Published in final edited form as:

Lancet Oncol. 2016 September ; 17(9): 1240–1247. doi:10.1016/S1470-2045(16)30148-6.

Genetic risk of extranodal natural killer T-cell lymphoma: a genome-wide association study

Zheng Li*,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Genome Institute of Singapore, Singapore

Yi Xia*,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Li-Na Feng*,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Jie-Rong Chen*,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Hong-Min Li*,

State Key Laboratory of Molecular Oncology, Beijing, China

Department of Etiology and Carcinogenesis, Beijing Key Laboratory for Carcinogenesis and Cancer Prevention, Beijing, China

Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Correspondence to: Prof Jin-Xin Bei, Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou 510060, China beijx@sysucc.org.cn.

*Contributed equally

†Senior authors who jointly directed the study

Contributors

CCK, J-XB, R-JP, WT, and ZL designed the study. CCK and J-XB procured financial support. ZL, YX, L-NF, J-RC, H-ML, Q-QC, WYM, L-ZC, Q-SF, FZ, Y-HL, Y-MG, CPP, LJC, SHC, RPE, JH, LPK, STC, Y-XZ, STL, SYT, RQ, MT, CPD, FM, TT, CKO, M-LN, YL, SBN, WT, J-XB, and R-JP collected and prepared the samples. ZL, YX, L-NF, J-RC, H-ML, WYM, W-SL, Y-MG, L-ZC, Q-SF, FZ, Y-HL, CPP, LJC, SHC, RPE, JL, JH, LPK, STC, Y-XZ, STL, SYT, RQ, MT, CPD, FM, TT, CKO, M-LN, YL, SBN, WT, R-JP, CCK, and J-XB collected the data. ZL, YX, L-NF, J-RC, H-ML, JC, Q-QC, KSS, WYM, W-SL, Y-MG, SHC, RPE, JNF, JL, TA, BC, Y-HL, BTT, WJC, YO, CKO, STL, WT, R-JP, CCK, and J-XB analysed and interpreted the data. ZL, KSS, JNF, SR, JC, YO, L-NF, CCK, and J-XB did the statistical analyses. ZL, WT, R-JP, SR, CCK, and J-XB wrote the report. All authors approved the final version.

Declaration of interests

We declare no competing interests.

For PyMOL see <https://www.pymol.org>

For GLOBOCAN 2012 see <http://globocan.iarc.fr>

See Online for appendix

For the SNP2HLA algorithm see <https://www.broadinstitute.org/mpg/snp2hla>

For the European Bioinformatics Institute GWAS catalogue see <http://www.ebi.ac.uk/gwas>

For the Genetic Power Calculator see <http://pngu.mgh.harvard.edu/~purcell/gpc>

Jing Cui,

Division of Rheumatology, Immunology, and Allergy, Division of Genetics, and Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Qing-Qing Cai,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Kar Seng Sim,

Genome Institute of Singapore, Singapore

Maarja-Liisa Nairismägi,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Yurike Laurensia,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Wee Yang Meah,

Genome Institute of Singapore, Singapore

Wen-Sheng Liu,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Yun-Miao Guo,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Li-Zhen Chen,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Qi-Sheng Feng,

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Chi Pui Pang,

Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

Li Jia Chen,

Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong Special Administrative Region, China

Soo Hong Chew,

Department of Economics and Department of Psychology, National University of Singapore, Singapore; Beijing Hospital, Beijing, China

Richard P Ebstein,

Department of Economics and Department of Psychology, National University of Singapore, Singapore; Beijing Hospital, Beijing, China

Jia Nee Foo,
Genome Institute of Singapore, Singapore

Jianjun Liu,
Genome Institute of Singapore, Singapore

Jeslin Ha,
Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Lay Poh Khoo,
Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Suk Teng Chin,
Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Yi-Xin Zeng,
Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Department of Economics and Department of Psychology, National University of Singapore, Singapore; Beijing Hospital, Beijing, China

Tin Aung,
Singapore Eye Research Institute, Singapore

Balram Chowbay,
Laboratory of Clinical Pharmacology, Division of Medical Sciences, Humphrey Oei Institute of Cancer Research, National Cancer Centre, Singapore

Clinical Pharmacology, SingHealth, Singapore

Office of Clinical Sciences, Program in Cancer and Stem Cell Biology, and Office of Education, Duke–National University of Singapore Medical School, Singapore

Colin Phipps Diong,
Department of Haematology, and Department of Pathology, Singapore General Hospital, Singapore; Department of Pathology, Guangdong General Hospital, Guangzhou, China

Fen Zhang,
Department of Haematology, and Department of Pathology, Singapore General Hospital, Singapore; Department of Pathology, Guangdong General Hospital, Guangzhou, China

Yan-Hui Liu,
Department of Haematology, and Department of Pathology, Singapore General Hospital, Singapore; Department of Pathology, Guangdong General Hospital, Guangzhou, China

Tiffany Tang,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Miriam Tao,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Richard Quek,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Farid Mohamad,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Soo Yong Tan,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Department of Haematology, and Department of Pathology, Singapore General Hospital, Singapore; Department of Pathology, Guangdong General Hospital, Guangzhou, China

Institute of Molecular and Cell Biology, A*STAR, Singapore

Department of Pathology, University of Malaya, Kuala Lumpur, Malaysia

Department of Pathology, and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Bin Tean Teh,

Office of Clinical Sciences, Program in Cancer and Stem Cell Biology, and Office of Education, Duke–National University of Singapore Medical School, Singapore

Institute of Molecular and Cell Biology, A*STAR, Singapore

Laboratory of Cancer Epigenome, Division of Medical Sciences, National Cancer Centre Singapore, Singapore

Cancer Science Institute of Singapore, National University of Singapore, Singapore

Siok Bian Ng,

Department of Pathology, and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Cancer Science Institute of Singapore, National University of Singapore, Singapore

Department of Pathology, National University Hospital, and Department of Haematology-Oncology, National University Cancer Institute of Singapore, National University Health System, Singapore

Wee Joo Chng,

Cancer Science Institute of Singapore, National University of Singapore, Singapore

Department of Pathology, National University Hospital, and Department of Haematology-Oncology, National University Cancer Institute of Singapore, National University Health System, Singapore

Choon Kiat Ong,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Yukinori Okada,

Department of Pathology, National University Hospital, and Department of Haematology-Oncology, National University Cancer Institute of Singapore, National University Health System, Singapore

Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan

Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan

Soumya Raychaudhuri,

Division of Rheumatology, Immunology, and Allergy, Division of Genetics, and Division of Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Partners Center for Personalized Genetic Medicine, Boston, MA, USA

Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

Institute of Inflammation and Repair, University of Manchester, Manchester, UK

Institute of Inflammation and Repair, University of Manchester, Manchester, UK

Soon Thye Lim,

Lymphoma Genomic Translational Research Laboratory, Division of Medical Oncology, National Cancer Centre Singapore, Singapore

Office of Clinical Sciences, Program in Cancer and Stem Cell Biology, and Office of Education, Duke-National University of Singapore Medical School, Singapore

Wen Tan[†],

Department of Etiology and Carcinogenesis, Beijing Key Laboratory for Carcinogenesis and Cancer Prevention, Beijing, China

State Key Laboratory of Molecular Oncology, Beijing, China

Cancer Institute and Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Rou-Jun Peng[†],

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Chiea Chuen Khor[†],

Genome Institute of Singapore, Singapore

Singapore Eye Research Institute, Singapore

Department of Pathology, and Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

Jin-Xin Bei[†]

Sun Yat-Sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, China

Genome Institute of Singapore, Singapore

Center for Precision Medicine, Sun Yat-Sen University, Guangzhou, China

Summary

Background—Extranodal natural killer T-cell lymphoma (NKTCL), nasal type, is a rare and aggressive malignancy that occurs predominantly in Asian and Latin American populations. Although Epstein-Barr virus infection is a known risk factor, other risk factors and the pathogenesis of NKTCL are not well understood. We aimed to identify common genetic variants affecting individual risk of NKTCL.

Methods—We did a genome-wide association study of 189 patients with extranodal NKTCL, nasal type (WHO classification criteria; cases) and 957 controls from Guangdong province, southern China. We validated our findings in four independent case-control series, including 75 cases from Guangdong province and 296 controls from Hong Kong, 65 cases and 983 controls from Guangdong province, 125 cases and 1110 controls from Beijing (northern China), and 60 cases and 2476 controls from Singapore. We used imputation and conditional logistic regression analyses to fine-map the associations. We also did a meta-analysis of the replication series and of the entire dataset.

Findings—Associations exceeding the genome-wide significance threshold ($p < 5 \times 10^{-8}$) were seen at 51 single-nucleotide polymorphisms (SNPs) mapping to the class II MHC region on chromosome 6, with rs9277378 (located in *HLA-DPB1*) having the strongest association with NKTCL susceptibility ($p = 4.21 \times 10^{-19}$, odds ratio [OR] 1.84 [95% CI 1.61–2.11] in meta-analysis of entire dataset). Imputation-based fine-mapping across the class II MHC region suggests that four aminoacid residues (Gly84-Gly85-Pro86-Met87) in near-complete linkage disequilibrium at the edge of the peptide-binding groove of HLA-DPB1 could account for most of the association between the rs9277378*A risk allele and NKTCL susceptibility (OR 2.38, p value for haplotype 2.32×10^{-14}). This association is distinct from MHC associations with Epstein-Barr virus infection.

Interpretation—To our knowledge, this is the first time that a genetic variant conferring an NKTCL risk is noted at genome-wide significance. This finding underlines the importance of HLA-DP antigen presentation in the pathogenesis of NKTCL.

Introduction

Natural killer T-cell lymphoma (NKTCL; formally known as extranodal NKTCL, nasal type), a rare and distinct malignancy with an aggressive clinical course,¹ predominantly occurs in Asian and Latin American populations compared with European populations.^{2,3} In two studies examining present best practice,^{4,5} 5-year overall survival was 72% in 155 patients with early-stage NKTCL and around 50% in 87 patients with mainly late-stage disease. Results from genome sequencing studies^{6–9} showed recurrent somatic mutations of *TP53*, *JAK3*, and *DDX3X* in NKTCL tumours, implicating the JAK–STAT, NF- κ B, and MAPK pathways in pathogenesis. Results from comparative genomic hybridisation studies^{10–12} showed an association between NKTCL and deletions on chromosome 6q21 that resulted in downregulated expression of tumour suppressor genes located in the region (*PRDM1*, *ATG5*, *AIM1*, *FOXO3*, and *HACE1*), and results from a population-based epidemiological study¹³ also suggested the involvement of environmental factors in NKTCL pathogenesis, as increased risk was reported in individuals exposed to pesticides and chemical solvents. However, few germline genetic variants have been positively identified to be directly linked to NKTCL risk, with one study¹⁴ reporting significant under-representation of the *HLA-A**0201 allele in patients with NKTCL who were also positive for Epstein-Barr virus (EBV) compared with healthy controls. Besides strong association with EBV-driven pathogenesis, the molecular and genetic mechanisms underpinning susceptibility to this malignancy remain poorly understood.

To examine the possibility that common germline polymorphisms might contribute to NKTCL susceptibility, we did a genome-wide association study (GWAS) on a discovery sample, followed by validation in four independent case-control series.

Methods

Study design and participants

In this GWAS, all cases were patients with extranodal NKTCL, nasal type, as defined by the WHO classification criteria for non-Hodgkin lymphoma.¹ For the discovery sample, we recruited 192 patients between March 1, 2002, and May 25, 2015, who were diagnosed at the Sun Yat-Sen University Cancer Center (SYSUCC), Guangzhou, Guangdong province, southern China, and 981 control individuals between Sept 1, 2009, and Nov 29, 2010, who were cancer free at enrolment from the same areas in Guangdong province.¹⁵

For the first replication series, we included 75 patients with NKTCL from patient records of the Department of Pathology, Guangdong General Hospital, Guangzhou, China, and 296 individuals from a subset of a previous cohort¹⁶ who were cancer-free at enrolment recruited from neighbouring Hong Kong. For the second replication series, we recruited 65 patients with NKTCL from SYSUCC between June 18, 2002, and Sept 24, 2015; in the control group, we included 983 individuals from SYSUCC who were cancer free at the point of enrolment and who were genotyped as the discovery sample in a previous GWAS.¹⁷ For the third replication series, we recruited 125 patients with NKTCL between Sept 1, 2008, and July 6, 2013, at the Cancer Hospital, Chinese Academy of Medical Sciences in Beijing, northern China; the control group included 1110 healthy students from universities in

Beijing who were part of a previous study.¹⁸ For the fourth replication series, 60 patients with NKTCL were enrolled from the National Cancer Centre Singapore, Singapore General Hospital, and National University Hospital, from 1995 to 2015; the control group included 2476 Singaporean Chinese individuals from a previous population-based study.¹⁹ All participants in the fourth replication series were of self-reported southern Han Chinese ancestry. Demographic data, including ancestry or race information, sex, and age, were obtained for all five sample series. We also retrieved hepatitis B virus infection status of patients in our discovery sample only from clinical records of serum antibody concentrations. Details of each recruitment site are shown in the appendix p 8.

Written informed consent was obtained from all participants enrolled in this study, and ethical approval was obtained from all the relevant local and hospital-based Institutional Review Boards, in accordance with the Declaration of Helsinki.

Procedures

We extracted genomic DNA from blood samples of all participants in the discovery sample using Qiagen blood midi or maxi kits (Qiagen, Hilden, Germany) according to manufacturer's instructions. For the replication case-control series, genomic DNA from participants were extracted from blood samples and from formalin-fixed, paraffin-embedded tumour tissues. Genome-wide genotyping was done at iGenostics Biotechnology (Guangzhou, China) and the Genome Institute of Singapore (Singapore); all cases were genotyped with Illumina Human OmniExpress ZhongHua-8 BeadChip, and population controls were scanned by Illumina OmniHumanExpress-24 V1.0 (both Illumina, San Diego, CA, USA).^{20,21} Replication genotyping for the samples in the replication series that were not part of previous studies was done with Taqman allelic discrimination assays (Applied Biosystems, Foster City, CA, USA), with cross-verification using standard Sanger sequencing.

We used stringent quality control filters to remove samples and SNP markers with poor performance in both the discovery (GWAS) and the replication (de-novo genotyping) phases, using PLINK version 1.07.²² SNPs with genotyping completeness of less than 95% were excluded from further analyses, as were SNPs with minor allele frequencies less than 1% and those showing significant deviation from Hardy-Weinberg equilibrium ($p < 1 \times 10^{-6}$). We also excluded SNPs with significantly different genotyping success rates (p value for the difference $< 1 \times 10^{-5}$, as calculated using Fisher's exact test implemented in PLINK). Quality control checks for individual samples were done in a similar manner, excluding samples with genotyping efficiency less than 95%. We removed all samples showing extremes of heterozygosity (which indicate possible cross-contamination) and outliers on principal component analysis of genetic ancestry from further statistical analyses. For pairs of individuals who showed evidence of cryptic relatedness (possibly because of unintended technical error or biologically related samples), we removed the samples with the lower call rate (ie, the proportion of genotypes without missing data per marker) before principal component analysis. We used principal component analysis, done with the R statistical program, to account for spurious associations resulting from ancestral differences of individual SNPs.

The laboratory method for classical HLA typing with next-generation re-sequencing has been described elsewhere.²³ To impute *HLA* alleles, the SNP2HLA algorithm was used with a publicly available and the most geographically matched reference panel, the Pan-Asian reference panel (n=530).^{24–26} The program covers both the common *HLA* alleles and the aminoacids of these classic alleles; subsequently, we directly tested the association of these imputed SNPs, alleles, and aminoacid variations with NKTCL. We applied additional quality control filters to exclude SNP markers with imputation scores less than 0.5. A call rate of less than 99% was used as a filtering criterion for SNPs with a minor allele frequency of less than 1% in either cases or controls. For common SNPs with a minor allele frequency of at least 1%, the filtering criterion was set at less than 95% call rate. GWAS catalogues (US National Human Genome Research Institute²⁷ and European Bioinformatics Institute) were interrogated for genetic variants associated with lymphomas and EBV-related diseases, as well as for disease phenotypes associated with HLA-DPB1. These reported genetic loci were searched against our GWAS results.

Statistical analysis

For both discovery and replication stages, association analysis comparing SNP genotypes between cases and controls was done with conditional logistic regression, which models a trend-per-copy effect of the minor allele on disease risk. Genome-wide significance was set as $p < 5 \times 10^{-8}$, as proposed previously by correcting for 1 million independent tests in the human genome.²⁸ For the discovery stage, we verified the absence of confounding of genetic results as a result of population stratification by incorporating up to 20 top principal components of genetic stratification into the association using logistic regression modelling in PLINK. Meta-analysis was done with the inverse-variance method, which weighs each strata according to sample size and allele frequency. Meta-analysis with fixed-effect was also done with PLINK. We measured heterogeneity within the meta-analysis with the I^2 index of heterogeneity and calculated the p value of heterogeneity (p_{het}) using Cochran's Q test. Genotyping cluster plots for genetic markers surpassing genome-wide significance on final meta-analysis of discovery and replication stages were visually checked and verified to be of good quality. Haplotype analysis was done with PLINK.

All statistical power calculations were done as previously described for two-staged GWAS and replication studies.²⁹ We present these power calculations for each of the following conditions: discovery cohort alone and discovery cohort plus replication cohorts (appendix p 9). Statistical power calculations were done with a standard Genetic Power Calculator for genetic effects,³⁰ which measures statistical power as a function of the per-copy odds ratio (OR) of the risk allele in the context of risk allele frequency. Statistical power calculation showed that our two-staged study design is powered to detect genetic ORs as low as 1.8 per copy of the risk allele at a minor allele frequency as low as 0.01 with 90% statistical power at genome-wide significance (appendix p 9).

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. CCK and J-XB had access to the raw data. The

corresponding author had full access to all the data and the final responsibility to submit for publication.

Results

After stringent quality checks on the post-experimental data on a per-SNP marker and per-sample basis, we included 189 (98%) of 192 patients with NKTCL, 957 (98%) of 981 controls, and 477 644 (98%) of 489 013 directly genotyped SNPs in the discovery sample for further analyses. Visualisation of the top principal component vectors showed that cases and controls were well matched in terms of genetic ancestry, and all individuals were located within 6 SDs from the mean of each principal component vector (appendix pp 1–5).^{17,19,31,32} A quantile–quantile plot revealed no genomic inflation (genomic inflation factor [λ_{GC}] 1.013), suggesting minimal confounding of the analysis by cryptic population stratification (appendix p 6). 51 directly genotyped SNP markers showed evidence of association with NKTCL surpassing genome-wide significance (appendix pp 6, 10–11).

All 51 SNPs mapped to the class II MHC region on chromosome 6, and the SNP with the strongest association (the so-called sentinel SNP) was rs9277378 ($p=5.84 \times 10^{-16}$, OR 2.65 [95% CI 2.08–3.37]; figure; table 1) located within *HLA-DPB1*. Because the class II MHC region was the only locus with resident markers surpassing genome-wide significance, we genotyped the sentinel SNP rs9277378 in four independent replication series comprising 325 patients and 4865 controls in total. We observed significant replications of the initial association with NKTCL in each of the four replication case-control groups and in the meta-analysis of all four replication strata without evidence of heterogeneity ($p=5.78 \times 10^{-8}$, OR 1.56 [1.33–1.84]; $p_{het}=0.86$, $I^2=0.0$; table 1). When data from both the discovery stage and the replication stage were included in the meta-analysis, we again observed a genome-wide significant association between *HLA-DPB1* rs9277378 and NKTCL ($p=4.21 \times 10^{-19}$, OR 1.84 [1.61–2.11]; table 1). Although there was some heterogeneity in the meta-analysis when the discovery and four replication series were included ($p_{het}=0.011$, $I^2=69.4\%$), it was considered moderate to high³³ and less extreme than the cutoff ($I^2>75\%$, $p_{het}<0.010$) suggested previously in a large GWAS reporting true associations.³⁴ More importantly, the ORs in the five sample collections showed a consistent direction of effect, suggesting that the heterogeneity was almost entirely driven by the larger estimate in the discovery stage, possibly because of sampling bias (also known as the winner's curse) for patients with NKTCL.

To pinpoint the functional *HLA* genes and alleles that might account for the significant associations of these SNPs, we did an *HLA* fine-mapping analysis of the discovery dataset with the Pan-Asian HLA reference panel as a template.^{25,26} The strongest association was seen at aminoacid positions 84–87 within *HLA-DPB1* ($p=1.01 \times 10^{-14}$; table 2), which are located at the edge of the peptide-binding groove of *HLA-DPB1*. These four aminoacids fully accounted for the association of the NKTCL phenotype and rs9277378, because the association became non-significant ($p=0.02$) after adjustment for allele dosage at *HLA-DPB1* Gly84 via conditional analysis (appendix pp 12–19). To confirm the accuracy of our imputation, we did direct HLA-typing for 45 patients with NKTCL from the discovery cohort. The allelic concordance between the directly typed and imputed alleles for

aminoacid positions 84–87 was 344 (96%) of 360 alleles. The concordance for HLA-typing was 87 (97%) of 90 alleles at four-digit level for *HLA-DPB1**0201, *0401, and *0501, which is consistent with results from recent studies.^{23,35,36}

The associations at aminoacid positions 84–87 of HLA-DPB1 were very similar in strength because of their near-complete pairwise linkage disequilibrium ($r^2 > 0.99$ for all comparisons), which makes them statistically indistinguishable from one another (table 2; appendix p 12).³⁷ Using haplotype estimation (also known as phasing), only two haplotypes were noted when aminoacids at positions 84–87 were phased. The first haplotype was Gly84-Gly85-Pro86-Met87 (in phase with the rs9277378*A risk allele), which was associated with an increased susceptibility to NKTCL. The second haplotype, Asp84-Glu85-Ala86-Val87, was in phase with the rs9277378*G (wild-type) allele and was associated with a decreased susceptibility to NKTCL (table 2). The risk haplotype (Gly84-Gly85-Pro86-Met87) partly defines the classic *HLA-DPB1**0201 and *0401 alleles, and the so-called protective haplotype (Asp84-Glu85-Ala86-Val87) partly defines the *HLA-DPB1**0501 classic allele (table 2). Modelling of the HLA-DPB1 protein using previously described protein structure³⁸ and the protein prediction software PyMOL indicated that the Gly84Asp alteration on the β chain of HLA-DPB1 affects the binding of the Phe1 residue in the p1 peptide of the ligand in terms of both space and charge (appendix p 7).

Despite being adequately powered to detect a difference of more than three times between cases and controls at 80% power using genome-wide significance thresholds (see power calculation in appendix p 9), we were unable to confirm an association at *HLA-A**0201 as seen in a previous study¹⁴ of 25 NKTCL cases and 303 controls of Japanese descent, which reported a substantially lower frequency of *HLA-A**0201 in NKTCL cases (1 [4%] of 25 patients) than in the baseline control population (122 [20%] of 606 alleles among 303 individuals). In our study, the frequency of the *HLA-A**0201 allele was similar in both cases (33 [9%] of 378 alleles in 189 individuals) and controls (158 [8%] of 1914 alleles in 957 individuals; $p=0.76$; appendix p 20).

In our GWAS discovery dataset, 26 (14%) of 189 patients with NKTCL had concurrent chronic hepatitis B virus infection. The association between the *HLA-DPB1* rs9277378*A risk allele and patients with NKTCL who do not have hepatitis B virus infection (OR 2.63, 95% CI 2.04–3.29, $p=6.68 \times 10^{-15}$) was stronger than that compared with patients with NKTCL who do have chronic hepatitis B virus infection (1.51, 0.87–2.72, $p=0.16$, appendix p 22).

We examined publicly available GWAS results for EBV-related malignancies, such as nasopharyngeal carcinoma in southern Chinese populations and Hodgkin's lymphoma in Europeans (appendix pp 20–22) to determine whether they had any association with *HLA-DPB1*. Neither disease was associated with *HLA-DPB1* (appendix pp 20–22). Additionally, we also examined publically available GWAS results to determine whether SNPs associated with lymphomas including diff use large B-cell lymphoma,^{18,39} marginal-zone lymphoma,⁴⁰ and follicular lymphoma are associated with risk of NKTCL.^{41–43} SNPs that are strongly associated with these three lymphomas do not show significant evidence of association with NKTCL (appendix pp 20–22).

Discussion

In our GWAS, we identified a strong association between HLA-DPB1 aminoacids and susceptibility to NKTCL. HLA-DPB1 is the β 1 subunit of the HLA-DP heterodimer, which plays a part in extracellular antigen presentation to CD4-positive T-cell lymphocytes.⁴⁴ Aminoacids 84–87 form one of the key hydrophobic anchor pockets (p1 binding pocket)³⁸ within the peptide-binding groove, accepting the side chains of peptide aminoacids for peptide binding. This step could be crucial for antigen presentation.^{45–48} Indeed, results from a previous study³⁷ showed that the Asp84 residue could repel peptide binding and potentially affect aminoacids in the antigen peptide that would bind in the groove, peptide-binding affinity, or peptide-binding configuration to provide a more optimal antigen presentation and subsequently a stronger immune and inflammatory response. Such a scenario would be consistent with the Asp84 residue being associated with a decreased risk of NKTCL, presumably because of better antigen recognition and tumour clearance.

NKTCL has been thought to be related to EBV infections.⁴⁹ We examined publicly available GWAS results for EBV-related malignancies, such as nasopharyngeal carcinoma in southern Chinese populations and Hodgkin's lymphoma in Europeans (appendix pp 20–22). Although both nasopharyngeal carcinoma and Hodgkin's lymphoma were very strongly associated with the MHC locus, the two associations were distinct from one another. Class I molecules, HLA-DQ, and HLA-DR were associated with nasopharyngeal carcinoma,¹⁷ whereas class I MHC molecules, HLA-DRA,^{50–52} and HLA-DPB2⁵³ have been found to be associated with Hodgkin's lymphoma. The absence of involvement of HLA-DP in patients with nasopharyngeal carcinoma from the same area in southern China as the patients with NKTCL in our study (which was strongly associated with HLA-DPB1) suggests that fundamental biological differences exist between nasopharyngeal carcinoma and NKTCL, despite the widely held view that both are mainly caused by EBV infection. Altogether, these distinct HLA signatures underlining the three malignancies suggest that these diseases are genetically heterogeneous and might have distinct molecular mechanisms of pathogenesis beyond EBV infection. Alternatively, these distinct HLA signatures might all converge onto a single immunological phenotype—the inability to effectively prevent chronic EBV infection and EBV-driven carcinogenesis in a tissue-specific manner. Similarly, we examined publicly available GWAS results for other lymphoid malignancies, such as diffuse large B-cell lymphoma,^{18,39} marginal-zone lymphoma,⁴⁰ and follicular lymphoma and found that SNPs associated with these lymphomas are not associated with NKTCL.^{41–43} These data suggest that all four lymphoid malignancies have distinct genetic architectures, which is consistent with the histopathological finding that NKTCL is of natural killer cell lineage, whereas diffuse large B-cell lymphoma, marginal-zone lymphoma, and follicular lymphoma originate from B cells. This finding is also consistent with the observation that patients with NKTCL have more aggressive disease and poorer clinical outcome than those with B-cell malignancies.¹

We examined previously published results for other diseases and the publicly available GWAS catalogues (US National Human Genome Research Institute²⁷ and European Bioinformatics Institute), and found locus pleiotropy with HLA-DPB1 Asp84Gly and SNP markers that are in strong linkage disequilibrium ($r^2 > 0.8$) with it. First, the *HLA-DPB1*

Asp84 allele has been associated with a more than two times increased risk of chronic beryllium disease, an inflammatory and immune-mediated lung condition.³⁷ This allele is associated with a decreased risk of NKTCL in our study. Conversely, the *HLA-DPB1 Gly84* allele is associated with a more than two times reduced risk of chronic beryllium disease and is associated with an increased risk of NKTCL.³⁷ For other immune-related diseases, the A allele of SNP rs9277535, previously reported to be associated with improved clearance of hepatitis B virus infection⁵⁴ and increased susceptibility to inflammatory granulomatosis with polyangiitis,⁵⁵ shares strong linkage disequilibrium (pairwise $r^2=0.93$) with the NKTCL risk A allele at rs9277378. Reconciling both findings, we thus suspect that individuals with greater hepatitis B virus clearance ability would have a higher risk of NKTCL. Indeed, in our discovery cohort, the association between the *HLA-DPB1* rs9277378*A risk allele and increased risk of NKTCL was much stronger in patients with NKTCL who do not have hepatitis B virus infection, compared with patients with NKTCL who had concurrent hepatitis B virus infection. The pleiotropy is consistent with potential counterbalancing effects between determinants of tumour immune response, infectious disease susceptibility, autoimmunity, and inflammation.^{23,56,57} However, a more definitive dissection of the individual contribution of these determinants to the genetic architecture of *HLA-DPB1* could be complicated by balancing selection,⁵⁸ so further research is needed to explore these contributing forces in depth.

To our knowledge, our study is the first GWAS of NKTCL. We identified a strong association at four sequential aminoacid positions within HLA-DPB1, implicating antigen processing and presentation in disease pathogenesis. Our findings further suggest potential common pathways among some common diseases—eg, clearance of hepatitis B virus infection and inflammatory conditions. Additionally, our findings also lower the possibility that NKTCL and other diseases previously suspected to be triggered by the same pathogen—eg, nasopharyngeal carcinoma and Hodgkin's lymphoma associated with EBV—share the same susceptibility pathways. Further attention should now be focused on determining the precise mechanisms underlying the association between aminoacids 84–87 in HLA-DPB1 and NKTCL, which could improve rational prevention, vaccination, and treatment of the disease. Additionally, we were unable to confirm an association between NKTCL risk and *HLA-A*0201* as seen in a previous study,¹⁴ suggesting that this allele might not have an important role in NKTCL in the context of our study.

Our study has some limitations. First, statistical power calculations showed that our study only had sufficient power to validate SNPs surpassing genome-wide significance in the GWAS discovery stage (appendix p 9) and that it was underpowered to detect genetic effects with per-allele ORs of 1.8 or less. We fully acknowledge that there might be incidence of cancer in the control group over time. Reassuringly, because the incidence of NKTCL is very rare in general population (crude annual rate <1 in 100 000 population according to GLOBOCAN 2012),² the potential loss of power caused by the inadvertent inclusion of misclassified controls is also low. If misclassified individuals were included in the control group, this loss of power would have reduced statistical significance in our study. Second, our study was focused specifically on NKTCL and did not include other lymphoma types with similar epidemiological distribution—eg, hydroa vacciniforme-like lymphoma, aggressive natural killer cell leukaemia, and chronic active EBV infection. Although the sole

focus on NKTCL allowed for the clear detection of a genetic risk allele for this disease, whether the genetic risk conferred by the *HLA-DPB1* risk alleles reported here can be generalised to the other lymphoma types is unclear. Such assessments should form the basis of future investigations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all the participants in the study; staff members at the biobank of Sun Yat-Sen University Cancer Center for their generous contribution in preparing patient samples; staff members at the SingHealth Tissue Repository, Dharambir Sethi Singh Surinder (Novena ENT-Head and Neck Surgery Specialist Center) and Jia Chuan Juliana Chen (Tan Tock Seng Hospital, Singapore) for their generous contribution of patient samples; Dongxin Lin (Chinese Academy of Medical Sciences and Peking Union Medical College) for contribution of samples and thoughtful inputs in the study; Song Gao (Sun Yat-Sen University Cancer Center) for input in illustration and interpretation of the protein structure; and Su Qin Peh (Genome Institute of Singapore), Xiao Yin Chen (Genome Institute of Singapore), and Ting Feng (Chinese Academy of Medical Sciences and Peking Union Medical College) for genotyping support.

Funding Top-Notch Young Talents Program of China, Special Support Program of Guangdong, Specialized Research Fund for the Doctoral Program of Higher Education (20110171120099), Program for New Century Excellent Talents in University (NCET-11-0529), National Medical Research Council of Singapore (TCR12DEC005), Tanoto Foundation Professorship in Medical Oncology, New Century Foundation Limited, Ling Foundation, Singapore National Cancer Centre Research Fund, and the US National Institutes of Health (1R01AR062886, 5U01GM092691-04, and 1R01AR063759-01A1).

References

1. Chan J, Quintanilla-Martinez L, Ferry J, Peh S. Extranodal NK/T-cell lymphoma, nasal type In: Swerdlow SH, Campo E, Harris NL, et al., eds. WHO classification of tumours of haematopoietic and lymphoid tissues, 4th edn. Geneva: World Health Organization, 2008: 285–88.
2. Lima M Aggressive mature natural killer cell neoplasms: from epidemiology to diagnosis. *Orphanet J Rare Dis* 2013; 8: 95. [PubMed: 23816348]
3. Adams SV, Newcomb PA, Shustov AR. Racial patterns of peripheral T-cell lymphoma incidence and survival in the United States. *J Clin Oncol* 2016; 34: 963–71. [PubMed: 26962200]
4. Yang Y, Zhu Y, Cao JZ, et al. Risk-adapted therapy for early-stage extranodal nasal-type NK/T-cell lymphoma: analysis from a multicenter study. *Blood* 2015; 126: 1424–32. [PubMed: 26109206]
5. Kwong YL, Kim WS, Lim ST, et al. SMILE for natural killer/T-cell lymphoma: analysis of safety and efficacy from the Asia Lymphoma Study Group. *Blood* 2012; 120: 2973–80. [PubMed: 22919026]
6. Quintanilla-Martinez L, Kremer M, Keller G, et al. p53 mutations in nasal natural killer/T-cell lymphoma from Mexico: association with large cell morphology and advanced disease. *Am J Pathol* 2001; 159: 2095–105. [PubMed: 11733360]
7. Ye Z, Cao Q, Niu G, et al. p63 and p53 expression in extranodal NK/T cell lymphoma, nasal type. *J Clin Pathol* 2013; 66: 676–80. [PubMed: 23626008]
8. Koo GC, Tan SY, Tang T, et al. Janus kinase 3-activating mutations identified in natural killer/T-cell lymphoma. *Cancer Discov* 2012; 2: 591–97. [PubMed: 22705984]
9. Jiang L, Gu ZH, Yan ZX, et al. Exome sequencing identifies somatic mutations of *DDX3X* in natural killer/T-cell lymphoma. *Nat Genet* 2015; 47: 1061–66. [PubMed: 26192917]
10. Huang Y, de Reynies A, de Leval L, et al. Gene expression profiling identifies emerging oncogenic pathways operating in extranodal NK/T-cell lymphoma, nasal type. *Blood* 2010; 115: 1226–37. [PubMed: 19965620]

11. Iqbal J, Kucuk C, Deleeuw RJ, et al. Genomic analyses reveal global functional alterations that promote tumor growth and novel tumor suppressor genes in natural killer-cell malignancies. *Leukemia* 2009; 23: 1139–51. [PubMed: 19194464]
12. Karube K, Nakagawa M, Tsuzuki S, et al. Identification of *FOXO3* and *PRDM1* as tumor-suppressor gene candidates in NK-cell neoplasms by genomic and functional analyses. *Blood* 2011; 118: 3195–204. [PubMed: 21690554]
13. Xu JX, Hoshida Y, Yang WI, et al. Life-style and environmental factors in the development of nasal NK/T-cell lymphoma: a case-control study in East Asia. *Int J Cancer* 2007; 120: 406–10. [PubMed: 17066445]
14. Kanno H, Kojya S, Li T, et al. Low frequency of *HLA-A*0201* allele in patients with Epstein-Barr virus-positive nasal lymphomas with polymorphic reticulosis morphology. *Int J Cancer* 2000; 87: 195–99. [PubMed: 10861473]
15. Li S, Qian J, Yang Y, et al. GWAS identifies novel susceptibility loci on 6p21.32 and 21q21.3 for hepatocellular carcinoma in chronic hepatitis B virus carriers. *PLoS Genet* 2012; 8: e1002791. [PubMed: 22807686]
16. Vithana EN, Khor CC, Qiao C, et al. Genome-wide association analyses identify three new susceptibility loci for primary angle closure glaucoma. *Nat Genet* 2012; 44: 1142–46. [PubMed: 22922875]
17. Bei JX, Li Y, Jia WH, et al. A genome-wide association study of nasopharyngeal carcinoma identifies three new susceptibility loci. *Nat Genet* 2010; 42: 599–603. [PubMed: 20512145]
18. Tan DE, Foo JN, Bei JX, et al. Genome-wide association study of B cell non-Hodgkin lymphoma identifies 3q27 as a susceptibility locus in the Chinese population. *Nat Genet* 2013; 45: 804–07. [PubMed: 23749188]
19. Aung T, Ozaki M, Mizoguchi T, et al. A common variant mapping to *CACNA1A* is associated with susceptibility to exfoliation syndrome. *Nat Genet* 2015; 47: 387–92. [PubMed: 25706626]
20. Khor CC, Chau TN, Pang J, et al. Genome-wide association study identifies susceptibility loci for dengue shock syndrome at *MICB* and *PLCE1*. *Nat Genet* 2011; 43: 1139–41. [PubMed: 22001756]
21. Chen Y, Lin Y, Vithana EN, et al. Common variants near *ABCA1* and in *PMM2* are associated with primary open-angle glaucoma. *Nat Genet* 2014; 46: 1115–19. [PubMed: 25173107]
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; 81: 559–75.23 [PubMed: 17701901]
23. Dunstan SJ, Hue NT, Han B, et al. Variation at *HLA-DRB1* is associated with resistance to enteric fever. *Nat Genet* 2014; 46: 1333–36. [PubMed: 25383971]
24. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013; 8: e64683. [PubMed: 23762245]
25. Okada Y, Kim K, Han B, et al. Risk for ACPA-positive rheumatoid arthritis is driven by shared HLA amino acid polymorphisms in Asian and European populations. *Hum Mol Genet* 2014; 23: 6916–26. [PubMed: 25070946]
26. Pillai NE, Okada Y, Saw WY, et al. Predicting *HLA* alleles from high-resolution SNP data in three Southeast Asian populations. *Hum Mol Genet* 2014; 23: 4443–51. [PubMed: 24698974]
27. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014; 42 (Database issue): D1001–06. [PubMed: 24316577]
28. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; 273: 1516–17. [PubMed: 8801636]
29. Kiryluk K, Li Y, Scolari F, et al. Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nat Genet* 2014; 46: 1187–96. [PubMed: 25305756]
30. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19: 149–50. [PubMed: 12499305]
31. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; 38: 904–09. [PubMed: 16862161]

32. Bailey JN, Loomis SJ, Kang JH, et al. Genome-wide association analysis identifies *TXNRD2*, *ATXN2* and *FOXC1* as susceptibility loci for primary open-angle glaucoma. *Nat Genet* 2016; 48: 189–94. [PubMed: 26752265]
33. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–60. [PubMed: 12958120]
34. Nalls MA, Plagnol V, Hernandez DG, et al., for the International Parkinson Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* 2011; 377: 641–49. [PubMed: 21292315]
35. International HIVCS, Pereyra F, Jia X, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 2010; 330: 1551–57. [PubMed: 21051598]
36. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012; 44: 291–96. [PubMed: 22286218]
37. Silveira LJ, McCanlies EC, Fingerlin TE, et al. Chronic beryllium disease, HLA-DPB1, and the DP peptide binding groove. *J Immunol* 2012; 189: 4014–23. [PubMed: 22972925]
38. Dai S, Murphy GA, Crawford F, et al. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proc Natl Acad Sci USA* 2010; 107: 7425–30. [PubMed: 20356827]
39. Cerhan JR, Berndt SI, Vijai J, et al. Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma. *Nat Genet* 2014; 46: 1233–38. [PubMed: 25261932]
40. Vijai J, Wang Z, Berndt SI, et al. A genome-wide association study of marginal zone lymphoma shows association to the HLA region. *Nat Commun* 2015; 6: 5751. [PubMed: 25569183]
41. Skibola CF, Berndt SI, Vijai J, et al. Genome-wide association study identifies five susceptibility loci for follicular lymphoma outside the *HLA* region. *Am J Hum Genet* 2014; 95: 462–71. [PubMed: 25279986]
42. Skibola CF, Bracci PM, Halperin E, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. *Nat Genet* 2009; 41: 873–75. [PubMed: 19620980]
43. Conde L, Halperin E, Akers NK, et al. Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet* 2010; 42: 661–64. [PubMed: 20639881]
44. Klein J, Sato A. The HLA system. First of two parts. *N Engl J Med* 2000; 343: 702–09. [PubMed: 10974135]
45. Diaz G, Amicosante M, Jaraquemada D, et al. Functional analysis of HLA-DP polymorphism: a crucial role for DPbeta residues 9, 11, 35, 55, 56, 69 and 84–87 in T cell allorecognition and peptide binding. *Int Immunol* 2003; 15: 565–76. [PubMed: 12697658]
46. Castelli FA, Buhot C, Sanson A, et al. HLA-DP4, the most frequent HLA II molecule, defines a new supertype of peptide-binding specificity. *J Immunol* 2002; 169: 6928–34. [PubMed: 12471126]
47. Patronov A, Dimitrov I, Flower DR, Doytchinova I. Peptide binding prediction for the human class II MHC allele *HLA-DP2*: a molecular docking approach. *BMC Struct Biol* 2011; 11: 32. [PubMed: 21752305]
48. Sethi DK, Schubert DA, Anders AK, et al. A highly tilted binding mode by a self-reactive T cell receptor results in altered engagement of peptide and MHC. *J Exp Med* 2011; 208: 91–102. [PubMed: 21199956]
49. Aozasa K, Takakuwa T, Hongyo T, Yang WI. Nasal NK/T-cell lymphoma: epidemiology and pathogenesis. *Int J Hematol* 2008; 87: 110–17. [PubMed: 18256789]
50. Cozen W, Timofeeva MN, Li D, et al. A meta-analysis of Hodgkin lymphoma reveals 19p13.3 TCF3 as a novel susceptibility locus. *Nat Commun* 2014; 5: 3856. [PubMed: 24920014]
51. Enciso-Mora V, Broderick P, Ma Y, et al. A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3). *Nat Genet* 2010; 42: 1126–30. [PubMed: 21037568]
52. Urayama KY, Jarrett RF, Hjalgrim H, et al. Genome-wide association study of classical Hodgkin lymphoma and Epstein-Barr virus status-defined subgroups. *J Natl Cancer Inst* 2012; 104: 240–53. [PubMed: 22286212]

53. Delahaye-Sourdeix M, Urayama KY, Gaborieau V, et al. A novel risk locus at 6p21.3 for Epstein-Barr virus positive Hodgkin lymphoma. *Cancer Epidemiol Biomarkers Prev* 2015; 24: 1838–43. [PubMed: 26404960]
54. Kamatani Y, Wattanapokayakit S, Ochi H, et al. A genome-wide association study identifies variants in the *HLA-DP* locus associated with chronic hepatitis B in Asians. *Nat Genet* 2009; 41: 591–95. [PubMed: 19349983]
55. Xie G, Roshandel D, Sherva R, et al. Association of granulomatosis with polyangiitis (Wegener's) with HLA-DPB1*04 and SEMA6A gene variants: evidence from genome-wide analysis. *Arthritis Rheum* 2013; 65: 2457–68. [PubMed: 23740775]
56. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. *Nat Rev Genet* 2014; 15: 379–93. [PubMed: 24776769]
57. Sfriso P, Ghirardello A, Botsios C, et al. Infections and autoimmunity: the multifaceted relationship. *J Leukoc Biol* 2010; 87: 385–95. [PubMed: 20015961]
58. de Bakker PI, McVean G, Sabeti PC, et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006; 38: 1166–72. [PubMed: 16998491]

Research in context

Evidence before this study

Extranodal natural killer T-cell lymphoma (NKTCL), nasal type, is a rare and distinct malignancy with a severely aggressive clinical course. Because of its rare and aggressive clinical presentation and the regional prevalence in Asia and Latin America, hereditary predisposition was thought to contribute to individual susceptibility to NKTCL. Epidemiological findings suggest an infectious trigger to NKTCL, with Epstein-Barr virus infection being the most likely cause. However, no germline genetic marker has been reliably linked to NKTCL susceptibility at stringent statistical thresholds.

Added value of this study

To our knowledge, our study is the first genome-wide association study of NKTCL. Our data suggest common genetic variation at *HLA-DPB1* as a strong contributor to the disease. On average, each copy of the risk allele increases the risk of NKTCL by 1.84 times compared with the baseline wild-type genotype.

Implications of all the available evidence

Results from this study increase our understanding of the risk factors and causes of NKTCL by revealing how the human body could potentially interact with tumour antigens. Despite improvements in NKTCL therapy, many patients have disease relapse. Our findings might offer insights in efforts to improve immunotherapies aimed at this rare form of lymphoma.

Figure. Association of SNPs within the broad *HLA* region with NKTCL
 (A) Unconditional analysis. (B) Conditional analysis on HLA-DPB1 aminoacid position Gly84. The chromosome position was based on the US National Center for Biotechnology Information human genome build 36.

NKTCL=natural killer T-cell lymphoma. OR=odds ratio. SNP=single-nucleotide polymorphism.

Table 1:
Association results between *HLA-DPB1* rs9277378 (A/G) and NKTCL, by study cohort

	Chinese population	NKTCL patients						Controls						p	Odds ratio (95% CI)	P _{het} (I ²)	
			Collection region	n	Male	Female	Age, years	RAF	Collection region	n	Male	Female	Age, years				RAF
Discovery	Southern		Guangdong	189	134 (71%)	55 (29%)	47 (33–57)	0.49	Guangdong	957	727 (76%)	230 (24%)	34 (27–43)	0.29	5.84 × 10 ^{−16}	2.65 (2.08–3.37)	..
Validation																	..
Replication 1	Southern		Guangdong	75	46 (61%)	29 (39%)	45 (28–63)	0.43	Hong Kong	296	118 (40%)	178 (60%)	73 (68–79)	0.32	0.010	1.61 (1.12–2.33)	..
Replication 2	Southern		Guangdong	65	46 (71%)	19 (29%)	52 (41–60)	0.45	Guangdong	983	718 (73%)	265 (27%)	48 (39–56)	0.32	0.0045	1.73 (1.22–2.46)	..
Replication 3	Northern		Beijing	125	83 (66%)	42 (34%)	41 (33–55)	0.55	Beijing	1110	655 (59%)	455 (41%)	22 (20–24)	0.45	0.00077	1.52 (1.17–1.98)	..
Replication 4	Southern*		Singapore	60	40 (67%)	20 (33%)	52 (44–64)	0.45	Singapore	2476	1263 (51%)	1213 (49%)	56 (49–62)	0.36	0.033	1.45 (1.03–2.05)	..
Validation sum	..	..	..	325	..	..	..	..	..	4865	..	..	..	..	5.78 × 10 ^{−8}	1.56 (1–33–1.84)	0.86 (0.0%)
Overall meta-analysis (all samples)	..	..	..	514	..	..	..	..	..	5822	..	..	..	..	4.21 × 10 ^{−19}	1.84 (1.61–2.11)	0.011 (69.4%)

Data are n, n (%), or median (IQR), unless otherwise indicated. NKTCL=natural killer T-cell lymphoma. RAF=risk allele frequency. P_{het}=p value for heterogeneity.

* Information on the participants' native region (defined as region self-reported by their parents) was collected.¹⁹

Table 2:

Associations between haplotypes at aminoacid positions 84–87 of HLA-DPB1 and NKTCL

	rs9277378 allele	Aminoacids at positions 84–87	Frequency in NKTCL cases	Frequency in controls	Odds ratio	P _{hap}
<i>*0201, *0401</i>	A	Gly84-Gly85-Pro86-Met87	0.460	0.264	2.38	2.32×10^{-14}
<i>*0201, *0401</i>	NA	Gly84-Gly85-Pro86-Met87	0.463	0.264	2.40	1.01×10^{-14}
<i>*0201</i> [†]	NA	NA	0.191	0.131	1.90	0.00017
<i>*0401</i> [†]	NA	NA	0.0939	0.0496	2.10	0.00072
<i>*0501</i>	A	Asp84-Glu85-Ala86-Val87	0.037	0.025	1.50	0.19
<i>*0501</i>	G	Asp84-Glu85-Ala86-Val87	0.503	0.711	0.41	2.60×10^{-15}
<i>*0501</i>	NA	Asp84-Glu85-Ala86-Val87	0.537	0.736	0.42	1.01×10^{-14}
<i>*0501</i> [†]	NA	NA	0.303	0.498	0.41	4.97×10^{-13}

NKTCL=natural killer T-cell lymphoma. P_{hap}=p value for haplotype. NA=not applicable—association analysis was irrespective of the allele at the given location.

[†] Association analysis of the classic *HLA-DPB1* alleles (**0201*, **0401*, and **0501*) is also shown, because the aminoacid haplotypes also partly define these classic alleles.