Title

Two HLA class II gene variants are independently associated with pediatric osteosarcoma risk

Journal

Cancer Epidemiology Biomarkers & Prevention, 27(10)

ISSN

1055-9965

Authors

Zhang, Chenan

Wiemels, Joseph L

Hansen, Helen M

et al.

Publication Date

2018-10-01

DOI

10.1158/1055-9965.epi-18-0306

Peer reviewed

HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2019 October 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2018 October ; 27(10): 1151–1158. doi: 10.1158/1055-9965.EPI-18-0306.

Two HLA class II gene variants are independently associated with pediatric osteosarcoma risk

Chenan Zhang[1,2], **Joseph L. Wiemels**[1,2], **Helen M. Hansen**[2], **Julio Gonzalez-Maya**[2], **Alyson A. Endicott**[2], **Adam J. de Smith**[1], **Ivan V. Smirnov**[2], **John S. Witte**[1], **Libby M. Morimoto**[3], **Catherine Metayer**[3], and **Kyle M. Walsh**[1,4,5,*]

[1]Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA 94158;

[2]Division of Neuroepidemiology, Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA 94158;

[3]School of Public Health, University of California, Berkeley, Berkeley, CA 94720;

[4]Division of Neuro-epidemiology, Department of Neurosurgery, Duke University, Durham, NC 27710;

[5]Children's Health and Discovery Institute, Duke University

Abstract

Background: The genetic etiology of osteosarcoma remains poorly understood despite the publication of a genome-wide association study. Association between human leukocyte antigen (HLA) gene variants and risk of several cancers has been observed, but HLA variation is not well-captured by standard SNP arrays.

Methods: We genotyped 207 Californian pediatric osteosarcoma cases and 696 controls of European ancestry using a custom genome-wide array supplemented with ~6,000 additional probes across the MHC region. We subsequently imputed four-digit classical HLA alleles using a reference panel of 5,225 individuals who underwent high-resolution HLA typing via next-generation sequencing. Case-control comparisons were adjusted for ancestry-informative principal components, and top associations from the discovery analysis underwent replication in an independent dataset of 657 cases and 1183 controls.

Results: Three highly correlated HLA class II variants (r^2=0.33–0.98) were associated with osteosarcoma risk in discovery analyses, including HLA-DRB1*0301 (OR=0.52; P=3.2×10^{-3}), HLA-DQA1*0501 (OR=0.74; P=0.031), and HLA-DQB1*0201 (OR=0.51; P=2.7×10^{-3}). Similar associations were observed in the replication data (P_{range}=0.011–0.037). Meta-analysis of the two datasets identified HLA-DRB1*0301 as the most significantly associated variant (OR_{meta}=0.62; P_{meta}=1.5×10^{-4}), reaching Bonferroni-corrected statistical significance. The meta-analysis also revealed a second significant independent signal at HLA-DQA1*01:01 (OR_{meta}=1.33,

*Corresponding Author: Dr. Kyle M. Walsh, Kyle.Walsh@Duke.edu, Phone: 919-684-8732.

Conflict of interest: The authors declare no potential conflicts of interest.

P_{meta}=1.2×10^{-3}), and a third suggestive association at HLA-DQB1*0302 (OR$_{meta}$=0.73, P_{meta}=6.4×10^{-3}).

Conclusions: Multiple independent HLA class II alleles may influence osteosarcoma risk.

Impact: Additional work is needed to extend our observations to other patient populations and to clarify the potential causal mechanisms underlying these associations. Understanding immunologic contributions to the etiology of osteosarcoma may inform rational therapeutic targets.

Introduction

Osteosarcoma is the most commonly diagnosed primary malignant bone tumor, with peak incidence occurring during adolescence. In the United States, approximately 400 new cases are diagnosed annually in individuals younger than 20 years of age(1), with similar rates of childhood and adolescent-onset osteosarcoma observed worldwide(2,3). Risk factors for osteosarcoma include taller stature and male sex(4–6), but the genetic etiology of the disease remains poorly understood. Although increased osteosarcoma risk is associated with Paget disease of the bone and inherited cancer predisposition syndromes (*e.g.* Li-Fraumeni syndrome), the majority of cases are sporadic(7,8). Candidate-gene studies have suggested a role for common genetic variants in the DNA repair pathway(9), growth hormone pathway(9,10) and telomere maintenance pathway(11) in contributing to risk of osteosarcoma, but only a single genome-wide association study (GWAS) has been published to-date(12). This GWAS reported two risk loci associated with osteosarcoma, one of which was located in the *GRM4* gene at 6p21.3, approximately 1 Mb from the major histocompatibility complex (MHC) class II region.

The MHC region, encompassing ~7.6Mb on chromosome 6p21, is one of the most polymorphic and gene-dense regions of the genome, including the highly variable human leukocyte antigen (HLA) genes. HLA variants have been associated with cancer risk in previous studies of chronic lymphocytic leukemia, lung squamous cell carcinoma, ovarian, breast, cervical, and nasopharyngeal cancers(13–18). Various mechanisms linking HLA genetic variation to cancer development have been hypothesized, including differential tumor tolerance by the immune system, efficiency of immunosurveillance, and cancer immunoediting resulting in certain HLA variants predisposing to particular malignancies(18–21). Investigating the role of heritable HLA variation in osteosarcoma risk may inform disease etiology, and inform potential strategies for immunotherapy.

Although the previous osteosarcoma GWAS did not report any signals in the MHC region, genetic variation and haplotype structure in HLA genes are not well-captured by standard genome-wide SNP arrays due to the unique characteristics of sequence and copy-number variation across the MHC region. Using a custom Affymetrix Axiom Array supplemented with ~6,000 additional probes spanning the MHC region, we typed classical HLA alleles at four-digit resolution *(e.g. HLA-DRB1*1501*) and tested for association with pediatric osteosarcoma risk. We subsequently attempted replication in an independent set of osteosarcoma cases and controls and performed meta-analysis of these two datasets, totaling 864 cases and 1879 controls. We further assessed non-additive genetic associations in the

MHC region as a complementary approach given the observation of widespread non-additive and epistatic effects among HLA variants for a number of diseases(22–24), as well as any potential modification by subject sex or clinical presentation (*e.g.* tumor location, presence of metastasis, age at diagnosis) to seek further insight into the etiology, progression, and prognosis of this aggressive cancer.

Materials and methods

Discovery sample:

The study was approved by the Institutional Review Boards at the Universities of California at Berkeley and San Francisco, and the California Department of Public Health (CDPH). The biospecimens and/or data used in this study were obtained from the California Biobank Program (SIS request number 550). The CDPH, Genetic Diseases Screening Branch obtains newborn blood samples from all neonates born within the state for the purpose of disease screening. The bloodspots that remain after screening have been archived at −20°C since 1982 and made available for approved research. We linked statewide birth records maintained by the CDPH (for years 1982–2009) to cancer diagnosis data from the California Cancer Registry (CCR, for years 1988–2011). Included in this analysis were non-Hispanic white children born in California during 1982–2009 and diagnosed with osteosarcoma (ICD-O-3 codes 9180–9183, 9185–9187, and 9192–9195) by age nineteen, per CCR record. Children born in California during the same period and not reported to CCR as having any childhood cancer were selected as controls. Characteristics of osteosarcoma cases appear in Supplementary Table S1.

DNA extraction:

A one-third portion of a 12mm dried bloodspot was partitioned into three uniform segments and placed in a 2mL microcentrifuge tube prior to the extraction, which was carried out using the QIAamp DNA Investigator Kit (Qiagen). In brief, 280μL of Buffer ATL and 20μL of Proteinase K were added to each sample. Samples were vortexed and then incubated in a dry-bath shaker at 900rpm and 56°C for one hour. After incubation, samples were briefly centrifuged and the lysate solution was transferred to a new 2mL microcentrifuge tube, while the solid remnants were discarded. 1μL of 1ng/μL carrier RNA was added to the lysate and then briefly vortexed. After adding carrier RNA, samples were placed in the Qiagen Qiacube automated work station for DNA isolation. The result of the Qiacube extraction protocol was a purified DNA sample in ATE buffer.

Discovery sample genotyping and quality control:

DNA specimens were assigned to genotyping plates using blocked randomization according to case-control status, reported ethnicity, and sex. DNA was genotyped on the Affymetrix Axiom Latino Array(25), with high coverage of European ancestries, supplemented with ~6,000 additional SNP probes in the MHC region. DNA samples were genotyped on an Affymetrix TITAN system, and raw image files were processed with Affymetrix Genetools to call genotypes.

Duplicate samples included on the array (n=34) had average genotype concordance >99%. Call-rate filtering for SNPs and samples was performed iteratively as previously described(26), excluding SNPs with call-rates <97% and samples with call-rates <97%. SNPs displaying significant departure from Hardy-Weinberg equilibrium ($P<1.0\times10^{-5}$) among European-ancestry controls and samples with mismatched reported versus genotyped sex were excluded. One individual from each subject pair with identity-by-decent (IBD) proportion >0.18 was excluded in the discovery data (27,28). Using genome-wide SNP array data from 1184 HapMap Phase 3 samples, we performed principal component analysis using unlinked autosomal biallelic SNPs with allele frequency>0.05 and removed from analysis any sample showing evidence of non-European ancestry (>3 SDs from mean CEPH values on PCs 1–3).

Replication dataset:

The osteosarcoma replication dataset was obtained from dbGaP study accession phs000734.v1.p1 (A Genome-wide Association Study (GWAS) of Risk for Osteosarcoma) and phs000381.v1.p1 (eMERGE Geisinger eGenomic Medicine MyCode Project Controls). Cases and controls were genotyped on the Illumina OmniExpress array and underwent quality-control filtering as previously described(11). Briefly, SNPs with call rates <0.98 and subjects with genotyping call rates <0.97 were removed. Ancestry-informative principal components were calculated using Eigenstrat(29) and HapMap reference samples and mean values of the first five principal components were calculated among HapMap CEPH samples. Subjects with evidence of non-European ancestry (>3 SDs from the mean CEPH values) were excluded. SNPs with Hardy–Weinberg equilibrium $P<1.0\times10^{-5}$ among controls were removed. Osteosarcoma cases from the discovery and replication datasets were compared and both duplicated and cryptically-related (IBD>0.18) samples were excluded from the replication dataset (n=22 duplicated patients, 0 cryptically-related patients). A total of 657 non-overlapping European-ancestry cases and 1183 controls were included in the final replication dataset. Osteosarcoma patients from dbGaP were predominantly children and adolescents (aged < 21), although individual-level age data were unavailable, and the original publication did not restrict to a specific age range. These osteosarcoma patients are a subset of those included in a previous GWAS publication(12).

Imputation of HLA genotypes:

We imputed four-digit classical HLA alleles of the three main class I genes (HLA-A, HLA-B, and HLA-C) and the five class II genes (HLA-DRB1, -DPA1, -DPB1, -DQA1, and -DQB1), as well as 5,695 HLA intragenic SNPs using SNP2HLA(30), which uses a reference panel of 5,225 individuals of European ancestry from the Type 1 Diabetes Genetics Consortium who underwent high-resolution HLA typing via next-generation sequencing. We used the SNP2HLA imputed allele dosage data for the primary MHC-wide association analyses, and phased best guess genotypes for the non-additive effect analyses, as previously described(22). Association analyses were restricted to classic four-digit HLA class I and class II alleles with imputation quality score INFO>0.80, allele frequency>0.05 (19 class I alleles, 27 class II alleles). For non-additive effect analyses, we restricted to individuals who had two best-guess haplotypes at each HLA gene analyzed.

Statistical analyses:

We performed logistic regression analyses for each of the HLA class I and class II alleles, using the imputed allele dosages output from SNP2HLA and the "logistic" command in PLINK 1.9, adjusting for the first 10 ancestry-informative principal components generated by Eigenstrat. As a secondary exploratory investigation, we also performed association analyses for 5,695 imputed HLA intragenic SNPs in the MHC region. The analyses were performed separately in the discovery and replication datasets and meta-analyzed using the "meta-analysis" command in PLINK. We conducted conditional analyses by repeating the logistic regressions adjusting for the top signals. We also adjusted for SNP rs1906953, a previously identified osteosarcoma GWAS hit at 6p21.3 that is located ~1 Mb from the MHC region in order to assess any attenuation of the top signals present in our analyses. For sex-stratified analyses, we examined association signals separately in males and females and tested for heterogeneity of effect. We also conducted case-only analyses, assessing whether top HLA association signals identified in case-control analyses were associated with differences in clinical presentation, including: age at diagnosis, tumor location, tumor size, extension, differentiation, and presence of metastasis. Information on clinical variables were available for only the discovery dataset, coded as previously described(6).

We investigated non-additive effects of the HLA alleles by assessing dominance-effect models as described in prior studies(22,24). In brief, we tested the improvement in model fit comparing the additive effect model with a model that additionally included a dominance term representing heterozygous status for each HLA allele investigated. A P-value for model improvement from the $\chi 2$ test represents the significance of the deviation from an additive model. We additionally tested for interaction between top variants in each independent signal detected and each of the remaining class II variants.

HLA class I and II alleles that achieved nominal statistical significance in the discovery analysis ($P<0.05$) were carried forward for replication. Variants that achieved nominal significance in both datasets and which achieved a P-value surpassing Bonferroni correction in the meta-analysis were considered to be statistically significant. Due to linkage disequilibrium (LD) between HLA variants, a traditional Bonferroni correction threshold would be overly conservative as the tests are not independent. We estimated independent tests performed after adjusting for LD to correct for multiple comparisons(31), reducing the number of "effective" tests from 46 to 39 primary HLA class I and II allele analyses and from 5,695 to 1,713 secondary HLA intragenic SNP analyses. Corresponding Bonferroni corrections of significance thresholds were then performed using the "effective" number of tests for HLA class I and II alleles ($0.05/39=1.3\times10^{-3}$), and HLA intragenic SNPs ($0.05/1713=2.9\times10^{-5}$).

RESULTS

A total of 13,103 SNPs in the MHC region were successfully genotyped on-array in the discovery dataset, which included 207 pediatric osteosarcoma cases and 696 controls of European ancestry. Using these data to impute HLA class I and class II alleles, we observed 19 class I alleles and 27 class II alleles with frequency>0.05 (Supplementary Table S2). In logistic regression analyses, three class I alleles were nominally associated with

osteosarcoma risk, including: HLA-A*0201 (OR=1.35; 95% CI=1.05–1.74; P=0.020), HLA-A*0301 (OR=0.66; 95% CI=0.46–0.94; P=0.023), and HLA-B*0801 (OR=0.60; 95% CI=0.38–0.92; P=0.021) (Table 1). Additionally, three class II alleles were nominally associated with osteosarcoma risk, including HLA-DRB1*0301 (OR=0.52; 95% CI=0.34–0.81; $P=3.2\times10^{-3}$), HLA-DQA1*0501 (OR=0.74; 95% CI=0.56–0.97; P=0.031), and HLA-DQB1*0201 (OR=0.51; 95% CI=0.33–0.79; $P=2.7\times10^{-3}$) (Table 1).

In the replication dataset of 657 non-overlapping osteosarcoma cases and 1183 controls of European ancestry, a total of 5,401 SNPs in the MHC region were successfully genotyped on-array and used to impute four-digit HLA variants. None of the HLA class I variants identified in the discovery set were associated with osteosarcoma risk in the replication data (P_{range}=0.12–1.0). However, all three HLA class II variants identified in the discovery dataset showed evidence of association in the replication data (P_{range}=0.011–0.037) (Table 1).

Meta-analysis of the two studies identified HLA-DRB1*0301 as top signal (OR_{meta}=0.62; 95% CI=0.48–0.79; $P_{meta}=1.5\times10^{-4}$), reaching Bonferroni-corrected significance. Other class II variants identified in the discovery analysis (*i.e.* DQA1*0501 and DQB1*0201) had similar p-values and effect sizes (Table 2). We observed that subject genotype at these three class II genes (DQB1*0201, DRB1*0301, and DQA1*0501) were highly correlated (r^2=0.33–0.98) (Supplementary Table S3). In logistic regression analyses conditioned on genotype at DRB1*0301, neither DQA1*0501 nor DQB1*0201 remained significantly associated with osteosarcoma risk in the meta-analysis (P_{range}=0.35–0.97), indicating that these three alleles represent a single protective haplotype.

The meta-analysis also revealed another association signal at three additional class II alleles: DRB1*0101 (OR_{meta}=1.37; 95% CI=1.10–1.71; $P_{meta}=4.9\times10^{-3}$), DQA1*0101 ($OR_{meta}$=1.33; 95% CI=1.12–1.58; $P_{meta}=1.2\times10^{-3}$), and DQB1*0501 ($OR_{meta}$=1.34; 95% CI=1.11–1.61; $P_{meta}=2.5\times10^{-3}$). High LD between these three alleles (r^2=0.57–0.99), together with conditional analyses showing attenuation of signals suggest that they represent a single risk haplotype. Conditional analyses adjusting for HLA-DRB1*0301 located on the protective haplotype described above did not substantially attenuate the associations with these three alleles ($P_{range}= 6.3\times10^{-3}$ - 0.014), suggesting that they represent a second independent signal (Supplementary Figure S1).

We observed a third independent risk locus at DQB1*0302 that was suggestively associated with osteosarcoma (OR_{meta}=0.73; 95% CI=0.58–0.91; $P_{meta}=6.4\times10^{-3}$). Conditional analyses adjusting for DRB1*0301 (the top signal in the protective haplotype) and DQA1*0101 (the top signal in the risk haplotype) did not substantially attenuate the association at DQB1*0302 ($P_{conditional}$=0.01) (Supplementary Figure S1). Additionally, all three HLA class II association signals remained when adjusting for rs1906953, the *GRM4* variant on 6p21.3 previously identified as an osteosarcoma risk locus(12) (Supplementary Figure S2), which showed no LD with any of our HLA class II risk variants (r^2<0.01).

Although the replication dataset genotyping array did not contain the additional ~6,000 probes for the MHC region included on the discovery dataset custom array, imputation

accuracy was similarly high in both discovery and replication datasets for the nominally significant osteosarcoma risk loci reported above (Supplementary Table S4).

Sex-specific analyses of the combined discovery and replication data suggested that the effect of the third HLA class II signal, DQB1*0302, may be modified by subject sex ($P_{interaction}$=0.02). Specifically, we observed a greater magnitude of effect in females (n=397; OR_{meta}=0.54; P_{meta}=2.4×10^{-3}) than in males (n=506; OR_{meta}=0.90; P_{meta}=0.49). No differences across strata of subject sex were observed for the other association signals. In case-only analyses of the discovery sample where clinical data were available, the lead HLA variant of the top signal (DQB1*0201) was not significantly associated with clinical features, although it was suggestively associated with higher tumor grade (P_{trend}=0.09).

A systematic search for non-additive effects in the region identified a departure from additivity at HLA-DQB1*0302. Inclusion of a dominance term improved the model fit in both the discovery and the replication datasets ($P_{improve}$=0.04 and $P_{improve}$=0.03, respectively). The dominance term (OR=0.48; 95% CI=0.25–0.92) suggests that for DQB1*0302, heterozygosity confers two-fold greater protection than expected with the effect observed in individuals homozygous for this allele. Given that DQB1*0302 had a female-specific association, we carried out sex-stratified non-additive analysis, observing the non-additive protective effect in male subjects ($P_{improve}$=0.09 and $P_{improve}$=6.4×10^{-3} for the discovery and replication datasets), but not female subjects ($P_{improve}$=0.23 and $P_{improve}$=0.51 for the discovery and replication datasets). Non-additive effects were not observed for any other DQB1 alleles (Supplementary Figure S3) or any other HLA class II alleles. Epistasis analyses of the top signal, represented by DRB1*0301, identified suggestive interactions with linked alleles DRB1*1301 ($P_{interaction}$=3.0×10^{-3}), DQA1*0103 ($P_{interaction}$=2.5×10^{-3}), and DQB1*0603 ($P_{interaction}$=4.8×10^{-3}), although these interactions were not significant after adjustment for all interactions tested (*i.e.* 0.05/26=1.9×10^{-3}) **(**Supplementary Table S5**)**. No evidence of interaction at the other top loci appearing in Table 2 was observed.

A previous study on the association between HLA variants and irritable bowel disease demonstrated that electrostatic potential differences at the peptide-binding groove correlate with the directionality of effect of DRB1 alleles on disease risk(23). We queried their published electrostatic potential data for DRB1*0301 and DRB1*0101, our two top signals showing opposing effects. We observed an electrostatic potential difference between these two alleles of approximately 0.8 on a scale ranging from 0 (identical) to 1 (maximum difference), indicating that the osteosarcoma risk allele (DRB1*0101) and protective allele (DRB1*0301) have substantially different electrostatic potentials, supporting a potential functional difference between these alleles.

In secondary analyses of imputed HLA intragenic SNPs, we observed several SNPs downstream of HLA-A showing evidence of association with osteosarcoma risk, including lead SNP rs2517612 (OR_{meta}=0.60, P_{meta}=5.5×10^{-5}). An additional independent signal was observed upstream of HLA-DRB1 at rs3129890 (OR_{meta}=0.71, P_{meta}=5.7×10^{-5}). The intragenic SNPs were in moderate LD with the top HLA class II signal DRB1*301 (r^2=0.41–0.42). However, neither SNP survived corrections for multiple testing (*i.e.*

$<2.9\times10^{-5}$). Association statistics, stratified by dataset, for SNPs with meta-analysis p-values $<1.0\times10^{-4}$ appear in Supplementary Table S6.

The California Department of Public Health is not responsible for the results or conclusions drawn by the authors of this publication.

DISCUSSION

In this first study to date of the association between HLA variation and osteosarcoma risk, we observed a signal associated with reduced risk consisting of the three linked alleles DQB1*0201, DRB1*0301, and DQA1*0501, which was replicated in an independent dataset and remained significant after Bonferroni correction. Two additional independent signals, one of reduced risk at DQB1*0302 and one of increased risk at three linked alleles DRB1*0101, DQA1*0101, and DQB1*0501, were revealed in meta-analysis of the discovery and replication datasets. None of the three signals were attenuated in conditional analyses adjusting for rs1906953, indicating that the effects are independent of a previously identified GWAS signal in *GRM4* at 6p21.3.

A growing body of literature has emerged on the complex interactions between bone and immune cells, with shared signal transduction pathways, common hematopoietic precursors, and cross-talk between the bone and the immune systems(32). The expression of HLA class II molecules on human osteoblasts has been observed in multiple studies(33,34), as has upregulation of the HLA class II receptor activity pathway in osteosarcoma(35), supporting a biological role for HLA class II genes in osteosarcoma development and progression. Indeed, in addition to immunotherapy targets currently in clinical development for patients with osteosarcoma (*e.g.* PD1, PDL1), HLA class II molecules, IL-10, and TGFβ are currently considered potential targets for breaking immune tolerance to the tumor(36).

A recent study found an association between variation in immune cell transcripts, specifically in monocytes and T cells, and osteosarcoma metastasis and survival(37). A role for immune cell infiltration and immune monitoring in bone is consistent with our results suggesting potential immunological contributions to osteosarcoma etiology, potentially operating through differences in antigen presentation. Although we were unable to assess patient survival, a suggestive association between DQB1*0201 and tumor grade suggests that germline variation in immune-related genes may influence patient outcomes. This complements a recent GWAS which identified an allele associated with reduced IL-33 expression and poorer osteosarcoma patient survival(38). IL-33 induces production of Th2-associated cytokines and is associated with atopic conditions(39,40). Because it is expressed by both osteoblasts and lymphocytes, the interaction of local and systemic immunologic changes in impacting osteosarcoma risk and outcome and interplay between IL-33 and HLA class II variation merit further attention.

Studies of other tumor sites have also revealed the involvement of the HLA alleles highlighted in our results, raising an interesting possibility of this variant as a source of genetic pleiotropy. Notably, the extended haplotype DRB1*0301:DQA1*0501:DQB1*0201 corresponding to our top association signal was previously associated with increased risk of

ovarian cancer(19). Increased risk of nasopharyngeal carcinoma, associated with the Epstein-Barr virus, was found with the extended haplotype HLA-A*3303-B*5801/2-DRB1*0301-DQB1*0201/2-DPB1*0401, specific to the Chinese ethnic group of the Taiwanese study(17). Previous examples of pleiotropic associations within the MHC region have been described among B-cell malignancies, suggesting the possibility of shared biological pathways influencing the development of different cancer types(41). Furthermore, opposing risk associations in different B-cell malignancies have previously been noted for pleotropic risk loci in the MHC region(41), similar to our observation of protection against osteosarcoma compared to increased risk for other cancer type for the same HLA variant. Individual alleles, DRB1*0101, DQA1*0101, DQB1*0501 of our second signal have also been linked to a variety of cancers including renal cell carcinoma, breast cancer, cervical cancer, and childhood acute lymphoblastic leukemia(42–46). Our third signal, DQB1*0302, has been linked to HPV-related cervical cancer, hypothesized to be due to differential affinity to HPV antigen binding associated with protection against infection(47).

Exploratory analyses of HLA intragenic SNPs identified lead SNP rs2517612, an expression quantitative trait locus (eQTL) for a large number of genes, including: *BTN3A2, SLC15A3, FAM20B, HLA-B, HLA-C, HLA-DQA1, HLA-DQA2*, among others. Due to the large number of HLA intragenic SNPs analyzed, no single-SNP association survived Bonferroni correction.

The potential interaction between DQB1*0302 and subject sex is intriguing given the well-documented increased incidence of osteosarcoma in males versus females(5). The larger and more significant decrease in risk for osteosarcoma seen in females with the DQB1*0302 allele than males, despite a smaller sample size, is consistent with this variant potentially contributing to lower incidence of osteosarcoma in girls. It is unclear why the association would be stronger in females, although sex-specific HLA associations were previously observed for gastric cancer and acute lymphoblastic leukemia(48,49). One possibility may relate to sex-differences in rates of bone growth during adolescence, previously suggested to impact osteosarcoma risk due to the susceptibility of rapidly proliferating cells to oncogenic agents and mitotic errors(50). However, the observed male-specific non-additive protective effect complicates interpretations and suggests a role for DRB1*0302 in osteosarcoma susceptibility across sexes. Since this HLA variant did not remain significant after Bonferroni correction, additional follow-up studies are necessary to confirm the association and explore any effect modification by sex.

The DQB1*0302 association displayed a non-additive dominance effect in both the discovery and replication datasets, implying a stronger protective effect in heterozygotes than homozygotes. Non-additive effects in the MHC region are hypothesized to play an important role in the necessary functional redundancy of the immune system(51), and given the overlapping signal transduction pathways and cross-talk between the bone and immune system, the detection of such effects in relation to osteosarcoma are intriguing.

Because our genotyping data were generated with custom SNP arrays rather than direct typing of HLA alleles via targeted sequencing, our genotyping is limited by imputation accuracy. However, imputation error would only lead to reduced power to detect signals and

should not lead to any increase in Type I error rate. In addition to rare variants, HLA imputation may also fail to detect variants in paralog genes and copy-number variable regions(52), as such variants could be excluded from analyses due to deviation from Hardy-Weinberg equilibrium. Higher-resolution HLA sequencing will be necessary to more conclusively identify the causative variants underlying our observed associations, potentially enabling specific protective HLA variants to guide approaches for breaking immune tolerance to the tumor.

The array-based platform utilized also limits our ability to detect rare or private germline mutations in genes associated with hereditary cancer predisposition syndromes, including *TP53*. While previous studies have not observed strong associations between common *TP53* variants and osteosarcoma risk(53), a recent study found that 9.5% of osteosarcoma patients diagnosed before age 30 harbored at least one rare exonic *TP53* variant(7). Given recent reports of a higher-than-expected population prevalence of TP53 mutations in controls unselected for cancer history(54), and potential polygenic contributions from multiple rare variants to sarcoma risk(55), the role of MHC class II variants and other immune factors in modifying cancer penetrance among *TP53* mutation carriers should be explored.

Our study has several additional limitations. Our analyses were performed in subjects of European ancestry and findings may not be generalizable to other ethnicities. The SNP2HLA method requires an ethnic-specific reference panel and non-European reference datasets of adequate sample size are currently lacking(30). We also cannot confidently ascertain the likeliest causative allele among multiple linked HLA variants in our top two signals, although our query of the electrostatic properties of DRB1 variants suggests that the DRB1 alleles of the two signals are strong candidates for functional relevance. However, additional functional studies are necessary to infer causality. Finally, although our two top signals remained significant after LD-adjusted Bonferroni-correction for multiple testing, additional datasets will be necessary to confirm our findings.

In summary, we genotyped HLA class I and class II alleles at 4-digit resolution and observed multiple independent class II alleles associated with risk of pediatric osteosarcoma, including two different alleles of DRB1. These findings improve our understanding of potential immunological contributions to osteosarcoma etiology and may also provide insights for development of future immunotherapies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

Funding for the MyCode® Project was provided by a grant from Commonwealth of Pennsylvania and the Clinic Research Fund of Geisinger Clinic. Funding support for the genotyping of the MyCode® cohort was provided by a Geisinger Clinic operating funds and an award from the Clinic Research Fund. This work was supported by the National Institutes of Health T32CA151022–06 (C. Zhang), R25T CA112355 (J. S. Witte), R01 CA155461 (J. L. Wiemels), and by ‘A’ Awards from The Alex’s Lemonade Stand Foundation (A. J. deSmith., K. M. Walsh).

References

1. Gurney J, Swensen A, Bulterys M. Malignant bone tumors Cancer Incid Surviv Child Adolesc U S SEER Program 1975–1995 Bethesda MD. Bethesda, MD;
2. Bridge J, Hogendoorn P, Bridge JA, Fletcher CD. WHO classification of tumours of soft tissue and bone. International Agency for Research on Cancer; 2013.
3. Stiller CA. International patterns of cancer incidence in adolescents. Cancer Treat Rev. 2007;33:631–45. [PubMed: 17329031]
4. Mirabello L, Pfeiffer R, Murphy G, Daw NC, Patiño-Garcia A, Troisi RJ, et al. Height at diagnosis and birth-weight as risk factors for osteosarcoma. Cancer Causes Control CCC. 2011;22:899–908. [PubMed: 21465145]
5. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: Data from the Surveillance, Epidemiology, and End Results Program. Cancer. 2009;115:1531–43. [PubMed: 19197972]
6. Endicott AA, Morimoto LM, Kline CN, Wiemels JL, Metayer C, Walsh KM. Perinatal factors associated with clinical presentation of osteosarcoma in children and adolescents. Pediatr Blood Cancer. 2017;64:e26349–n/a.
7. Mirabello L, Yeager M, Mai PL, Gastier-Foster JM, Gorlick R, Khanna C, et al. Germline TP53 variants and susceptibility to osteosarcoma. J Natl Cancer Inst. 2015;107.
8. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2015;373:2336–46. [PubMed: 26580448]
9. Mirabello L, Yu K, Berndt SI, Burdett L, Wang Z, Chowdhury S, et al. A comprehensive candidate gene approach identifies genetic variation associated with osteosarcoma. BMC Cancer. 2011;11:209. [PubMed: 21619704]
10. Musselman JR, Bergemann TL, Ross JA, Sklar C, Silverstein KA, Langer EK, et al. Case-parent analysis of variation in pubertal hormone genes and pediatric osteosarcoma: a Children's Oncology Group (COG) study. Int J Mol Epidemiol Genet. 2012;3:286–93. [PubMed: 23205180]
11. Walsh KM, Whitehead TP, de Smith AJ, Smirnov IV, Park M, Endicott AA, et al. Common genetic variants associated with telomere length confer risk for neuroblastoma and other childhood cancers. Carcinogenesis. 2016;37:576–82. [PubMed: 27207662]
12. Savage SA, Mirabello L, Wang Z, Gastier-Foster JM, Gorlick R, Khanna C, et al. Genome-wide association study identifies two susceptibility loci for osteosarcoma. Nat Genet. 2013;45:799–803. [PubMed: 23727862]
13. Chaudhuri S, Cariappa A, Tang M, Bell D, Haber DA, Isselbacher KJ, et al. Genetic susceptibility to breast cancer: HLA DQB*03032 and HLA DRB1*11 may represent protective alleles. Proc Natl Acad Sci. 2000;97:11451–4. [PubMed: 11027344]
14. Castro FA, Haimila K, Sareneva I, Schmitt M, Lorenzo J, Kunkel N, et al. Association of HLA-DRB1, interleukin-6 and cyclin D1 polymorphisms with cervical cancer in the Swedish population--a candidate gene approach. Int J Cancer. 2009;125:1851–8. [PubMed: 19585495]
15. Gragert L, Fingerson S, Albrecht M, Maiers M, Kalaycio M, Hill BT. Fine-mapping of HLA associations with chronic lymphocytic leukemia in US populations. Blood. 2014;124:2657. [PubMed: 25232063]
16. Kohno T, Kunitoh H, Shimada Y, Shiraishi K, Ishii Y, Goto K, et al. Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. Carcinogenesis. 2010;31:834–41. [PubMed: 20061363]
17. Hildesheim A, Apple RJ, Chen C-J, Wang SS, Cheng Y-J, Klitz W, et al. Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan. J Natl Cancer Inst. 2002;94:1780–9. [PubMed: 12464650]
18. Bateman AC, Howell WM. Human leukocyte antigens and cancer: is it in our genes? J Pathol. 1999;188:231–6. [PubMed: 10419588]
19. Kübler K, Arndt PF, Wardelmann E, Krebs D, Kuhn W, van der Ven K. HLA-class II haplotype associations with ovarian cancer. Int J Cancer. 2006;119:2980–5. [PubMed: 17016821]
20. Mapara MY, Sykes M. Tolerance and Cancer: Mechanisms of Tumor Evasion and Strategies for Breaking Tolerance. J Clin Oncol. 2004;22:1136–51. [PubMed: 15020616]

21. Kim R, Emi M, Tanabe K. Cancer immunoediting from immune surveillance to immune escape. Immunology. 2007;121:1–14. [PubMed: 17386080]
22. Zhang C, de Smith AJ, Smirnov IV, Wiencke JK, Wiemels JL, Witte JS, et al. Non-additive and epistatic effects of HLA polymorphisms contributing to risk of adult glioma. J Neurooncol. 2017;
23. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. Nat Genet. 2015;47:172–9. [PubMed: 25559196]
24. Lenz TL, Deutsch AJ, Han B, Hu X, Okada Y, Eyre S, et al. Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. Nat Genet. 2015;47:1085–90. [PubMed: 26258845]
25. Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. Genomics. 2011;98:422–30. [PubMed: 21903159]
26. Wiemels JL, Walsh KM, de Smith AJ, Metayer C, Gonseth S, Hansen HM, et al. GWAS in childhood acute lymphoblastic leukemia reveals novel genetic associations at chromosomes 17q12 and 8q24.21. Nat Commun. 2018;9:286. [PubMed: 29348612]
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81:559–75. [PubMed: 17701901]
28. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015;4:7. [PubMed: 25722852]
29. 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;38:904–9. [PubMed: 16862161]
30. Jia X, Han B, Onengut-Gumuscu S, Chen W-M, Concannon PJ, Rich SS, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. PLOS ONE. 2013;8:e64683. [PubMed: 23762245]
31. Li M-X, Yeung JMY, Cherny SS, Sham PC. Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. Hum Genet. 2012;131:747–56. [PubMed: 22143225]
32. Terpos E, Dimopoulos MA. Interaction between the skeletal and immune systems in cancer: mechanisms and clinical implications. Cancer Immunol Immunother. 2011;60:305–17. [PubMed: 21243489]
33. Stanley KT, VanDort C, Motyl C, Endres J, Fox DA. Immunocompetent properties of human osteoblasts: interactions with T lymphocytes. J Bone Miner Res Off J Am Soc Bone Miner Res. 2006;21:29–36.
34. Skjødt H, Hughes DE, Dobson PR, Russell RG. Constitutive and inducible expression of HLA class II determinants by human osteoblast-like cells in vitro. J Clin Invest. 1990;85:1421–6. [PubMed: 2110190]
35. Cleton-Jansen A-M, Anninga JK, Briaire-de Bruijn IH, Romeo S, Oosting J, Egeler RM, et al. Profiling of high-grade central osteosarcoma and its putative progenitor cells identifies tumourigenic pathways. Br J Cancer. 2009;101:1909–18. [PubMed: 19888226]
36. Kansara M, Teng MW, Smyth MJ, Thomas DM. Translational biology of osteosarcoma. Nat Rev Cancer. 2014;14:722–35. [PubMed: 25319867]
37. Scott MC, Temiz NA, Sarver AE, LaRue RS, Rathe SK, Varshney J, et al. Comparative Transcriptome Analysis Quantifies Immune Cell Transcript Levels, Metastatic Progression, and Survival in Osteosarcoma. Cancer Res. 2018;78:326–37. [PubMed: 29066513]
38. Koster R, Panagiotou OA, Wheeler WA, Karlins E, Gastier-Foster JM, Caminada de Toledo SR, et al. Genome-wide association study identifies the GLDC/IL33 locus associated with survival of osteosarcoma patients. Int J Cancer. 2018;142:1594–601. [PubMed: 29210060]
39. Mirchandani AS, Salmond RJ, Liew FY. Interleukin-33 and the function of innate lymphoid cells. Trends Immunol. 2012;33:389–96. [PubMed: 22609147]

40. Hinds DA, McMahon G, Kiefer AK, Do CB, Eriksson N, Evans DM, et al. A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. Nat Genet. 2013;45:907–11. [PubMed: 23817569]
41. Law PJ, Sud A, Mitchell JS, Henrion M, Orlando G, Lenive O, et al. Genome-wide association analysis of chronic lymphocytic leukaemia, Hodgkin lymphoma and multiple myeloma identifies pleiotropic risk loci. Sci Rep. 2017;7:41071. [PubMed: 28112199]
42. Özdemir E, Kakehi Y, Nakamura E, Kinoshita H, Terachi T, Okada Y, et al. HLA-DRB1*0101 and *0405 as Protective Alleles in Japanese Patients with Renal Cell Carcinoma. Cancer Res. 1997;57:742. [PubMed: 9044854]
43. Mahmoodi M, Nahvi H, Mahmoudi M, Kasaian A, Mohagheghi M-A, Divsalar K, et al. HLA-DRB1,-DQA1 and -DQB1 Allele and Haplotype Frequencies in Female Patients with Early Onset Breast Cancer. Pathol Oncol Res. 2012;18:49–55. [PubMed: 21720852]
44. Maciag PC, Schlecht NF, Souza PSA, Franco EL, Villa LL, Petzl-Erler ML. Major Histocompatibility Complex Class II Polymorphisms and Risk of Cervical Cancer and Human Papillomavirus Infection in Brazilian Women. Cancer Epidemiol Biomark Amp Prev. 2000;9:1183.
45. Ivansson EL, Magnusson JJ, Magnusson PKE, Erlich HA, Gyllensten UB. MHC loci affecting cervical cancer risk: distinguishing the effects of HLA-DQB1 and non-HLA genes TNF, LTA, TAP1 and TAP2. Genes Immun. 2008;9:613–23. [PubMed: 18650831]
46. Taylor GM, Dearden S, Payne N, Ayres M, Gokhale DA, Birch JM, et al. Evidence that an HLA-DQA1-DQB1 haplotype influences susceptibility to childhood common acute lymphoblastic leukaemia in boys provides further support for an infection-related aetiology. Br J Cancer. 1998;78:561–5. [PubMed: 9744491]
47. Hildesheim A, Wang SS. Host and viral genetics and risk of cervical cancer: a review. Virus Res. 2002;89:229–40. [PubMed: 12445662]
48. Wu M-S, Hsieh R-P, Huang S-P, Chang Y-T, Lin M-T, Chang M-C, et al. Association of HLA-DQB1*0301 and HLA-DQB1*0602 with different subtypes of gastric cancer in Taiwan. Jpn J Cancer Res Gann. 2002;93:404–10. [PubMed: 11985790]
49. Dorak MT, Lawson T, Machulla HK, Darke C, Mills KI, Burnett AK. Unravelling an HLA-DR association in childhood acute lymphoblastic leukemia. Blood. 1999;94:694–700. [PubMed: 10397736]
50. Ottaviani G, Jaffe N. The Etiology of Osteosarcoma. In: Jaffe N, Bruland OS, Bielack S, editors. Pediatr Adolesc Osteosarcoma [Internet]. Boston, MA: Springer US; 2010 page 15–32. Available from: 10.1007/978-1-4419-0284-9_2
51. Rose AM, Bell LCK. Epistasis and immunity: the role of genetic interactions in autoimmune diseases. Immunology. 2012;137:131–8. [PubMed: 22804709]
52. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004;5:889–99. [PubMed: 15573121]
53. Savage SA, Burdett L, Troisi R, Douglass C, Hoover RN, Chanock SJ, et al. Germ-line genetic variation of TP53 in osteosarcoma. Pediatr Blood Cancer. 2007;49:28–33. [PubMed: 17096406]
54. de Andrade KC, Mirabello L, Stewart DR, Karlins E, Koster R, Wang M, et al. Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. Hum Mutat. 2017;38:1723–30. [PubMed: 28861920]
55. Ballinger ML, Goode DL, Ray-Coquard I, James PA, Mitchell G, Niedermayr E, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. Lancet Oncol. 2016;17:1261–71. [PubMed: 27498913]

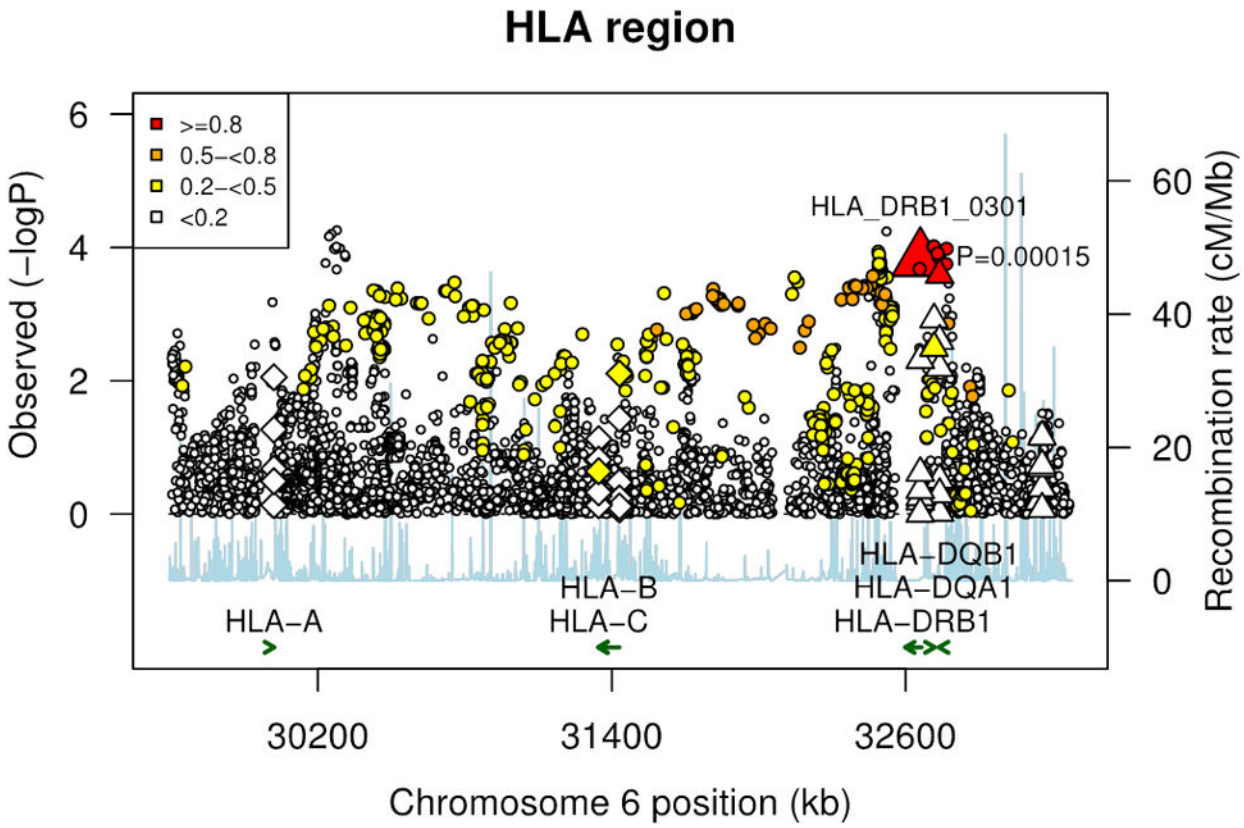

Figure 1: Regional association plot of the meta-analysis for HLA variants.

Diamonds represent HLA class I variants, triangles represent HLA class II variants, and circles represent HLA intragenic SNPs. The large triangle represents the top signal in DRB1*0301.

Table 1.

Association results (P<0.05) for discovery stage of 207 osteosarcoma cases and 696 controls, and replication stage of 657 cases and 1183 controls.

		Discovery			Replication		
Class	**HLA allele**	**Freq**[a]	**OR (95% CI)**	**P**	**Freq**[b]	**OR**	**P**
I	A*0201	0.26	1.35 (1.05, 1.74)	0.020	0.28	1.07(0.89, 1.28)	0.48
	A*0301	0.15	0.66 (0.46, 0.94)	0.023	0.14	1.00(0.79, 1.27)	1.0
	B*0801	0.11	0.60(0.38, 0.92)	0.021	0.10	0.75(0.52, 1.08)	0.12
II	DRB1*0301	0.12	0.52 (0.34, 0.81)	3.2×10^{-3}	0.12	0.67 (0.50, 0.91)	0.011
	DQA1*0501	0.25	0.74 (0.56, 0.97)	0.031	0.28	0.81 (0.67, 0.99)	0.037
	DQB1*0201	0.12	0.51 (0.33, 0.79)	2.7×10^{-3}	0.12	0.70 (0.52, 0.94)	0.018

[a]Allele frequency in discovery dataset controls

[b]Allele frequency in replication dataset controls

Table 2.

Meta-analysis results for HLA class II variants nominally associated with osteosarcoma (p<0.05), in 3 statistically independent loci.

HLA allele	$Freq_{discovery}$ [a]	OR_{meta} (95% CI)	P_{meta} [b]	I^2 [c] (%)	Signal
DRB1*0301	0.12	0.62 (0.48, 0.79)	**1.5×10^{-4}**	0	1
DQA1*0501	0.25	0.79 (0.67, 0.92)	3.2×10^{-3}	0	1
DQB1*0201	0.12	0.63 (0.49, 0.81)	**2.6×10^{-4}**	23.8	1
DRB1*0101	0.08	1.37 (1.10, 1.71)	4.9×10^{-3}	0	2
DQA1*0101	0.15	1.33 (1.12, 1.58)	**1.2×10^{-3}**	0	2
DQB1*0501	0.13	1.34 (1.11, 1.61)	2.5×10^{-3}	0	2
DQB1*0302	0.11	0.73 (0.58, 0.91)	6.4×10^{-3}	0	3

[a] Allele frequency in discovery dataset controls, with similar frequencies in replication set

[b] Fixed-effects meta-analysis p-value adjusted for ancestral components. Bold highlighting indicates significance after Bonferroni correction with adjustment for linkage disequilibrium between variants ($P<1.3\times10^{-3}$)

[c] I2 heterogeneity index