Published in final edited form as:

Cancer Immunol Immunother. 2018 July ; 67(7): 1123–1133. doi:10.1007/s00262-018-2168-2.

Genetic variants in the HLA class II region associated with risk of cutaneous squamous cell carcinoma

Wei Wang¹, Hanna M. Ollila², Alice S. Whittemore¹, Shadmehr Demehri³, Nilah M. Ioannidis¹, Eric Jorgenson⁴, Emmanuel Mignot², Maryam M. Asgari^{3,5}

¹Department of Health Research and Policy, Stanford University School of Medicine, Stanford, CA, USA

²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

³Department of Dermatology, Massachusetts General Hospital, 50 Staniford Street, Suite 270, 02114 Boston, MA, USA

⁴Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA

⁵Department of Population Medicine, Harvard Medical School, Boston, MA, USA

Abstract

Background—The immune system has been implicated in the pathophysiology of cutaneous squamous cell carcinoma (cSCC) as evidenced by the substantially increased risk of cSCC in immunosuppressed individuals. Associations between cSCC risk and single nucleotide polymorphisms (SNPs) in the HLA region have been identified by genome-wide association studies (GWAS). The translation of the associated HLA SNPs to structural amino acids changes in HLA molecules has not been previously elucidated.

Methods—Using data from a GWAS that included 7238 cSCC cases and 56,961 controls of non-Hispanic white ancestry, we imputed classical alleles and corresponding amino acid changes in HLA genes. Logistic regression models were used to examine associations between cSCC risk and genotyped or imputed SNPs, classical HLA alleles, and amino acid changes.

Results—Among the genotyped SNPs, cSCC risk was associated with rs28535317 (OR = 1.20, $p = 9.88 \times 10^{-11}$) corresponding to an amino-acid change from phenylalanine to leucine at codon 26

Maryam M. Asgari, masgari@partners.org.

Author contributions WW: drafting and revisions to manuscript, acquisition of data, and/or analysis and interpretation of results. HMO: drafting and revisions to manuscript, acquisition of data, and/or analysis and interpretation of results. ASW: conception and design, revisions to manuscript. SD: interpretation of results and revisions to manuscript. NMI: revisions to manuscript. EJ: interpretation of results and revisions to manuscript. EM: conception and design, data analysis, revisions to manuscript. MMA: conception, funding, data acquisition, revisions to manuscript. Emmanuel Mignot and Maryam M. Asgari equally contributing as last author

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00262-018-2168-2>) contains supplementary material, which is available to authorized users.

Conflict of interest Dr. Asgari has received institutional research funding (role: Principal Investigator) from Valeant Pharmaceuticals for an unrelated topic (atopic dermatitis). All other authors declare that they have no conflict of interest.

Ethical approval and ethical standards This study was approved by the Kaiser Permanente Northern California Institutional review board (approval number CN-12MAsga-03-H) and was conducted in accordance with the Helsinki declaration.

Informed consent Informed consent was waived.

of *HLA-DRB1* ($OR = 1.17$, $p = 2.48 \times 10^{-10}$). An additional independent association was observed for a threonine to isoleucine change at codon 107 of *HLA-DQA1* ($OR = 1.14$, $p = 2.34 \times 10^{-9}$). Among the classical HLA alleles, cSCC was associated with *DRB1*01* ($OR = 1.18$, $p = 5.86 \times 10^{-10}$). Conditional analyses revealed additional independent cSCC associations with *DQA1*05:01* and *DQA1*05:05*. Extended haplotype analysis was used to complement the imputed haplotypes, which identified three extended haplotypes in the *HLA-DR* and *HLA-DQ* regions.

Conclusions—Associations with specific *HLA-DR* and *-DQ* alleles are likely to explain previously observed GWAS signals in the HLA region associated with cSCC risk.

Keywords

Skin cancer; Immunology; HLA; Genetics; Genome-wide association study

Introduction

Cutaneous squamous cell carcinoma (cSCC) is the second most common cancer among non-Hispanic whites in the United States [1]. Risk factors for cSCC include skin pigmentation, sun exposure, smoking, and immunosuppression [2, 3]. A role for immunosuppression in cSCC risk is supported by higher incidence among recipients of solid organ transplants [4] and individuals with human immunodeficiency virus infection [5].

The immune response is modulated by human leukocyte antigens (HLAs), which are encoded by a cluster of highly polymorphic genes located on chromosome 6. Direct links between HLA allelic variants and cSCC risk have been suggested by previous small studies comparing HLA serotypes. In a recent genome-wide association study (GWAS), we found cSCC risk to be associated with the intronic single nucleotide polymorphism (SNP) rs4455710 in *HLA-DQA1* [6]. A suggestive association of a single nucleotide deletion in the upstream of *HLA-DQB1* (rs28993540) was also reported by another GWAS [7]; however this association did not reach genome-wide significance. Both variants may be correlated with a single underlying causal variant that was not directly tested in either study. A limitation of GWAS data for examining the HLA region is its focus on genotyped or imputed common SNPs, which cannot identify underlying causal HLA polymorphism. Moreover, the limited number of subjects in the fully sequenced reference panel used to impute GWAS genotypes is insufficient to cover all the allelic variability of the highly polymorphic HLA region.

We identified functional polymorphisms in HLA genes responsible for the HLA association signal using an HLA-specific imputation tool called SNP2HLA [8]. This tool uses a large reference panel genotyped at a dense set of tagging markers for the entire HLA region to impute the two-digit and four-digit HLA alleles (“classical HLA alleles”) and corresponding amino acid changes. By linking the association findings to the known or predicted functionality of the HLA molecules, results from our SNP2HLA analysis can facilitate further understanding of causal mechanisms in this region.

Materials and methods

Study subjects

Study subjects were participants of the Kaiser Permanente Northern California (KPNC) GERA (Genetic Epidemiology Research in Adult Health and Aging) cohort. Subjects with one or more cSCC diagnoses after study entry were identified from KPNC pathology records. Controls were those with no skin cancer history at study entry and no subsequent skin cancer diagnosis during follow-up. Additional details of subject selection can be found elsewhere [6].

SNP genotyping, quality control and genetic ancestry estimation

Genotyping was performed on custom Affymetrix Axiom arrays with Axiom v1 or Axiom v2 reagent kits, as described previously [9, 10]. Of 67,867 eligible subjects [6], we included 64,199 who were genotyped on the European array. A total of 7,689 genotyped SNPs are in the HLA region. After removing 1,606 SNPs with minor allele frequency (MAF) < 0.01 or Hardy–Weinberg equilibrium $< 1 \times 10^{-6}$ in controls, 6,083 genotyped SNPs were included in the analysis. Principal component analysis of genetic ancestry was performed as described previously [11].

Imputation of HLA genotypes

Classical HLA alleles and corresponding amino acid changes at class I and class II loci were imputed using the SNP2HLA package, with reference data of 5,225 individuals of European descent from the Type 1 Diabetes Genetics Consortium (T1DGC) [8]. We generated genotype probabilities and dosages of 424 classical alleles, 1,276 amino acid changes (at 399 sites) in the *HLA-A*, *-B*, *-C*, *-DRB1*, *-DQA1*, *-DQB1*, *-DPA1*, and *-DPB1* genes and 7258 SNPs (2787 genotyped and 4471 imputed) [12].

Statistical analysis

Analysis of single SNPs For genotyped SNPs in the HLA region, minor allele counts were used in marker-specific logistic regressions to examine their log-additive effects on cSCC risk. Similarly, for all classical alleles, bi-allelic amino acid changes and imputed SNPs, the imputed dosage (of the given classical allele or amino acid residue, or the minor allele of the SNP) was used in logistic regressions. Markers with MAF < 0.01 were excluded. The first 10 principal components (PCs) of ancestry, sex, and reagent kit were included as covariates in all logistic regression models. Further adjustment for age and immunosuppressive status (defined as subject's previous history of organ transplant, bone marrow transplant, HIV infection, or chronic lymphocytic leukemia) resulted in minimal changes in the risk estimates and, therefore, they were not included as covariates in the final analyses. Due to small number of immunosuppressed subjects in this study (1.4% in cases and 0.6% in controls), we were unable to perform meaningful stratified analysis by immunosuppressive status.

Analysis of multiple-allelic amino acid changes For amino acid sites with $m > 2$ residues, we first performed the single-marker analysis as described above for any given residue at the site. We then performed an additional logistic regression that included $m - 1$ dosage terms

(one for each of the amino acid residues other than the most frequent one) and used a likelihood ratio test of $m - 1$ degrees of freedom to examine the global effect of the amino acid changes at the particular site [8].

Analysis of extended haplotypes For each subject, the extended haplotypes were extracted from the fully phased data generated by SNP2HLA. The effect of haplotype on cSCC risk was estimated by including individual haplotype counts in a logistic regression model.

Conditional analysis To examine independent effects of genotyped SNPs, we performed stepwise logistic regressions to identify SNPs that were independently associated with SCC risk after adjusting for the most significant SNP. Similar conditional logistic regressions were applied to examine independent effects of amino acid changes, classical alleles, extended haplotypes or imputed SNPs. We also performed conditional analysis by sequentially removing carriers of the most significant classical alleles or extended haplotypes and examining the other markers within the remaining subject sets. Single marker analyses were performed using PLINK v1.9 [13]. All other analyses were performed with R.

Data visualization Manhattan plots were generated using the QQman package in R. LocusZoom was used to generate mini-Manhattan plots. We used dbMHC sequence Alignment Viewer to examine amino acid changes corresponding to the classical alleles in the HLA genes.

Results

We performed three conditional logistic regression analyses for cSCC association: SNPs, amino acids, and classical HLA alleles. To address multiple signals, conditional analyses were conducted both within variant type and across associated variants.

Association of individual genotyped SNPs with cSCC Of 6083 genotyped SNPs in the HLA region, 86 were associated with cSCC at the genome-wide significance level ($p < 5 \times 10^{-8}$) (Table 1; Fig. 1). The most strongly associated SNP was rs28535317 (OR = 1.20, $p = 9.88 \times 10^{-11}$), an intergenic SNP located 16 kb upstream of *HLA-DQA1* and 31 kb upstream of *HLA-DRB1*. Linkage disequilibrium (LD) between rs28535317 and the other 85 associated SNPs varied greatly, with R^2 ranging from 0.1 to 1. After adjustment for rs28535317, the 85 SNPs were no longer significant at the genome-wide level, but 22 SNPs that were originally borderline significant ($p < 7 \times 10^{-7}$) reached genome-wide significance (Table 1; Fig. 1). SNP rs9272063 had the smallest p value after adjusting for rs28535317 (adjusted OR = 1.15, adjusted $p = 7.55 \times 10^{-10}$); this SNP is in almost complete LD ($R^2 = 0.95$) with a coding polymorphism in *HLA-DQA1* (rs707952, adjusted OR = 1.14, $p = 2.34 \times 10^{-9}$), located 10 kb downstream. SNP rs707952 results in a threonine (T) to isoleucine (I) change at codon 107 (amino acid 130) in exon 3 of *HLA-DQA1*, which is characteristic of the *DQA1*05* group. After adjustment for both rs28535317 and rs9272063, no other association reached genome-wide significance; the strongest remaining association was with rs13220896, an intronic SNP in the gene *MICA* (OR = 1.18, $p = 1.49 \times 10^{-4}$, Fig. 2).

Association of imputed amino acid changes in HLA genes with cSCC Among the 1,276 imputed HLA amino acid changes, we found that changes in *HLA-DRB1* codon 26 in exon 2 were the most significantly associated with cSCC, with $p = 2.29 \times 10^{-10}$ for the omnibus test (Fig. 3a); OR = 0.88 and $p = 1.33 \times 10^{-11}$ for phenylalanine (F); and OR = 1.17 and $p = 2.48 \times 10^{-10}$ for leucine (L). Tyrosine (Y) at this codon was not significantly associated with cSCC risk. This exon 2 codon variant results in a complex association, as *DRB1_26_F* is shared by numerous classical *HLA-DRB1* alleles. It is also located in one of the most polymorphic regions of exon 2, so neighboring amino acids are also highly variable. Amino acid changes at codons 11, 13, 31, and 96 in *DRB1* were also significantly associated with cSCC risk (Fig. 4a). No amino acid changes in other HLA genes reached genome-wide significance.

DRB1_26_F was moderately correlated with our most significant SNP, rs28535317 ($R^2 = 0.29$). After adjustment, the association levels for both rs28535317 and *DRB1_26_F* were reduced but still borderline significant (OR = 0.91, $p = 8.81 \times 10^{-5}$ for *DRB1_26_F*, and OR = 1.12, $p = 0.001$ for rs28535317). A group of amino acid changes in *DQA1* that are in perfect LD with each other became significant after adjustment for rs28535317 (Fig. 4b). They occur at codons 40, 51, 75, 107, 156, 163, and 175, characteristic of the *DQA1*05* allele. After adjustment for *DRB1_26_F*, no amino acid change was associated with cSCC at the genome-wide significant level. The strongest adjusted association was for *DRB1* at codon 58 in exon 2 [$p = 3.01 \times 10^{-6}$ (Fig. 3b), OR = 1.18 and $p = 2.31 \times 10^{-7}$ for glutamic acid (E), an amino-acid shared by the *DRB1*11* group, vs. alanine (A)]. Considering the complexity of these associations, which occur at amino acid sites that are often multiallelic, we next examined the imputed classical HLA alleles in the *DR* and *DQ* region.

Association of imputed classical alleles in HLA genes with cSCC among the HLA classical alleles, the strongest association with cSCC was found for *DRB1*01* (OR = 1.18, $p = 5.86 \times 10^{-10}$, Supplementary Table 1). This finding is consistent with the amino acid associations described above, since *DRB1_26_L* is mostly present on *DRB1*01*, a frequent subtype in comparison to other alleles also carrying *DRB1_26_L* (*DRB1*10*, frequency < 0.01 and *DRB1*12*, frequency = 0.02). As *DRB1*12* is not associated with cSCC, a primary association with *DRB1*01* or linked DQ alleles (*DQA1*01:01* and *DQB1*05:01*) is likely. After adjustment for *DRB1_26_L*, the association with *DRB1*01* was no longer significant (OR 1.08, $p = 0.21$, Supplementary Table 2). After conditioning for *DRB1*01*, a genome-wide significant association with classical alleles *DQA1*05:01/5* (these two subtypes were not distinguished by SNP2HLA) was revealed (adjusted OR = 1.15, adjusted $p = 9.42 \times 10^{-10}$) (Supplementary Table 1).

Association of extended haplotypes in the DRB1, DQA1, and DQB1 region with cSCC Given that SNP2HLA does not differentiate all known HLA classical alleles, known extended haplotypes were used to complement the imputed haplotypes (Table 2). As expected, *DRB1*01:01-DQA1*01:01-DQB1*05:01* showed the strongest cSCC risk association (OR = 1.19, $p = 1.38 \times 10^{-8}$). Interestingly, although *DRB1*01:03-DQA1*01:01-DQB1*05:01* also appeared to be associated with cSCC (OR = 1.30, $p = 0.01$), *DRB1*01:02-DQA1*01:01-DQB1*05:01* and *DRB1*10-DQA1*01:05-DQB1*05:01* were not ($p > 0.05$), suggesting a *DR*, and not a *DQ*, effect. However, sample size was small,

and *DRB1*01:03-DQA1*01:01-DQB1*05:01* did not reach genome-wide significance level. Adjusting for the *DRB1*01* allele in the extended haplotype analysis revealed additional cSCC associations with both *DRB1*11:01-DQA1*05:05* and *DRB1*03:01-DQA1*05:01* (which track with the *DQA1*05:01/05* risk allele). An association was also observed for the closely related *DRB1*11:04-DQA1*05:05-DQB1*03:01* haplotype, but not for *DRB1*12-DQA1*05:05-DQB1*03:01*, suggesting an effect of both *DRB1*11:01* and *DRB1*11:04* alleles, two closely related alleles differentiated by a glycine to valine change at codon 86 in exon 2, or an effect of the *DQA1*05:05/DQB1*03:01* heterodimer.

Additional adjustment for *DQA1*05-DQB1*03:01* revealed a suggestive third association with *DRB1*03:01-DQA1*05:01-DQB1*02:01* ($p = 3 \times 10^{-3}$). Because both *DQA1*05:01/5* and *DQB1*02:01* or *DQB1*03:01* are associated with cSCC, it was impossible to further decipher these effects. *DRB1*07:01-DQA1*02:01-DQB1*02:02* was not associated with cSCC risk (Table 2).

Additional HLA associations with cSCC beyond DRB1, DQA1 and DQB1 loci Independent effects of HLA Class I and HLA DP alleles were examined after adjusting for all associated *DRB1*, *DQA1* and *DQB1* classical alleles. None of the DP or class I classical alleles reached genome-wide significance (data not shown). After adjustment for the *DRB1*, *DQA1* and *DQB1* classical alleles, none of the genotyped SNPs in the HLA region reached genome-wide significance level. The strongest remaining association was for rs9261191, an intronic SNP in *ZNRD1-AS1* (OR = 1.15, $p = 4.01 \times 10^{-3}$, Supplementary Fig. 1). It was not in high LD with any of the DP or class I classical alleles listed above.

Discussion

We identified an association between cSCC risk and a phenylalanine to leucine change at codon 26 of *HLA-DRB1*. The leucine amino acid occurs most frequently on the *DRB1*01* classical allele, which also shows the strongest association among classical DR alleles. An additional independent association was observed for the *DQA1*05* classical allele, characterized by a threonine to isoleucine change at codon 107 of *HLA-DQA1*. Analysis of extended haplotypes spanning the *DRB1*, *DQA1* and *DQB1* regions suggested positive associations with *DRB1*01-DQA1*01:01-DQB1*05:01*, *DRB1*11:01/04-DQA1*05:05-DQB1*03:01* and *DRB1*03:01-DQA1*05:01-DQB1*02:01*, with trans and haplotype effects suggesting functional relevance of specific alleles rather than linkage disequilibrium. These extended haplotype associations are likely to explain previous GWAS-observed association signals in the HLA region for cSCC risk [6, 7], and could benefit from additional trans-ethnic studies to strengthen individual allele association [14]. Alternatively, effects in the HLA region may be mediated by intronic polymorphisms such as rs28535317, as some of these have been shown to be eQTL for the expression of HLA genes [15]. A previously published positive association in a small study between the DR1 serotype and cSCC risk is consistent with the present findings [16–18].

The exact mechanisms underlying the observed associations between variations in HLA class II genes and cSCC risk needs to be elucidated. The immune system plays complex roles over the entire process of cancer initiation, promotion and progression. On the one

hand, the immune system can function as an extrinsic tumor suppressor by identifying and destroying transformed cells. Presentation of tumor antigens to CD8 + cytotoxic T cells and CD4 + helper T cells by HLA class I and class II molecules, respectively, is a key component of this process. Specifically, as the upstream activators of immune response, CD4 + T helper cells can engage a large array of adaptive and innate immune effector cells to eliminate cancerous cells. On the other hand, inflammation can facilitate cell transformation by providing pro-tumorigenic cytokines and growth factors to tumor cells and forming an immune suppressive microenvironment within the tumor, which ultimately lead to immune escape and clinical manifestation of the tumors [19, 20]. HLA-DR-expressing CD4 + regulatory T cells are one of the major immunosuppressive leukocyte populations that play key roles in inhibiting host-protective antitumor responses [19, 21]. Antigen presentation by HLA-DR molecules on CD4 + T cells to autologous T cells has been shown to be a dominant tolerance that persists even in the presence of macrophages and other antigen-presenting cells [21]. Therefore, the observed associations between the amino acid changes in the binding grooves of HLA-DRB1 and DQA1 (and their associated classical alleles) may be explained by the functional impact on antigenic peptide presentation by APC to CD4 + T helper cells or by CD4 + regulatory T cells to other T cells. The complex role of HLAs in tumor development and progression is exemplified by a study of HLA class II expression and promoter methylation in SCC arising in different class of epithelium, namely esophageal SCC. Findings comparing esophageal SCC to adjacent normal esophagus tissue suggested that aberrant *HLA-DRB1* and *HLA-DQB1* methylation contributes to changes in HLA class II expression, and that these molecular changes may influence the immune response to specific tumor epitopes, promoting SCC development and progression [22].

Several cSCC-associated classical alleles and long-range haplotypes in the *DRB1*, *DQA1*, and *DQB1* regions identified in this study have been associated with autoimmune/inflammatory diseases [23], which are characterized by enhanced and poorly regulated immune response. These HLA variants may predispose to cSCC through inflammation-independent effects such as increasing antigen presenting capacity of HLA class II molecules expressed on regulatory T cells, which could lead to immune tolerance; and placing constant immune selection pressure on genetically unstable tumor cells, which leads to the emergence of poorly immunogenetic tumor cell variants, which can then escape immunity and result in clinically apparent disease.

Several specific HLA-DRB1 amino acid positions may account for the observed association between the HLA alleles with autoimmune and/or inflammatory diseases: codons 11, 13, and the shared epitope positions (70–74) with rheumatoid arthritis [24, 25], codon 11 with sarcoidosis, codons 13 and 60 with multiple sclerosis, codon 37 with psoriasis vulgaris, codon 78 with primary biliary cirrhosis, and codon 74 with Graves' disease (summarized in [26]). The strongest association between amino acid changes and cSCC risk was found for codon 26 of *HLA-DRB1*. Although we also observed significant associations for codons 11, 13, 31 and 96, adjustment for codon 26 nullified these associations and, therefore, our analysis does not support an independent role for these other codons in modulating cSCC risk. Codon 26 lies on the β -sheet with its side chain oriented into the peptide-binding groove. The phenylalanine to leucine change could potentially affect antigen binding and

presentation. However, the true functional impact of the amino acid change at codon 26 on *HLA-DRB1* molecule remains to be elucidated.

Strengths of this study include the large sample size and high-quality genotyping data [27]. One limitation is that the amino acid changes, classical alleles and long-range haplotypes were imputed, not directly genotyped. However, previous studies have shown that the SNP2HLA imputation is reasonably accurate, based on high concordances between genotyped and imputed HLA alleles [8, 26, 28]. Another limitation is the low resolution of the SNP2HLA package and relatively homogeneous study population, which both results in lack of diversity in HLA alleles and thus difficulty in pinpointing the causal loci associated with cSCC.

This study adds evidence to the known importance of the immune system in determining cSCC predisposition. The identification of specific amino acid changes in the HLA class II genes, if confirmed, helps provide mechanistic clues to the relationship between HLA-mediated immune response and cSCC tumorigenesis. Future studies that examine the mechanism underlying the association between HLA class II and cSCC risk need to be performed. These can include in vitro exposure of CD4 + T cells to antigen-presenting cells expressing certain HLA class II alleles co-cultured with irradiated autologous cSCC cells [29]. The degree of CD4 + T cell activation and its polarization state in this context will help elucidate the anti- versus pro-tumor effect of the HLA class II alleles on cSCC. Understanding these mechanisms may be critical for developing therapeutic tools to modulate the immune responses to inhibit the growth of cSCC.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by a grant from the National Cancer Institute at the National Institutes of Health (Grant number R01CA166672).

Abbreviations

cSCC	Cutaneous squamous cell carcinoma
GERA	Genetic Epidemiology Research in Adult Health and Aging
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
KPNC	Kaiser Permanente Northern California
LD	Linkage disequilibrium
MAF	Minor allele frequency
OR	Odds ratio

SNP

Single nucleotide polymorphism

References

1. Jennings L, Schmults CD (2010) Management of high-risk cutaneous squamous cell carcinoma. *J Clin Aesthet Dermatol* 3(4):39–48
2. Lindström LS, Yip B, Lichtenstein P et al. (2007) Etiology of familial aggregation in melanoma and squamous cell carcinoma of the skin. *Cancer Epidemiol Biomarkers Prev* 16(8):1639–1643 [PubMed: 17684139]
3. Trakatelli M, Ulrich C, del Marmol V et al. (2007) Epidemiology of nonmelanoma skin cancer (NMSC) in Europe: accurate and comparable data are needed for effective public health monitoring and interventions. *Br J Dermatol* 156(Suppl 3):1–7
4. Euvrard S, Kanitakis J, Claudy A (2003) Skin cancers after organ transplantation. *N Engl J Med* 348(17):1681–1691 [PubMed: 12711744]
5. Silverberg MJ, Leyden W, Warton EM et al. (2013) HIV infection status, immunodeficiency, and the incidence of non-melanoma skin cancer. *J Natl Cancer Inst* 105(5):350–60 [PubMed: 23291375]
6. Asgari MM, Wang W, Ioannidis NM et al. (2016) Identification of susceptibility loci for cutaneous squamous cell carcinoma. *J Invest Dermatol* 136(5):930–937 [PubMed: 26829030]
7. Chahal HS, Lin Y, Ransohoff KJ et al. (2016) Genome-wide association study identifies novel susceptibility loci for cutaneous squamous cell carcinoma. *Nat Commun* 7:12048 [PubMed: 27424798]
8. Jia X, Han B, Onengut-Gumuscu S et al. (2013) Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 8(6):e64683 [PubMed: 23762245]
9. Hoffmann TJ, Zhan Y, Kvale MN et al. (2011) 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* 98(6):422–430 [PubMed: 21903159]
10. Hoffmann TJ, Kvale MN, Hesselton SE et al. (2011) Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics* 98(2):79–89 [PubMed: 21565264]
11. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190 [PubMed: 17194218]
12. Jorgenson E, Melles RB, Hoffmann TJ et al. (2016) Common coding variants in the HLA-DQB1 region confer susceptibility to age-related macular degeneration. *Eur J Hum Genet* 24(7):1049–1055 [PubMed: 26733291]
13. Chang CC, Chow CC, Tellier LC et al. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7 [PubMed: 25722852]
14. Ollila HM, Ravel JM, Han F et al. (2015) HLA-DPB1 and HLA class I confer risk of and protection from narcolepsy. *Am J Hum Genet* 96(1):136–46 [PubMed: 25574827]
15. Raj P, Rai E, Song R et al. Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *Elife* 2016;5:e12089, [PubMed: 26880555]
16. Glover MT, Bodmer J, Bodmer W et al. (1993) HLA antigen frequencies in renal transplant recipients and non-immunosuppressed patients with non-melanoma skin cancer. *Eur J Cancer* 29A(4):520–524 [PubMed: 8435203]
17. Cerimele D, Contu L, Carcassi C et al. (1988) HLA and multiple skin carcinomas. *Dermatologica* 176(4):176–81 [PubMed: 3378654]
18. Czarnecki DB, Lewis A, Nicholson I et al. (1991) Multiple non-melanoma skin cancer associated with HLA DR7 in southern Australia. *Cancer* 68(2):439–40 [PubMed: 2070340]
19. Schreiber RD, Old LJ, Smyth MJ (2011) Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science* 331(6024):1565–1570 [PubMed: 21436444]
20. Vinay DS, Ryan EP, Pawelec G et al. (2015) Immune evasion in cancer: mechanistic basis and therapeutic strategies. *Semin Cancer Biol* 35 S185–98 [PubMed: 25818339]

21. Baecher-Allan C, Wolf E, Hafler DA (2006) MHC class II expression identifies functionally distinct human regulatory T cells. *J Immunol* 176(8):4622–4631 [PubMed: 16585553]
22. Hu JM, Li L, Chen YZ, et al. HLA-DRB1 and HLA-DQB1 methylation changes promote the occurrence and progression of Kazakh ESCC. *Epigenetics* 2014;9(10):1366–1373 [PubMed: 25437052]
23. Jones EY, Fugger L, Strominger JL et al. (2006) MHC class II proteins and disease: a structural perspective. *Nat Rev Immunol* 6(4):271–82 [PubMed: 16557259]
24. Raychaudhuri S, Sandor C, Stahl EA et al. (2012) Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 44(3):291–296 [PubMed: 22286218]
25. Feitsma AL, van der Helm-van Mil AH, Huizinga TW et al. (2008) Protection against rheumatoid arthritis by HLA: nature and nurture. *Ann Rheum Dis* 67(Suppl 3):iii61–3 [PubMed: 19022816]
26. Fernandez CA, Smith C, Yang W et al. (2014) HLA-DRB1*07:01 is associated with a higher risk of asparaginase allergies. *Blood* 124(8):1266–1276 [PubMed: 24970932]
27. Kvale MN, Hesselton S, Hoffmann TJ et al. (2015) Genotyping Informatics and Quality Control for 100,000 Subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA). *Cohort Genetics* 200(4):1051–1060 [PubMed: 26092718]
28. Pereyra F, Jia X, McLaren PJ et al. (2010) The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* 330(6010):1551–1557 [PubMed: 21051598]
29. Butler MO, Ansén S, Tanaka M et al. (2010) A panel of human cell-based artificial APC enables the expansion of long-lived antigen-specific CD4 + T cells restricted by prevalent HLA-DR alleles. *Int Immunol* 22(11):863–73 [PubMed: 21059769]

Fig. 1.

Manhattan plot of typed SNPs in the HLA region, before and after adjustment for rs28535317

Fig. 2.
Manhattan plot of typed SNPs in the HLA region, after adjustment for both rs28535317 and rs707952

Fig. 3. Associations between SNP2HLA imputed amino acid changes and cSCC (**a**) and the remaining association after adjustment for the presence of Phe at HLA_DRB1 codon 26 (**b**)

Fig. 4.

Associations between cSCC risk and SNP2HLA imputed amino acid changes in DRB1 (**a**), and in DQA1 after adjustment for rs28535317 (**b**). Colors represent the correlations between the amino acid changes (as measured by the highest squared Pearson correlation coefficient between the imputed amino acid dosages at the selected codon and at the top codon in the gene)

Table 1
 Typed SNPs associated with cSCC ($p < 5 \times 10^{-8}$), before and after adjustment for rs28535317

SNP	hg18_pos	Minor allele	OR	P value	OR_adj	p_adj	Functional impact	Gene
rs28535317	32,697,043	G	1.204	9.88×10^{-11}	NA	NA	Regulatory/intergenic	
rs12526612	32,761,576	A	1.188	1.14×10^{-10}	1.082	0.18	Regulatory/intergenic	
rs112627226	32,690,947	C	1.203	1.29×10^{-10}	0.842	0.81	Intergenic	
rs72847197	32,782,230	A	1.191	1.32×10^{-10}	1.075	0.22	NA	NA
rs12528797	32,453,064	G	1.194	1.40×10^{-10}	1.075	0.31	Intron/intergenic	LOC101929163
rs10484561	32,773,398	G	1.187	1.53×10^{-10}	1.072	0.24	Intergenic	
rs3998180	32,758,875	A	1.186	1.62×10^{-10}	1.080	0.19	Intergenic	
rs72847174	32,776,004	T	1.186	1.65×10^{-10}	1.081	0.18	Regulatory/intergenic	
rs17427564	32,775,045	C	1.186	1.74×10^{-10}	1.069	0.25	Intergenic	
rs67887332	32,772,377	A	1.186	1.78×10^{-10}	1.071	0.24	Intergenic	
rs72843270	32,454,191	T	1.195	1.86×10^{-10}	1.081	0.26	Intron/intergenic	LOC101929163
rs9271899	32,703,985	A	0.892	2.14×10^{-10}	0.925	9.35×10^{-5}	Regulatory/intergenic	
rs28530648	32,635,057	C	1.180	2.15×10^{-10}	1.039	0.53	Intron, non-coding RNA	HLA-DRB6
rs35998847	32,774,975	A	1.185	2.16×10^{-10}	1.064	0.29	Intergenic	
rs17206182	32,773,482	T	1.185	2.26×10^{-10}	1.077	0.20	Intergenic	
rs17202892	32,542,457	A	1.191	2.47×10^{-10}	0.981	0.84	Intergenic	
rs4530903	32,689,867	T	1.187	2.50×10^{-10}	1.059	0.47	Intergenic	
rs12524667	32,497,312	T	1.199	3.26×10^{-10}	0.967	0.83	Intergenic	
rs6930777	32,459,544	T	1.189	3.39×10^{-10}	1.065	0.38	Downstream/intron/intergenic	LOC101929163
rs9271824	32,702,894	G	1.119	3.54×10^{-10}	1.081	9.93×10^{-5}	Regulatory/intergenic	
rs2395154	32,453,754	T	1.190	3.58×10^{-10}	1.084	0.25	Intron/intergenic	LOC101929163
rs35678932	32,774,841	T	1.183	3.62×10^{-10}	1.060	0.31	Intergenic	
rs17211356	32,686,404	G	1.186	3.65×10^{-10}	1.031	0.69	Regulatory/intergenic	
rs9271300	32,689,560	C	0.893	3.75×10^{-10}	0.920	2.60×10^{-5}	Intergenic	
rs9271926	32,704,544	T	1.119	4.04×10^{-10}	1.081	1.08×10^{-4}	Regulatory/intergenic	
rs13217979	32,549,104	T	1.192	4.06×10^{-10}	0.981	0.84	Intergenic	
rs72847183	32,777,091	C	1.185	4.10×10^{-10}	1.063	0.29	Intergenic	

SNP	hg18_pos	Minor allele	OR	P value	OR_adj	p_adj	Functional impact	Gene
rs2894253	32,453,518	G	1.188	4.25×10^{-10}	1.072	0.33	Intron/intergenic	LOC101929163
rs9271857	32,703,125	C	0.894	4.37×10^{-10}	0.926	1.31×10^{-4}	Regulatory/intergenic	
rs9271897	32,703,932	G	0.894	4.40×10^{-10}	0.926	1.10×10^{-4}	Regulatory/intergenic	HLA-DQA1
rs9272130	32,708,616	A	0.894	4.41×10^{-10}	0.927	1.30×10^{-4}	Upstream	
rs17212553	32,773,665	T	1.181	4.61×10^{-10}	1.063	0.29	Intergenic	HLA-DQA1
rs9272117	32,708,389	C	0.894	4.65×10^{-10}	0.926	1.10×10^{-4}	Upstream	
rs35954087	32,548,027	T	1.189	4.73×10^{-10}	0.984	0.87	Regulatory/intergenic	HLA-DQA1
rs116603449	32,703,172	C	0.894	4.84×10^{-10}	0.927	1.52×10^{-4}	Regulatory/intergenic	
rs78825896	32,775,159	G	1.181	4.91×10^{-10}	1.067	0.26	Intergenic	HLA-DQA1
rs9272120	32,708,479	A	0.895	4.92×10^{-10}	0.926	1.31×10^{-4}	Upstream	
rs35522883	32,658,326	T	1.183	5.02×10^{-10}	1.055	0.48	Intron/upstream	HLA-DRB1/LOC105369230
rs12528622	32,766,237	A	1.178	5.14×10^{-10}	1.062	0.27	Regulatory/intergenic	
rs59108952	32,682,222	T	1.183	5.18×10^{-10}	1.030	0.70	Regulatory/intergenic	HLA-DQA1
rs9272113	32,708,269	C	1.118	5.44×10^{-10}	1.081	1.14×10^{-4}	Upstream	
rs28377518	32,540,628	T	1.188	5.65×10^{-10}	0.962	0.69	Intergenic/upstream	LOC105379570/LOC105379648/LOC105379652/LOC105379658/LOC105379665
rs113979555	32,691,166	A	1.182	6.10×10^{-10}	1.031	0.70	Intergenic	
rs113711404	32,691,735	A	1.182	6.11×10^{-10}	1.024	0.76	Regulatory/intergenic	LOC105379570/LOC105379648/LOC105379652/LOC105379658/LOC105379665
rs35472814	32,767,875	C	1.180	6.46×10^{-10}	1.068	0.26	Regulatory/intergenic	
rs36233212	32,683,798	C	1.182	6.62×10^{-10}	1.036	0.64	Regulatory/intergenic	LOC105379570/LOC105379648/LOC105379652/LOC105379658/LOC105379665
rs9271720	32,701,485	G	0.895	6.68×10^{-10}	0.923	6.83×10^{-5}	Regulatory/intergenic	
rs113585425	32,541,103	C	1.195	7.00×10^{-10}	0.985	0.93	Intergenic/upstream	HLA-DQA1
rs72850016	32,680,918	G	1.182	7.78×10^{-10}	1.032	0.67	Regulatory/intergenic	
rs112270441	32,690,958	T	1.182	7.85×10^{-10}	1.008	0.91	Intergenic	HLA-DQA1
rs35473146	32,771,691	G	1.179	8.03×10^{-10}	1.060	0.32	Intergenic	
rs9272070	32,707,463	T	0.896	8.12×10^{-10}	0.927	1.48×10^{-4}	Intergenic	HLA-DQA1
rs9272254	32,711,076	C	0.896	8.21×10^{-10}	0.928	1.62×10^{-4}	Upstream	
rs35122881	32,771,655	T	1.176	8.57×10^{-10}	1.062	0.270	Intergenic	NA
rs35900855	32,772,016	G	1.177	9.46×10^{-10}	1.064	0.29	Intergenic	
rs36226783	32,707,383	A	1.180	9.55×10^{-10}	1.021	0.79	NA	NA

SNP	hg18_pos	Minor allele	OR	P value	OR_adj	p_adj	Functional impact	Gene
rs28380113	32,690,470	T	1.180	9.73×10^{-10}	1.013	0.87	Regulatory/intergenic	HLA-DRB1/LOC105369230
rs3830123	32,656,624	T	1.180	9.94×10^{-10}	1.043	0.57	Intron	
rs35696625	32,739,230	C	1.177	1.02×10^{-9}	1.062	0.30	Intron	
rs9271893	32,703,837	G	0.896	1.06×10^{-9}	0.928	1.95×10^{-4}	Regulatory/intergenic	HLA-DQB1
rs9272118	32,708,426	C	0.896	1.07×10^{-9}	0.928	1.84×10^{-4}	Upstream	
rs9271848	32,703,005	A	0.897	1.20×10^{-9}	0.928	1.96×10^{-4}	Regulatory/intergenic	
rs36022997	32,681,339	C	1.179	1.22×10^{-9}	1.015	0.84	Regulatory/intergenic	HLA-DQA1
rs72844253	32,664,241	G	1.172	1.27×10^{-9}	1.083	0.22	Intron/regulatory	
rs55873471	32,689,690	G	1.178	1.34×10^{-9}	1.024	0.76	Intergenic	
rs60545961	32,550,134	T	1.185	1.78×10^{-9}	0.968	0.72	Intergenic	HLA-DRB1/LOC105369230
rs72847591	32,678,472	T	1.176	2.40×10^{-9}	1.013	0.87	Regulatory/intergenic	
rs28638783	32,680,715	A	1.176	2.53×10^{-9}	1.007	0.92	Regulatory/intergenic	
rs35875335	32,499,969	C	1.176	3.61×10^{-9}	0.921	0.31	Intergenic	HLA-DRA
rs35637654	32,694,119	A	1.169	5.35×10^{-9}	0.999	0.98	Intergenic	
rs34106969	32,699,812	T	1.151	6.08×10^{-9}	1.033	0.42	Regulatory/intergenic	
rs35040284	32,694,046	C	1.171	7.11×10^{-9}	0.982	0.81	Intergenic	HLA-DRA
rs17496549	32,517,686	T	1.173	9.22×10^{-9}	0.894	0.19	Intron/regulatory/intergenic	
rs17533090	32,698,700	T	1.147	9.62×10^{-9}	1.025	0.54	Regulatory/intergenic	
rs34113780	32,683,236	G	1.169	1.15×10^{-8}	1.005	0.94	Regulatory/intergenic	HLA-DRA
rs34411532	32,700,272	A	1.146	1.23×10^{-8}	1.028	0.49	Regulatory/intergenic	
rs35216590	32,701,224	G	1.145	1.38×10^{-8}	1.024	0.55	Regulatory/intergenic	
rs17202717	32,505,841	T	1.168	1.71×10^{-8}	0.883	0.14	Intergenic	NA
rs13207836	32,704,352	G	1.144	2.17×10^{-8}	1.014	0.74	Intergenic	
rs35472536	32,701,075	G	1.143	2.28×10^{-8}	1.022	0.59	NA	
rs35658013	32,703,200	T	1.143	2.37×10^{-8}	1.019	0.64	NA	NA
rs35030589	32,780,881	A	1.147	3.09×10^{-8}	1.022	0.61	Intergenic	
rs35014460	32,699,084	C	1.142	3.49×10^{-8}	1.023	0.57	Regulatory/intergenic	
rs35675538	32,701,105	T	1.141	3.82×10^{-8}	1.024	0.55	Regulatory/intergenic	NA
rs35399661	32,698,968	C	1.139	4.31×10^{-8}	1.010	0.80	Regulatory/intergenic	
rs35201643	32,703,683	C	1.14	4.52×10^{-8}	1.015	0.71	NA	

SNP	hg18_pos	Minor allele	OR	P value	OR_adj	p_adj	Functional impact	Gene
rs9272063	32,707,281	T	1.114	1.83×10^{-7}	1.146	7.55×10^{-10}	Intergenic	
rs9272066	32,707,352	A	1.111	3.46×10^{-7}	1.143	1.12×10^{-9}	Intergenic	
rs9272065	32,707,330	G	1.110	4.69×10^{-7}	1.142	1.91×10^{-9}	Intergenic	
rs9272145	32,708,859	G	1.115	1.21×10^{-7}	1.140	2.34×10^{-9}	Upstream	HLA-DQA1
rs707952	32,717,784	T	1.112	2.48×10^{-7}	1.140	2.34×10^{-9}	Missense (DQA1 codon 107)	HLA-DQA1
rs35344500	32,717,503	C	1.111	2.83×10^{-7}	1.137	3.71×10^{-9}	Intron	HLA-DQA1
rs9272308	32,711,884	C	1.111	2.99×10^{-7}	1.136	4.45×10^{-9}	Upstream/regulatory	HLA-DQA1
rs9272224	32,710,330	T	1.102	1.22×10^{-6}	1.130	8.79×10^{-9}	Upstream	HLA-DQA1
rs9271895	32,703,884	C	1.104	2.14×10^{-6}	1.135	1.28×10^{-8}	Regulatory/intergenic	
rs9271668	32,700,752	T	1.103	2.91×10^{-6}	1.135	1.34×10^{-8}	Regulatory/intergenic	
rs3129763	32,698,903	A	1.102	3.36×10^{-6}	1.134	1.53×10^{-8}	Regulatory/intergenic	
rs9272577	32,715,303	A	1.100	1.61×10^{-6}	1.126	1.86×10^{-8}	Intron/regulatory	HLA-DQA1
rs77936036	32,716,335	A	1.108	1.20×10^{-6}	1.135	1.90×10^{-8}	Intron	HLA-DQA1
rs9272630	32,716,128	A	1.098	2.17×10^{-6}	1.126	2.14×10^{-8}	Intron	HLA-DQA1
rs617058	32,682,457	T	1.102	2.79×10^{-6}	1.132	2.25×10^{-8}	Regulatory/intergenic	
rs9271528	32,697,901	T	1.101	4.15×10^{-6}	1.132	2.42×10^{-8}	Regulatory/intergenic	
rs9270493	32,667,088	C	1.101	3.89×10^{-6}	1.131	2.89×10^{-8}	Upstream	HLA-DRB1/LOC105369230
rs9272276	32,711,612	C	1.097	3.31×10^{-6}	1.123	4.05×10^{-8}	Upstream	HLA-DQA1
rs9272840	32,718,661	T	1.097	3.39×10^{-6}	1.123	4.34×10^{-8}	Intron	HLA-DQA1
rs9271524	32,697,798	C	1.099	5.84×10^{-6}	1.130	4.36×10^{-8}	Regulatory/intergenic	
rs9271545	32,698,131	T	1.099	6.68×10^{-6}	1.130	4.49×10^{-8}	Regulatory/intergenic	
rs9271541	32,698,048	C	1.098	6.87×10^{-6}	1.130	4.56×10^{-8}	Regulatory/intergenic	

Table 2

Associations between the DRB1, DQA1 and DQB1 haplotypes and cSCC

haplotype names	Frequency reported ^a (%)	Frequency in controls (%)	Frequency in cases (%)	OR ^b	P value ^b	OR ^c	P value ^c	OR ^d	P value ^d	OR ^e	P value ^e
DRB1*01:01-DQA1*01:01-DQB1*05:01	8.43	8.25	9.93	1.187	1.38 × 10 ⁻⁸	1.044	0.47	1.207	6.92 × 10 ⁻¹⁰	1.201	1.87 × 10 ⁻⁹
DRB1*01:02-DQA1*01:01-DQB1*05:01	1.73	1.66	1.43	1.060	0.44	0.902	0.20	1.082	0.30	1.068	0.39
DRB1*01:03-DQA1*01:01-DQB1*05:01	0.64	0.60	0.84	1.300	0.01	1.134	0.22	1.315	6.52 × 10 ⁻³	1.311	0.01
DRB1*10:01-DQA1*01:05-DQB1*05:01	2.03	0.73	0.73	1.150	0.18	0.998	0.99	1.168	0.14	1.157	0.17
DRB1*11:01-DQA1*05:05-DQB1*03:01 ^f	6.21	5.80	5.72	1.114	0.01	1.140	7.79 × 10 ⁻⁴	1.017	0.74	1.123	2.84 × 10 ⁻³
DRB1*11:04-DQA1*05:05-DQB1*03:01 ^f	4.5	2.91	2.56	1.109	0.07	1.131	0.03	1.012	0.84	1.117	0.05
DRB1*12:01-DQA1*05:05-DQB1*03:01 ^f	1.43	1.69	1.64	1.019	0.79	1.043	0.55	0.918	0.25	1.026	0.71
DRB1*03:01-DQA1*05:01-DQB1*02:01	14.09	10.91	12.03	1.070	0.01	1.096	9.77 × 10 ⁻⁴	1.086	3.00 × 10 ⁻³	NA	NA
DRB1*07:01-DQA1*02:01*02:02	11.08	9.83	9.37	0.947	0.08	0.968	0.29	0.961	0.20	0.955	0.14

^aFrequency in white populations reported on allelefrequencies.net^bAdjusted for the first 10 PCs, sex and kitO^cWith additional adjustment for DQA1*01:01/5-DQB1*05:01^dWith additional adjustment for DQA1*05-DQB1*03:01^eWith additional adjustment for DRB1*03-DQA1*05:01-DQB1*02:01