNA sequence, chromosome location sequence, chromosome location, and more, as well as biological pathways.

2.1.3.2. Candidate genes. Candidates for gene network construction in this study come from 3 aspects:

1. Genes related to narcolepsy from literature review, which were partly obtained by genome-wide association analysis and partly from case studies;
2. Potential susceptibility genes identified by the genome-wide association analysis conducted in this study;
3. The first 20 genes with the lowest P value identified by the case-control differential gene expression study using public databases.

2.1.3.3. Pathway information arrangement. The KEGG database was used to retrieve candidate pathway information, analyze the gene pathway information, screen the pathway information for common pathways between genes, and eliminate the unrelated pathways.

2.1.3.4. Construction of the gene network. Pathways in accordance with biological logic were kept, and pathways with approximate functions were eliminated. Then, a gene network with pathways as links and genes as nodes was established.

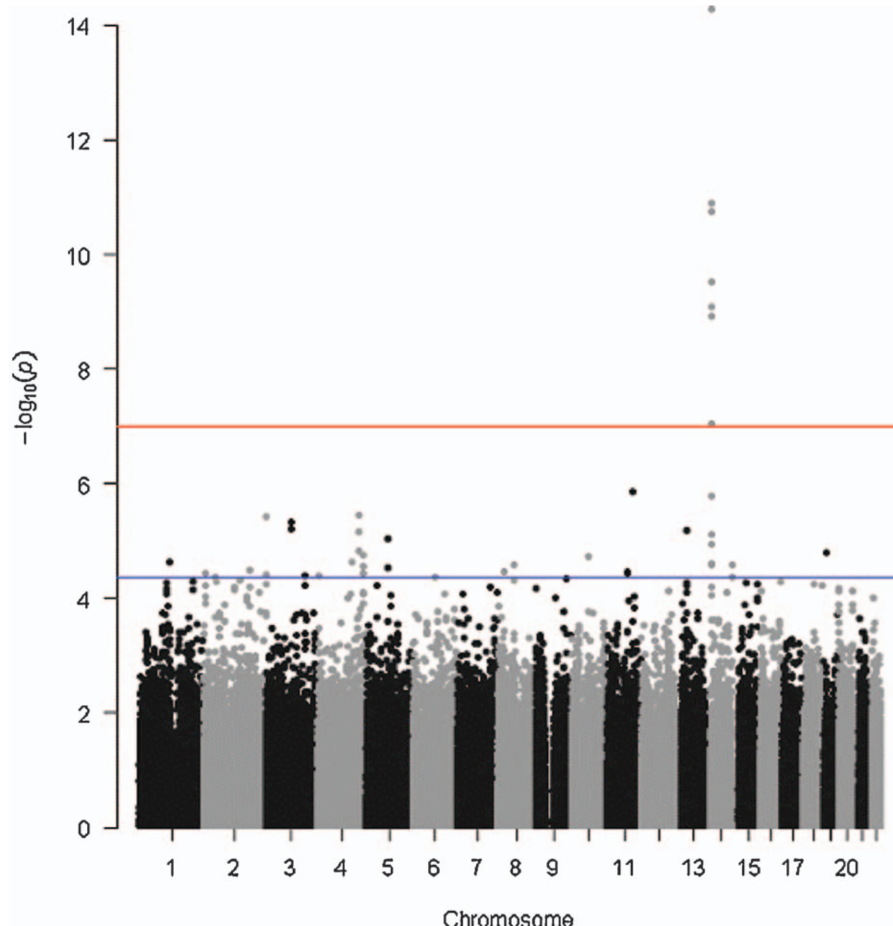

Figure 1. Manhattan plot of the results from the genome-wide association study.

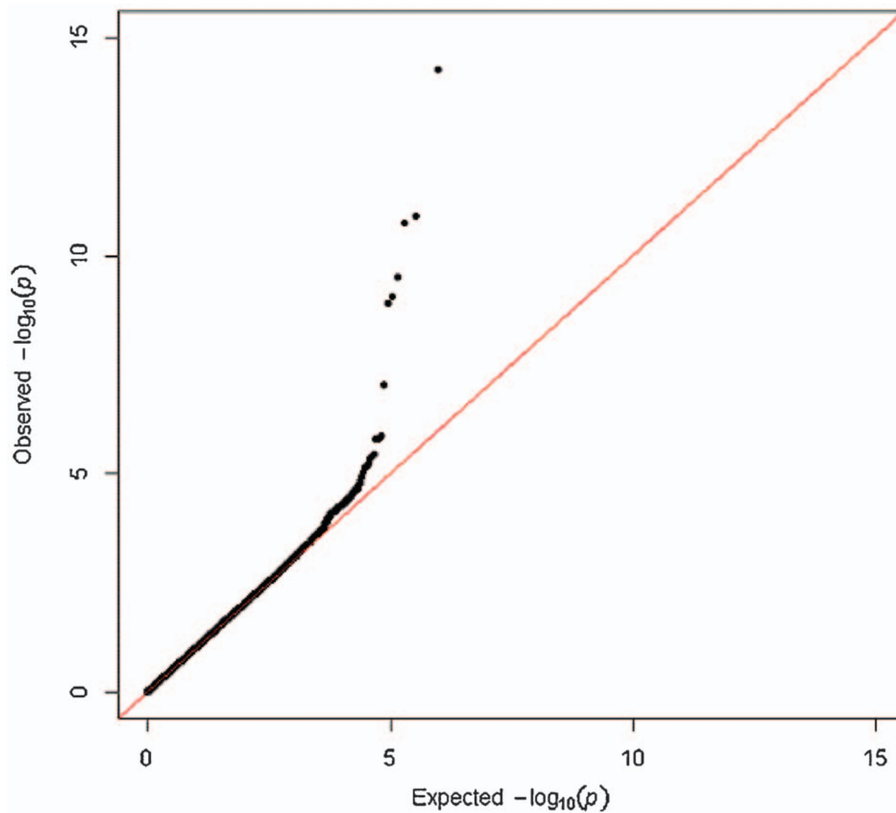

Figure 2. Quantile-Quantile plot.

Table 2

Significant SNPs after FDR correction.

CHR	SNP	Position	Locus	Allele	MAF	*P* value	OR	95%CI
1	rs7524853	116716281		T	0.497226	2.26E-05	0.768	0.6797–0.8677
2	rs6723800	8544487		A	0.441279	3.61E-05	1.297	1.1470–1.4680
2	rs6761161	180158510		A	0.062674	3.17E-05	1.658	1.3070–2.1050
2	rs34931532	242773145		T	0.478624	3.78E-05	0.7782	0.6907–0.8768
2	rs34071003	242779862		G	0.4565582	3.80E-06	1.328	1.1789–1.4980
3	rs11925912	98354506		A	0.224419	6.10E-06	1.387	1.2040–1.5980
3	rs58475214	98354662		G	0.225705	4.57E-06	1.393	1.2090–1.6060
3	rs1439010	152546595		G	0.28678	4.03E-05	0.7529	0.6575–0.8621
4	rs4331842	6481781		C	0.258333	4.14E-05	1.33	1.1610–1.5240
4	rs62328493	161654024		G	0.398432	1.47E-05	0.7535	0.6630–0.8564
4	rs3908007	161659385		A	0.390739	3.54E-06	0.7397	0.6511–0.8475
4	rs3846248	161662411		A	0.377342	6.93E-06	0.7457	0.6561–0.8475
4	rs2581761	176641467		G	0.378819	2.80E-05	0.764	0.6736–0.8665
4	rs6824335	176643742		T	0.380196	1.76E-05	0.7584	0.6685–0.8605
4	rs2715388	179408291		T	0.098472	3.57E-05	0.6303	0.5064–0.7845
5	rs10462313	83223012		T	0.344576	2.97E-05	1.311	1.1540–1.4880
5	rs10942340	83233019		G	0.412153	9.01E-06	1.322	1.1690–1.4960
8	rs7816586	27369334		A	0.45248	3.34E-05	1.291	1.1440–1.4560
8	rs10957264	63905899		T	0.390066	2.59E-05	0.7622	0.6716–0.8650
10	rs10995245	64391375		A	0.306424	1.87E-05	1.329	1.1670–1.5140
11	rs72933139	78795830		A	0.279673	3.51E-05	0.7513	0.6561–0.8603
11	rs17137691	78803340		A	0.434896	3.57E-05	0.7725	0.6835–0.8731
11	rs17882041	102390343		A	0.164873	1.36E-06	0.6593	0.5568–0.7807
13	rs932831	40825382		A	0.302366	6.55E-06	1.346	1.1830–1.5310
19	rs1551570	10218030		C	0.286732	1.61E-05	0.7466	0.6537–0.8526

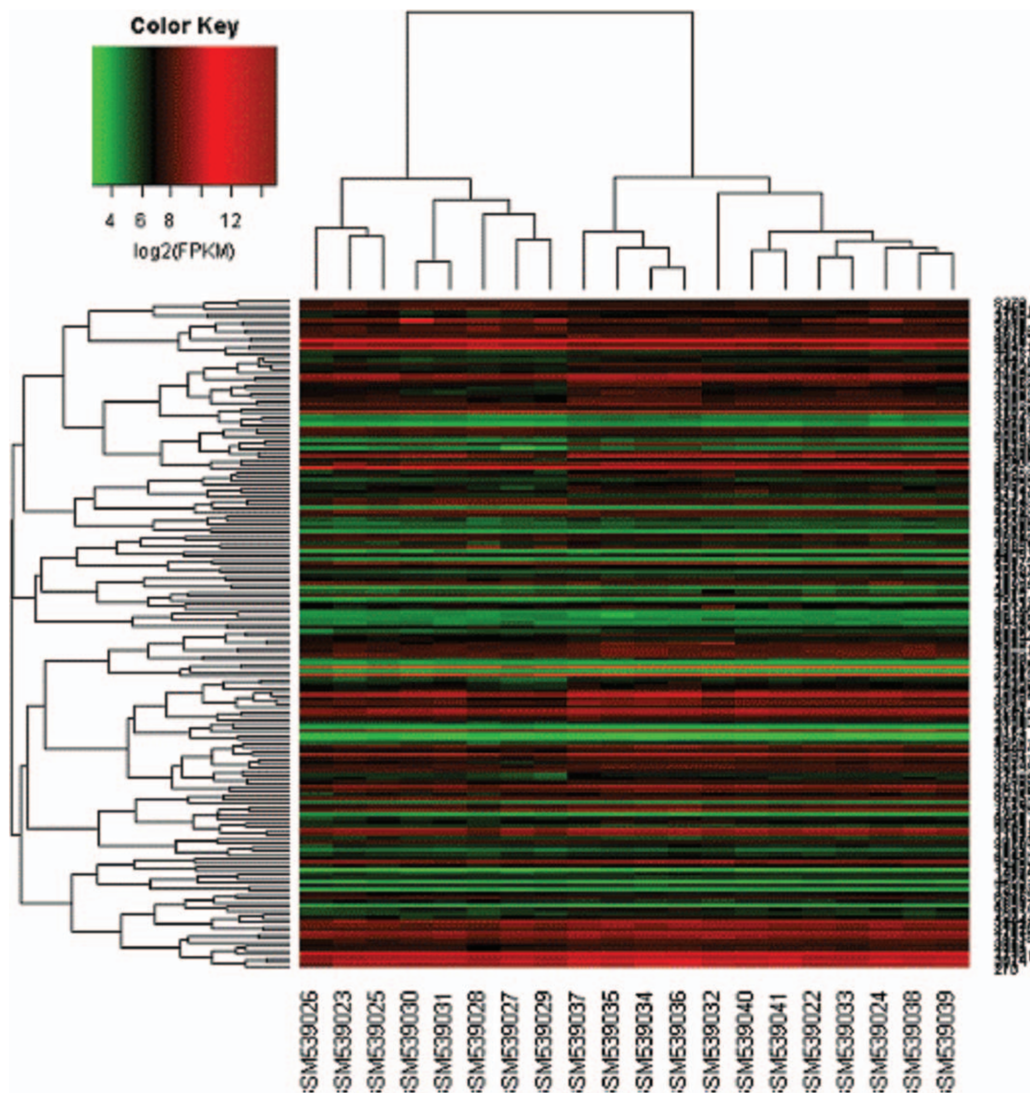

Figure 3. The thermal map of significant expression gene.

3. Results

3.1. Genome-wide association analysis

3.1.1. The sample and clinical data. After a series of strict quality control measures were taken, 903 narcolepsy patients were included in the study. Table 1 shows their basic and clinical information.

3.1.2. The results of the principle component analysis. To avoid the impact of population stratification on the results, principle component analysis (PCA) was performed using genotype data from 237955 SNP loci. Most of the cases were concentrated in the same region as the control group, showing that the genetic background of the case group matched that of the control group. Some sample points deviate from the group, indicating that some population stratification existed. Fifteen sample points that exceed the average+ 6 SD (+0.114) were excluded. These 15 sample points belong to the control group. Therefore, 903 cases and 1982 controls were included in the final correlation analysis.

3.1.3. Manhattan Diagram and Q-Q. In the case-control study, the genotype of SNP loci in non-HLA segments of autosomal chromosomes was correlated with the incidence of disease. Figure 1 is a Manhattan chart showing the overall results of genome-wide association analysis. The genetic variants of 7 SNPs exceeded the threshold for genome-wide significance in this GWAS after Bonferroni correction, and the genetic variants of more SNPs reached the significant threshold after false discovery rate (FDR) correction. The locus with lead SNPs within the TRA gene were among the top signals.

Figure 2 shows the quantile-quantile plot (Q-Q plot) based on the *P* value of each SNP locus in the association analysis. The Q-Q plot is mainly used to show whether the observed values are significantly different from the predicted values. If the SNP loci are not related to disease, the *P* value obtained by analysis should accord with normal distribution, so the theoretical *P* value can be calculated. In the Q-Q plot, since most of the loci are not related to disease, the actual *P* value obtained is basically consistent with the theoretical *P* value, so most of the points in the plot are near

Table 3

The information of the selected gene.

Gene.symbol	*P*.value	logFC	Gene.title	Location
BRD8	.000127	−0.26245737	bromodomain containing 8	5q31
COPA	.000372	0.44129252	coatomer protein complex subunit alpha	1q23.2
TPP1	.000399	0.46648331	Tripeptidyl peptidase 1	11p15
IQCE	.000444	0.32950234	IQ motif containing E	7p22.3
ZNF652	.000479	0.4739675	Zinc finger protein 652	17q21.32
CD44	.000661	0.5359138	CD44 molecule (Indian blood group)	11p13
UTRN	.000708	0.35946283	utrophin	6q24
BPTF	.000914	0.50219388	bromodomain PHD finger transcription factor	17q24.3
WNK1	.000925	0.52634354	WNK lysine deficient protein kinase 1	12p13.3
CHRNA7	.000935	−0.2265192	Cholinergic receptor nicotinic alpha 7 subunit	9q32
PTBP3	.000955	0.50801558	polypyrimidine tract binding protein 3	9q32
GCHFR	.001088	−0.28114824	GTP cyclohydrolase I feedback regulator	15q15
ISG15	.001104	−083426395	ISG15 ubiquitin-like modifier	1p36.33
ZBTB6	.001142	0.28264334	Zinc finger and BTB domain containing t	9q33.2
PCM1	.001151	0.38942709	Pericentriolar material 1	8p22-p21.3
MXD4	.001168	−0.19732428	MAX dimerization protein 4	4p16.3
ARL6IP4	.00117	−0.2692821	ADP ribosylation factor like GTPase 6 interacting protein 4	12q24.31
MZT2A	.001191	−0.41661053	Mitotic spindle organizing protein 2A	2q21.1
MARK4	.001217	0.29174318	Microtubule affinity regulating kinase 4	19q13.3
SERTAD3	.0013	−0.27053366	SERTA domain containing 3	19q13.2

Log2 (FC): log2 (Fold Change), the logarithm based on two is used to express the expression of gene change.

the 45° diagonal line. Only a few of the data points are significantly related to the disease, so the actual *P* value obtained is larger than the theoretical *P* value. In addition, the genomic inflation factor (GIF) can be calculated according to the *P* value of the correlation analysis. If the λ value is too large, the population stratification value is too large in the experimental sample, which has a great impact on the experimental results. In genome-wide association analysis, it is generally believed that when $1 < \lambda < 1.05$, the population structure is relatively small and within acceptable range. In this study, the genome dilator λ value was 1.024. The Q-Q plot and λ value showed that mixed factors such as population stratification were well controlled in the association analysis.

The plot displays 903 individuals and 1982 controls. Single-nucleotide polymorphisms in black are in linkage disequilibrium with the index single-nucleotide polymorphisms and have a *P* value of less than .001. The abscissa represents the SNP loci on chromosome 1–22 in this study, and the longitudinal coordinate is –log 10(p). P is the *P* value of each SNP locus obtained from the correlation analysis. Larger values for -1og 10(p) indicate a greater the correlation between the SNP locus and the disease. The red line in the graph indicates the significant *P* value threshold of 1*10–7 after Bonferonni correction, and the blue line indicates the significant *P* value threshold of 4.2*10–5 after FDR correction.

After FDR correction and excluding the significant *TRA* locus on chromosome 14, the information for the remaining 25 SNP loci is shown in Table 2. Some loci are located in non-coding regions. Due to the difficulty of determining which genes are associated with these loci according to the number of existing loci, this was not analyzed in the follow-up study. Located within the gene, the functions *PPP2R2C, GPM6A, EPHX2, NKAIN3, ZNF365, TENM4,* and *PR2Y11* are related to narcolepsy. As

Table 4

Summary of genes and sources.

Source of gene	Gene			
GWAS	TRA	PPP2R2C	EPHX2	GPM6A
	NKAIN3	ZNF365	TENM4	PR2Y11
Differential expression analysis	BRD8	COPA	TPP1	IQCE
	ZNF652	CD44	UTRN	BPTF
	WNK1	CHRNA7	PTBP3	GCHFR
	ISG15	ZBTB6	PCM1	MXD4
	ARL6IP4	MZT2A	MARK4	SERTAD3
Literature review	DNMT	HLA-DQB1	TNF	HLA-DPB1
	HCRT	TEAD4	BDNF	CXCL8
	IGFBP3	CPT1B	IL10RB	HLA-DRB1
	UBXN2B	HLA-DQA2	CACNA1C	CHKB
	CCR3	NFATC2	CCR1	POLE
	PPAN	HCRTR2	HCRTR1	SCP2
	CTSH	TNFSF4	TRB	MOG
	GFAP	CLOCK	EIF3G	FAM3D

Table 5

The path information in gene networks.

Number of pathway	Name of pathway	Gene		
hsa04062	Chemokine signaling pathway	CXCL8	CCR3	CCR1
hsa04218	Cellular senescence	IGFBP3	NFATC2	CXCL8
hsa05206	MicroRNAs in cancer	DNMT1	CD44	
hsa04640	Hematopoietic cell lineage	CD44	TNF	MHC II
hsa04728	Dopaminergic synapse	CACNA1C	PPP2R2C	CLOCK
hsa04390	Hippo signaling pathway	PPP2R2C	TEAD4	
hsa04152	AMPK signaling pathway	PPP2R2C	CPT1B	
hsa04151	PI3K-Akt signaling pathway	PPP2R2C	BDNF	
hsa04024	cAMP signaling pathway	BDNF	CACNA1C	
hsa04060	Cytokine-cytokine receptor interaction	IL10RB	TNF	CCR3
		CCR1	CXCL8	TNFSF4
hsa04630	Jak-STAT signaling pathway	GFAP	IL10RB	
hsa05168	Cytokine-cytokine receptor interaction	CLOCK	MHC II	TNF
hsa04622	RIG-I-like receptor signaling pathway	TNF	CXCL8	ISG15
hsa04657	IL-17 signaling pathway	TNF	CXCK8	
hsa05166	HTLV-I infection	POLE	NFATC2	MHC
		TNF		
hsa03320	PPAR signaling pathway	SCP2	CPT1B	
hsa04146	Peroxisome	EPHX2	SCP2	
hsa04080	Neuroactive ligand-receptor interaction	HCRTR1	HCRTR2	P2PY11
		CHRNA7		
hsa04725	Cholinergic synapse	CACNA1C	CHRNA7	
hsa04612	Antigen processing and presentation	MHC II	TNF	
hsa04658	Th1 and Th2 cell differentiation	MHC II	TNF	
hsa04659	Th17 cell differentiation	MHC II	TNF	
hsa04620	Toll-like receptor signaling pathway	TNF	CXCL8	
hsa04621	NOD-like receptor signaling pathway	TNF	CXCL8	
hsa04064	NF-kappa B signaling pathway	TNF	CXCL8	
hsa04660	T cell receptor signaling pathway	TNF	NFATC2	
hsa04650	Natural killer cell mediated cytotoxicity	NFATC2	TNF	
hsa04010	MAPK signaling pathway	TNF	CACNA1C	BDNF
hsa04142	Lysosome	CTSH	TPP1	
hsa04210	Apoptosis	CTSH	TNF	
hsa05165	Human papillomavirus infection	ISG15	TNF	PPP2R2C

these genes are potentially associated with narcolepsy, they will be incorporated into subsequent network construction.

3.2. Differential gene expression analysis

There were 168 genes in the differential gene expression analysis ($P<.01$). Thermal maps were drawn using these expression quantities (Fig. 3). The differences in gene expression between the control group and the case group are not noticeable in the figure. The most significant top 20 genes were selected for subsequent gene network construction. The selected gene information is shown in Table 3.

3.3. Construction of gene network

3.3.1. Pathway information collection. The candidates for gene network construction in this study come from 3 aspects:

1. Thirty two genes related to narcolepsy were sorted out in the literature review stage;
2. Eight potential susceptible genes were identified by genome-wide association analysis in this study;
3. Differential gene expression was carried out using a public database.

The top 20 most prominent genes were selected. A total of 60 genes were identified and included in the subsequent analysis, which is summarized in Table 4.

Querying the KEGG database revealed that 33 of the 60 genes contain the pathway information. The results are shown in Table 5. Among them, the *HLA-DQB1, HLA-DRB1, HLA-DPB1,* and *HLA-DQA2* genes co-encode MHC class II molecules, so their pathway information was merged. After merging, the pathways involved in these 30 genes were sorted out to find the pathways in which they participated. A total of 28 genes were found to potentially connect with other genes through the pathways involved. The remaining two genes, *EIF3G* and *CHK*, were considered as isolated genes for the time being, and their pathways were not studied in the following steps.

3.3.2. Construction of Gene network. Of the pathways in which the genes involved were sorted out, only the biologically meaningful pathways were kept. Then, a network of 28 genes and 32 pathways was constructed. The gene network is shown in Figure 4.

4. Discussion

4.1. Analysis of GWAS results

In this study, a GWAS was conducted on 2885 case-control samples. The healthy controls were university staff and students, which may not reflect the normal control population. In order to avoid the potential impact of population stratification on the

Figure 4. The gene network.

results, a principle component analysis was used to verify that the genetic background of the case group matched that of the control group. There may be some false positive loci in the results. In subsequent studies, it would be beneficial to use other datasets or carry out experimental analysis to test the loci, so as to increase the reliability of the results. Some SNPs located in non-gene-coding regions were also found in GWAS research. It is difficult to identify the functions of these regulatory genes. In the follow-up study, we may consider using eQTL (expression Quantitative Trait Loci) analysis, introducing gene expression data from different tissues, and establishing the relationship between gene expression and SNP loci in order to learn which genes are regulated by the SNPs of the non-gene-coding regions found in this study.

4.2. Biological significance of the gene network map

The gene network map is based on MHC II (major histocompatibility complex II) and *TNF* genes. The pathways involved include immune-related viral infections, antigen-presentation-related viral infections, antigen-presentation-related T cell signaling pathways, lectin receptor cytokine interaction, and a series of other pathways. Pathway types include dopamine synapses, cholinergic synapses, and ligand-receptor interactions with neurotransmitters associated with the nervous system. Existing evidence shows that narcolepsy type I may be caused by an autoimmune response in which hypothalamic secretin neurons are attacked. However, no antibodies or T cells corresponding to this process have been found, and the presence of the blood-brain barrier prevents antibodies from easily contacting the brain, so the specifics of pathogenesis are not clear. In this study, the genes associated with narcolepsy type I were linked by gene pathways, and the gene network map was drawn accordingly. Linkage between the nervous and immune systems was highlighted, which brought to light potential pathogenesis pathways for narcolepsy and provided a reference for studying its mechanisms and treatment methods.

4.3. Limitations

This study encountered several limitations. Only gene expression data from European patients were included, since the genetic

variation underlying narcolepsy may differ across different ethnicities. Future studies with Asian narcoleptic patients will be possible when the necessary data are available. Another limitation of the study was that the healthy controls were university staff and students, rather than a completely random sample, In order to avoid bias, principal component analysis was carried out in this study. The results showed that the genetic background of the case group and the control group matched. Additionally, there may be some false positive loci in the results. In the follow-up study, different datasets may be used or experimental analysis to test the loci may be carried out, so as to increase the reliability of the results.

Acknowledgments

We thank Dr Fang Han, Prof. Yiqun Wu and Prof. Dafang Chen for their valuable contributions.

Author contributions

Study design: Jun Zhang, Hui Ouyang, Study performance: Hui Ouyang, Shiying Wang and Jun Zhang. Data analysis and interpretation: Shiying Wang and Qiwen Zheng. Paper writing: Hui Ouyang. All authors approved the final version of the paper.

References

[1] Dauvilliers Y, Arnulf I, Mignot E. Narcolepsy with cataplexy. Lancet 2007;369:499–511.
[2] Longstreth WTJr, Koepsell TD, Ton TG, et al. The epidemiology of narcolepsy. Sleep 2007;30:13–26.
[3] Nishino S, Okura M, Mignot E. Narcolepsy: genetic predisposition and neuropharmacological mechanisms. Review article. Sleep Med Rev 2000;4:57–99.
[4] Mignot E. Genetic and familial aspects of narcolepsy. Neurology 1998;50(2 Suppl 1):S16–22.
[5] Peyron C, Faraco J, Rogers W, et al. A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. Nat Med 2000;6:991–7.
[6] Nishino S, Ripley B, Overeem S, et al. Hypocretin (orexin) deficiency in human narcolepsy. Lancet 2000;355:39–40.
[7] Mignot E, Lin L, Rogers W, et al. Complex HLA-DR and -DQ interactions confer risk of narcolepsy-cataplexy in three ethnic groups. Am J Hum Genet 2001;68:686–99.
[8] Han F, Lin L, Schormair B, et al. HLA DQB106:02 negative narcolepsy with hypocretin/orexin deficiency. Sleep 2014;37:1601–8. doi:10.5665/sleep.4066.
[9] Yamasaki M, Miyagawa T, Toyoda H, et al. Evaluation of polygenic risks for narcolepsy and essential hypersomnia. J Hum Genet 2016;61: 873–8.
[10] Ahmed SS, Schur PH, MacDonald NE, et al. Narcolepsy, 2009 A(H1N1) pandemic influenza, and pandemic influenza vaccinations: what is known and unknown about the neurological disorder, the role for autoimmunity, and vaccine adjuvants. J Autoimmun 2014;50:1–1.
[11] Rose NR, Bona C. Defining criteria for autoimmune diseases (Witebsky's postulates revisited). Immunol Today 1993;14:426–30.
[12] Miyagawa T, Kawashima M, Nishida N, et al. Variant between CPT1B and CHKB associated with susceptibility to narcolepsy. Nat Genet 2008;40:1324–8.
[13] Hallmayer J, Faraco J, Lin L, et al. Narcolepsy is strongly associated with the T-cell receptor alpha locus. Nat Genet 2009;41:708–11.
[14] Hor H, Kutalik Z, Dauvilliers Y, et al. Genome-wide association study identifies new HLA class II haplotypes strongly protective against narcolepsy. Nat Genet 2010;42:786–9.
[15] Han F, Faraco J, Dong XS, et al. Genome wide analysis of narcolepsy in China implicates novel immune loci and reveals changes in association prior to versus after the 2009 H1N1 influenza pandemic. PLoS Genet 2013;9:e1003880.
[16] Junger WG. Immune cell regulation by autocrine purinergic signalling. Nat Rev Immunol 2011;11:201–12.
[17] Faraco J, Lin L, Kornum BR, et al. ImmunoChip study implicates antigen presentation to T cells in narcolepsy. PLoS Genet 2013;9: e1003270.
[18] Toyoda H, Miyagawa T, Koike A, et al. A polymorphism in CCR1/ CCR3 is associated with narcolepsy. Brain Behav Immun 2015;49: 148–55.
[19] Moreira F, Pedrazzoli M, Dos Santos Coelho FM, et al. Clock gene polymorphisms and narcolepsy in positive and negative HLA-DQB1*0602 patients. Brain Res Mol Brain Res 2005;140:150–4.
[20] Luca G, Haba-Rubio J, Dauvilliers Y, et al. Clinical, polysomnographic and genome-wide association analyses of narcolepsy with cataplexy: a European Narcolepsy Network study. J Sleep Res 2013;22:482–95. doi:10.1111/jsr.12044.
[21] Holm A, Lin L, Faraco J, et al. EIF3G is associated with narcolepsy across ethnicities. Eur J Hum Genet 2015;23:1573–80.
[22] Shimada M, Miyagawa T, Kawashima M, et al. An approach based on a genome-wide association study reveals candidate loci for narcolepsy. Hum Genet 2010;128:433–41.
[23] Hor H, Bartesaghi L, Kutalik Z, et al. A missense mutation in myelin oligodendrocyte glycoprotein as a cause of familial narcolepsy with cataplexy. Am J Hum Genet 2011;89:474–9.
[24] Winkelmann J, Lin L, Schormair B, et al. Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. Hum Mol Genet 2012;21:2205–10. doi:10.1093/hmg/dds035.
[25] Klein AB, Jennum P, Knudsen S, et al. Increased serum brain-derived neurotrophic factor (BDNF) levels in patients with narcolepsy. Neurosci Lett 2013;544:31–5.
[26] Feneberg E, Steinacker P, Lehnert S, et al. Elevated glial fibrillary acidic protein levels in the cerebrospinal fluid of patients with narcolepsy. Sleep Med 2013;14:692–4.
[27] Tanaka S, Honda M, Toyoda H, et al. Increased plasma IL-6, IL-8, TNF-alpha, and G-CSF in Japanese narcolepsy. Hum Immunol 2014;75: 940–4.
[28] Honda M, Eriksson KS, Zhang S, et al. IGFBP3 colocalizes with and regulates hypocretin (orexin). PLoS One 2009;4:e4254.
[29] Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
[30] Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
[31] Benjamini YHY. Controlling the false discovery rate: a pratical and powerful approach to multiple testing. R Stat Soc Ser 1995; 289–300.
[32] Yang J, Weedon MN, Purcell S, et al. Genomic inflation factors under polygenic inheritance. Eur J Hum Genet 2011;19:807–12.
[33] Pruim RJ, Welch RP, Sanna S, et al. Locus Zoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26: 2336–7.
[34] Bernardini C, Lattanzi W, Bosco P, et al. Genome-wide gene expression profiling of human narcolepsy. Gene Expr 2012;15:171–81.
[35] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.