

## SHORT REPORT

# A *cis*-eQTL of *HLA-DRB1* and a frameshift mutation of *MICA* contribute to the pattern of association of HLA alleles with cervical cancer

Dan Chen &amp; Ulf Gyllensten

Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Science for Life Laboratory Uppsala, Uppsala University, Uppsala, Sweden

**Keywords**Cervical cancer, *cis*-eQTL, frameshift mutation, HLA, MICA**Correspondence**

Dan Chen, Department of Immunology, Genetics and Pathology, Rudbeck Laboratory, Uppsala University, S-751 85 Uppsala, Sweden.  
Tel: +46(0)737174449; Fax: +46(0)184714931;  
E-mails: dan.chen@igp.uu.se;  
simpledandan@gmail.com

**Funding Information**

This work was supported by grants from the Swedish Cancer Society (U. G.), the Medical Faculty of Uppsala University (D. C.), The Swedish Foundation for Strategic Research (SSF) (U. G.) and the Swedish Research Council (U. G.).

Received: 2 October 2013; Revised: 1 November 2013; Accepted: 1 December 2013

*Cancer Medicine* 2014; 3(2): 445–452

doi: 10.1002/cam4.192

**Abstract**

The association of classic human leukocyte antigen (HLA) alleles with risk of cervical cancer has been extensively studied, and a protective effect has consistently been found for *DRB1\*1301*, *DQA1\*0103*, and/or *DQB1\*0603* (these three alleles are in perfect linkage disequilibrium [LD] and often occur on the same haplotype in Europeans), while reports have differed widely with respect to the effect of *HLA-B\*07*, *DRB1\*1501*, and/or *DQB1\*0602* (the last two alleles are also in perfect LD in Europeans). It is not clear whether the reported HLA alleles are responsible for the differences in cervical cancer susceptibility, or if functional variants at other locations within the major histocompatibility complex (MHC) region may explain the effect. In order to assess the relative contribution of both classic HLA alleles and single-nucleotide polymorphisms (SNPs) within the MHC region to cervical cancer susceptibility, we have imputed classic HLA alleles in 1034 cervical cancer patients and 3948 controls in a Swedish population for an integrated analysis. We found that the protective haplotype *DRB1\*1301-DQA1\*0103-DQB1\*0603* has a direct effect on cervical cancer and always occurs together with the C allele of a *HLA-DRB1 cis*-eQTL (rs9272143), which increases the expression of *HLA-DRB1*. The haplotype rs9272143C-*DRB1\*1301-DQA1\*0103-DQB1\*0603* conferred the strongest protection against cervical cancer (odds ratio [OR] = 0.41, 95% confidence interval [CI] = 0.32–0.52,  $P = 6.2 \times 10^{-13}$ ). On the other hand, the associations with *HLA-B\*0702* and *DRB1\*1501-DQB1\*0602* are attributable to the joint effects of both the *HLA-DRB1 cis*-eQTL (rs9272143) and a frameshift mutation (G insertion of rs67841474, also known as A5.1) of the MHC class I polypeptide-related sequence A gene (*MICA*). Variation in LD between the classic HLA loci, rs9272143 and rs67841474 between populations may explain the different associations of *HLA-B\*07* and *DRB1\*1501-DQB1\*0602* with cervical cancer between studies. The mechanism suggested may also explain similar inconsistent results for other HLA-associated diseases.

**Introduction**

Worldwide, cervical cancer is the third most common cancer among women [1]. Persistent infection with high-risk human papillomavirus (HPV) is the main causal factor of cervical cancer and its precursor lesions, cervical intraepithelial neoplasia (CIN), where CIN III is considered the same as carcinoma in situ (CIS) or Stage 0 cervical cancer [2]. The genetic variability in the host plays an important

role in the persistence of HPV infections and progression to cervical cancer, especially genetic factors that control the immune response [3]. Human leukocyte antigen (HLA) molecules are responsible for the presentation of foreign antigens to the immune system, and play a central role in the immune recognition and subsequent clearance of virally infected cells [4]. Classic HLA alleles, defined by genetic variation within the peptide-binding site of HLA genes, are believed to influence the ability to bind and present

different peptide antigens [5, 6]. Numerous studies have studied the association of specific HLA alleles with risk of (pre) neoplastic cervical disease, but with variable results between studies. Some studies have found the HLA class II *DRB1\*1501* and/or *DQB1\*0602* (two alleles that are in perfect linkage disequilibrium [LD] and often occur on the same haplotype in Europeans) [5, 6] as well as class I *HLA-B\*07* [7–9] to increase risk of cervical disease, while other studies have failed to confirm these associations [10–12] and some have even found inverse associations [8, 13, 14]. In contrast, studies have consistently reported *DRB1\*1301*, *DQA1\*0103*, and/or *DQB1\*0603* (these three alleles are also in perfect LD in Europeans) to reduce risk of cervical disease [5, 6]. Most of these studies have not taken into account the complex LD pattern that extends across multiple HLA and non-HLA genes within the major histocompatibility complex (MHC) region [15], and the need to control rigorously for population stratification. It is, therefore, not clear whether the reported HLA alleles are responsible for the differences in cervical cancer susceptibility, or if functional variants at other locations may explain the effect.

It has recently been shown that single-nucleotide polymorphism (SNP) data within the 6p21 region can be used to impute alleles at key classic HLA class I (*HLA-A*, *HLA-B*, and *HLA-C*) and class II (*HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1*) loci with accuracy that exceeds 90% at the four-digit level [16]. In order to assess the relative contribution of both classic HLA alleles and SNP variation to cervical cancer susceptibility, we used existing SNP data from a previous genome-wide association study (GWAS) of cervical cancer [17] to impute classic HLA alleles in 1034 cervical cancer patients and 3948 controls for an integrated analysis.

## Materials and Methods

### Study population and genotyping

Subjects included in this study were from a previous GWAS of cervical cancer in a Swedish population, and were recruited from two studies, the CervixCan I study and the TwinGene study [17]. In the CervixCan study, cervical cancer patients were selected from families with at least two affected women from Sweden who were born after 1940 and reported to the Swedish Cancer Registry before 1993. Details on study designs and subject recruitment have been described previously [18]. The CervixCan study was further divided into two parts, that is, the CervixCan I study that comprises cases who are the sole participants of their family and the CervixCan II study that comprises individuals with more than one first-degree relative also participating. 766 sole participants (720 CIS

and 46 invasive carcinoma) from the CervixCan I study were included in this study. Direct genotyping of *HLA-B*, *-DQB1*, and *-DRB1* was performed in 576 cervical cancer patients from the CervixCan II study [18]. The TwinGene study is a population-based Swedish study of twins born between 1911 and 1958. In total, 9896 subjects were genotyped consecutively with those from the CervixCan I study using Illumina HumanOmniExpress BeadChip (731,422, SNPs). Among these subjects, 309 unrelated cervical cancer cases (288 CIS and 21 invasive carcinoma) were further included in this study. One female singleton was then randomly selected from each twin pair without cervical cancer, resulting in 4014 unrelated cervical cancer-free females who were included as controls. The details of quality control (QC) have been described previously [17]. Briefly, after stringent QC, data from 1034 cervical cancer patients (971 CIS and 63 invasive carcinoma) and 3948 control subjects were available for 632,668 SNPs with an overall call rate of 99.92%. Utilizing a set of 17,386 SNPs evenly distributed across the genome and in low LD (pairwise  $r^2 < 0.02$ ) that passed QC, principal components analysis (PCA) was performed using EIGENSTRAT package (Broad Institute of Harvard and MIT, New York, USA) [19] to identify population stratification. Nine significant eigenvectors were identified based on the Tracy–Widom statistic ( $P < 0.05$ ) from the PCA.

### Statistical analysis for SNPs

For single-SNP analysis, we considered the extended MHC to be defined by a 7.64 Mb region bordered by the histone cluster 1, H2aa gene (*HIST1H2AA*) and ribosomal protein L12 pseudogene 1 (*RPL12P1*) [20] (rs4711095 at 25,726,774 bp and rs1547668 at 33,775,446 bp, respectively) at the telomeric and centromeric ends of 6p21, respectively, which include 5976 genotyped SNPs in this study. All positions are with respect to the GRCh37.p10/hg19 assembly of the Human Genome. The association between each SNP and risk of cervical cancer was estimated by the odds ratio (OR) per minor allele and 95% confidence interval (CI) using multivariate unconditional logistic regression in allelic test with adjustment for population stratification by including nine informative eigenvectors as covariates, generated by PCA.

### Imputation of classic HLA alleles and statistical analysis

From a reference database of SNP haplotypes carrying known HLA alleles, we imputed HLA genotypes for all the subjects using HLA\*IMP [16]. The reference database combines classic HLA data from the HapMap Project and the 1958 Birth Cohort [16]. Imputation was performed

for three class I (*HLA-A* [ $n = 2474$  reference samples], *-B* [ $n = 3\ 090$ ], *-C* [ $n = 2022$ ]), and three class II (*DQA1* [ $n = 175$ ], *-DQB1* [ $n = 2629$ ], and *-DRB1* [ $n = 2665$ ]) loci. Imputation accuracy was assessed through a cross-validation analysis of the training data. Training was performed on two-thirds of the reference panel with known HLA types and tested on the remaining third at the SNPs chosen for imputation in this study. Thresholding calls at a posterior probability of 0.7 provided call rates of between 0.91 (*HLA-DRB1*) and 0.99 (*HLA-DQB1*) and accuracy of  $\geq 0.95$  for all loci, suggesting that the imputed HLA data are reliable. To ensure the accuracy of prediction, we restricted our case-control analysis to those alleles that were predicted with a posterior probability more than 90%. The association between each HLA allele/haplotype and cervical cancer was then estimated by ORs and 95% CIs using unconditional logistic regression in allelic test. Incorporating these probabilities directly into the general logistic regression framework resulted in similar association results (data not shown). Adjustment for population stratification was performed by including nine informative eigenvectors as covariates in the logistic regression, generated by PCA. Statistical analyses were all performed using SAS 9.3 software (SAS Institute, Cary, NC), with two-sided tests. Pairwise LD was estimated by Haploview [21].

## Results and Discussion

After imputation of classic HLA loci, 27 *HLA-A* alleles, 48 *HLA-B* alleles, 21 *HLA-C* alleles, seven *HLA-DQA1* alleles, 18 *HLA-DQB1* alleles, and 34 *HLA-DRB1* alleles were observed in this study. Association analysis was performed for 5976 SNPs and 155 classic HLA alleles within the extended MHC region, resulting in a significance threshold of  $P = 8.1 \times 10^{-6}$  based on Bonferroni correction.

A total of 161 SNPs were associated with risk of cervical cancer at  $P < 8.1 \times 10^{-6}$ . The strongest signal of association was attained for rs9272143 between *HLA-DRB1* and *HLA-DQA1* (OR = 0.65, 95% CI = 0.59–0.72,  $P = 2.8 \times 10^{-17}$ ). When conditioning upon rs9272143, residual association was detected at some SNPs in this region, with the most significant signal occurring at rs2516448 adjacent to the MHC class I polypeptide-related sequence A gene (*MICA*) (OR = 1.46, 95% CI = 1.32–1.61,  $P = 8.6 \times 10^{-14}$  compared with OR = 1.44, 95% CI = 1.30–1.58,  $P = 5.8 \times 10^{-13}$  in the unconditional analysis). Further conditioning on rs2516448 still left residual association at rs3117027, which is located within the pseudogene *HLA-DPB2* (OR = 1.27, 95% CI = 1.14–1.41,  $P = 7.0 \times 10^{-6}$  compared with OR = 1.28, 95% CI = 1.15–1.42,  $P = 3.1 \times 10^{-6}$  in the unconditional analysis) (Table 1).

The LD between these three SNPs is very weak (pairwise  $r^2 = 0.001$ ), suggesting three independently acting loci. Conditioning on all three SNPs jointly left little evidence for residual association at 6p21.3. The associations with the three loci were further replicated in an independent study including 1140 cervical cancer patients and 1058 controls in the Swedish population [17].

One HLA class I allele (*B\*0702*) and five class II alleles (*DRB1\*1301*, *DRB1\*1501*, *DQA1\*0103*, *DQB1\*0603*, and *DQB1\*0602*) showed evidence of association with cervical cancer at the  $P < 8.1 \times 10^{-6}$  threshold (Table 1). The *DRB1\*1301-DQA1\*0103-DQB1\*0603* haplotype conferred reduced risk of cervical cancer (OR = 0.47, 95% CI = 0.37–0.60,  $P = 8.8 \times 10^{-10}$ ), whereas the *B\*0702* and *DRB1\*1501-DQB1\*0602* haplotype independently conferred increased risk (OR = 1.42, 95% CI = 1.25–1.61,  $P = 7.9 \times 10^{-8}$ ; OR = 1.39, 95% CI = 1.23–1.58,  $P = 3.8 \times 10^{-7}$ , respectively). When conditioning on *B\*0702*, *DRB1\*1301-DQA1\*0103-DQB1\*0603* or *DRB1\*1501-DQB1\*0602*, similar strength of association was observed for rs9272143, rs2516448, and rs3117027 as compared to the unconditional analysis, suggesting that the effects of the three SNP loci are independent from the above HLA alleles/haplotypes [17].

SNP rs9272143 was recently reported to be a *cis*-expression quantitative trait locus (eQTL) that affects the expression of *HLA-DRB1* in adipose tissue from 84 Finnish individuals, with the C allele being associated with increased expression of *HLA-DRB1* ( $P = 4.0 \times 10^{-7}$ ) [22]. The expression of *HLA-DRB1* in adipose tissue is probably due to the presence of macrophages, the antigen presenting cells (APCs), in the inflamed adipose tissue. We cannot, however, rule out that the adipocytes themselves may become more proinflammatory with the *HLA-DRB1* expression increasing allele, as the adipocytes are already known to secrete proinflammatory cytokines [23]. Consistently, rs9271699, which is in perfect LD ( $D' = 1$ ,  $r^2 = 1$ ) with rs9272143 in this study, was also found to be a *cis*-eQTL for *HLA-DRB1* in human lymphoblastoid cell line samples from 373 subjects with European ancestry ( $P = 1.1 \times 10^{-16}$ ) [24], suggesting a role of rs9272143 or its highly correlated variants in regulating *HLA-DRB1* expression. *HLA-DRB1* belongs to the HLA class II beta-chain paralogs, which encodes the  $\beta$ -chain of the peptide-antigen receptor HLA-DR and plays a central role in the cell-mediated immune response by presenting processed foreign antigens to CD4<sup>+</sup> helper T-lymphocytes [4–6]. Impaired HLA class II gene expression has been reported in genital HPV infections and in lesions due to HPV [25, 26]. It is, hence, biologically plausible that carriers of the C allele of rs9272143, which have higher expression level of *HLA-DRB1*, are less susceptible to cervical cancer. However, further studies are warranted to identify the

**Table 1.** Association of SNPs and HLA alleles/haplotypes with risk of cervical cancer.

Variants	Alleles <sup>1</sup>	Allele frequency		Association		
		Case	Control	OR	95% CI	P
SNPs <sup>2</sup>						
rs9272143	T>C	0.39	0.49	0.65	0.59–0.72	$2.8 \times 10^{-17}$
rs2516448 (or rs67841474)	C>T (or →G)	0.59	0.50	1.44	1.30–1.58	$5.8 \times 10^{-13}$
rs3117027	C > A	0.35	0.29	1.28	1.15–1.42	$3.1 \times 10^{-6}$
HLA alleles/haplotypes <sup>3</sup>						
HLA-B						
<i>B*0702</i>	–	0.19	0.14	1.42	1.25–1.61	$7.9 \times 10^{-8}$
HLA-DRB1						
<i>DRB1*1301</i>	–	0.04	0.08	0.47	0.37–0.59	$3.9 \times 10^{-10}$
<i>DRB1*1501</i>	–	0.19	0.15	1.39	1.23–1.57	$2.7 \times 10^{-7}$
HLA-DQA1						
<i>DQA1*0103</i>	–	0.04	0.08	0.49	0.39–0.62	$1.9 \times 10^{-9}$
HLA-DQB1						
<i>DQB1*0603</i>	–	0.04	0.08	0.48	0.38–0.60	$5.7 \times 10^{-10}$
<i>DQB1*0602</i>	–	0.19	0.15	1.39	1.22–1.58	$3.5 \times 10^{-7}$
HLA haplotypes						
<i>DRB1*1301-DQA1*0103-DQB1*0603</i> <sup>4</sup>	–	0.04	0.08	0.47	0.37–0.60	$8.8 \times 10^{-10}$
<i>DRB1*1501-DQB1*0602</i> <sup>5</sup>	–	0.19	0.15	1.39	1.23–1.58	$3.8 \times 10^{-7}$

CI, confidence interval; HLA, human leukocyte antigen; OR, odds ratio; P, two-sided P value corresponding to the OR; SNPs, single-nucleotide polymorphisms.

<sup>1</sup>Major allele > minor allele for each SNP.

<sup>2</sup>Allele frequency corresponds to the minor allele frequency of each SNP, and association results are derived from unconditional logistic regression for each minor allele compared to the wild-type allele with adjustment for the nine informative eigenvectors generated by principal components analysis.

<sup>3</sup>Association results are derived from unconditional logistic regression model for each allele/haplotype compared to all the others with adjustment for the nine informative eigenvectors generated by principal components analysis.

<sup>4</sup>*DRB1\*1301*, *DQA1\*0103*, and *DQB1\*0603* are in strong LD with each other and often occur on the same haplotype (pairwise  $r^2 \geq 0.97$  in the cancer-free controls).

<sup>5</sup>*DRB1\*1501* and *DQB1\*0602* are in strong LD and often occur on the same haplotype (pairwise  $r^2 = 0.98$  in the cancer-free controls).

causal regulatory variation at this locus and evaluate its functional effect in leukocytes and cervical tissues.

Variant rs2516448 is in perfect LD ( $D' = 1$ ,  $r^2 = 1$ ) with a deletion–insertion polymorphism rs67841474, where the guanine (G) insertion causes a frameshift mutation (known as A5.1) in exon 5 of *MICA*, which in turn results in a truncated *MICA* protein lacking part of the transmembrane domain (TMD) and the whole cytoplasmic tail. The risk allele T of rs2516448 always occurs together with the frameshift mutation A5.1. *MICA* encodes a membrane-bound protein which acts as a ligand to stimulate an activating receptor, NKG2D, expressed on the surface of essentially all human natural killer (NK),  $\gamma\delta$  T, and  $CD8^+ \alpha\beta$  T cells [27–29]. Normally, *MICA* is constitutively expressed in low levels on epithelial cells in the gut and thymus, endothelial cells, fibroblasts, and monocytes [30–32], but is upregulated or expressed de novo in stressed conditions, such as during viral and bacterial infections [29, 33, 34], heat shock [22], DNA damage response [35], oncogenic transformation [27, 28], and in autoimmune conditions [36]. *MICA*

serves as signal of cellular stress, and engagement of NKG2D by *MICA* triggers NK cells, and costimulates some  $\gamma\delta$  T cells and antigen-specific  $CD8^+ \alpha\beta$  T cells, resulting in a range of immune effector functions, such as cytotoxicity and cytokine production [30, 37]. The recognition of the *MICA* molecule by the NKG2D receptor enables immune cells to identify and attack infected or transformed cells without the need of MHC class I expression or antigen recognition [38]. Thus, the *MICA*/NKG2D interaction is an effective mechanism for immunosurveillance. The cytoplasmic tail-deleted *MICA*-A5.1 gene product is aberrantly transported to the apical surface of human intestinal epithelial cells instead of the basolateral surface, where the interaction with intraepithelial T and NK lymphocytes takes place [39]. Furthermore, cervical neoplasia patients carrying the A5.1 allele have less membrane-bound *MICA* in their lesions [17], which may comprise their ability to alert the immune system of HPV infection or neoplastic change, leading to impaired immune activation and increased risk of tumor development.

The LD estimates between the associated SNPs and classic HLA alleles are shown in Table 2. Although there is a small difference in the LD pattern between the 576 cervical cancer patients with HLA typing data and the 1034 cervical cancer patients with imputed HLA data, the two groups showed comparable correlations between SNPs and HLA alleles. Correlations were observed between rs9272143 and all the cervical cancer-associated HLA alleles, as well as between rs67841474 and *B\*0702*,

**Table 2.** Linkage disequilibrium (LD) between associated SNPs and HLA alleles.

HLA allele	rs9272143		rs67841474		rs3117027	
	<i>D'</i>	<i>r</i> <sup>2</sup>	<i>D'</i>	<i>r</i> <sup>2</sup>	<i>D'</i>	<i>r</i> <sup>2</sup>
Controls with imputed HLA data <sup>1</sup>						
<i>B*0702</i>	0.50	0.04	0.99	0.16	0.03	0
<i>DRB1*1501</i>	1.0	0.17	0.41	0.03	0.01	0
<i>DQB1*0602</i>	1.0	0.17	0.43	0.03	0.01	0
<i>DRB1*1301</i>	1.0	0.08	0.25	0	0.05	0
<i>DQA1*0103</i>	0.99	0.09	0.23	0	0.05	0
<i>DQB1*0603</i>	1.0	0.09	0.24	0	0.05	0
Cases with imputed HLA data <sup>1</sup>						
<i>B*0702</i>	0.54	0.04	1.0	0.16	0.09	0
<i>DRB1*1501</i>	1.0	0.15	0.49	0.04	0.03	0
<i>DQB1*0602</i>	1.0	0.15	0.51	0.04	0.03	0
<i>DRB1*1301</i>	1.0	0.06	0.15	0	0.03	0
<i>DQA1*0103</i>	1.0	0.06	0.07	0	0.05	0
<i>DQB1*0603</i>	1.0	0.06	0.12	0	0	0
Cases with direct genotyping data of HLA <sup>2</sup>						
<i>B*0702</i>	0.41	0.02	0.96	0.19	0	0
<i>DRB1*1501</i>	0.94	0.12	0.61	0.07	0	0
<i>DQB1*0602</i>	1.0	0.14	0.61	0.07	0	0
<i>DRB1*1301</i>	1.0	0.04	0.28	0	0.48	0
<i>DQB1*0603</i>	1.0	0.05	0.27	0	0.42	0

HLA, human leukocyte antigen; SNP, single-nucleotide polymorphism.

<sup>1</sup>Classic HLA alleles were imputed from SNP data using HLA\*IMP [30].

<sup>2</sup>Estimated in 278 randomly selected singletons stemming from each family in CervixCan II study by Haploview [21] which have overlapping genotyping data of classic HLA alleles and SNPs.

*DRB1\*1501* and *DQB1\*0602* ( $r^2 > 0$ ). To assess the extent of possible confounding of the association with HLA alleles/haplotypes by the SNP associations, conditional logistic regression analysis was performed for these HLA alleles/haplotypes conditioning on the top SNPs in LD with them. As shown in Table 3, upon conditioning on rs9272143 or rs67841474, associations with both *B\*0702* and *DRB1\*1501-DQB1\*0602* were significantly attenuated. When conditioning on both rs9272143 and rs67841474, no statistically significant association was observed for *B\*0702* (conditional analysis: OR = 1.02, 95% CI = 0.88–1.18,  $P = 0.80$ ; unconditional analysis: OR = 1.42, 95% CI = 1.25–1.61,  $P = 7.9 \times 10^{-8}$ ) or *DRB1\*1501-DQB1\*0602* (conditional analysis: OR = 0.99, 95% CI = 0.86–1.15,  $P = 0.90$ ; unconditional analysis: OR = 1.39, 95% CI = 1.23–1.58,  $P = 3.8 \times 10^{-7}$ ), suggesting that the associations of these HLA allele/haplotypes are driven by the joint effects of both rs9272143 and rs67841474. The effect of *DRB1\*1301-DQA1\*0103-DQB1\*0603* was also attenuated upon conditioning on rs9272143, but there was still statistically significant residual association with decreased risk of cervical cancer (conditional analysis: OR = 0.60, 95% CI = 0.45–0.74,  $P = 2 \times 10^{-5}$ ; unconditional analysis: OR = 0.47, 95% CI = 0.37–0.60,  $P = 8.8 \times 10^{-10}$ ), indicating a direct effect on cervical cancer. The class II HLA molecules encoded by this haplotype may have higher affinity of binding HPV antigens than other HLA alleles.

These results provide an explanation of both the conflicting findings with regard to variable pattern of association with *HLA-B\*07* and *DRB1\*1501-DQB1\*0602* and the consistent pattern of association with *DRB1\*1301-DQA1\*0103-DQB1\*0603*. As shown in Table S1, the LD between the HLA loci (*HLA-B\*0702* and *DRB1\*1501*) and the SNPs (rs9272143 and rs67841474) varies between ethnically distinct populations. As *HLA-B\*0702* and *DRB1\*1501-DQB1\*0602* have indirect effects on cervical cancer, variation in the LD pattern between these classic

**Table 3.** Analysis of possible confounding of the HLA allele/haplotypes associations by SNPs.

Conditioning on <sup>1</sup>	Case <sup>2</sup>	Control <sup>2</sup>	<i>B*0702</i>			<i>DRB1*1501-DQB1*0602</i>			<i>DRB1*1301-DQA1*0103-DQB1*0603</i>		
			OR	95% CI	<i>P</i>	OR	95% CI	<i>P</i>	OR	95% CI	<i>P</i>
rs9272143	1034	3948	1.27	1.12–1.45	$3 \times 10^{-4}$	1.13	0.99–1.30	0.08	0.60	0.45–0.74	$2 \times 10^{-5}$
rs67841474	1034	3946	1.20	1.04–1.38	0.01	1.30	1.12–1.46	$2 \times 10^{-4}$	–	–	–
rs9272143 and rs67841474	1034	3946	1.02	0.88–1.18	0.80	0.99	0.86–1.15	0.90	–	–	–

CI, confidence interval; HLA, human leukocyte antigen; SNP, single-nucleotide polymorphism OR, odds ratio; *P*, two-sided *P* value corresponding to the OR.

<sup>1</sup>Derived from conditional logistic regression conditioning on specified SNP for each allele/haplotype compared to all the others with adjustment for the nine informative eigenvectors generated by principal components analysis.

<sup>2</sup>Number of cervical cancer patients and cancer-free controls, respectively.

**Table 4.** Haplotype analysis between rs9272143 and *DRB1\*1301-DQA1\*0103-DQB1\*0603*.

Haplotype	Allele frequency		Association <sup>1</sup>		
	Case	Control	OR	95% CI	P
<i>T-others</i>	0.61	0.51	Reference		
<i>C-others</i>	0.35	0.41	0.70	0.63–0.78	$1.7 \times 10^{-11}$
<i>C-DRB1*1301-DQA1*0103-DQB1*0603</i>	0.04	0.08	0.41	0.32–0.52	$6.2 \times 10^{-13}$

CI, confidence interval; OR, odds ratio.

<sup>1</sup>Derived from unconditional logistic regression for each haplotype compared to *T-others* with adjustment for the nine informative eigenvectors generated by principal components analysis.

HLA alleles and the causal variants across different populations, could result in different strengths as well as direction of associations. By contrast, *DRB1\*1301-DQA1\*0103-DQB1\*0603* has a direct effect and is likely to be pathogenetic. In accordance with the hypothesis that the mechanism of action for causal variants is shared across human populations [40], the biological impact of *DRB1\*1301-DQA1\*0103-DQB1\*0603* on the risk for cervical cancer appears to be consistent across populations.

The *HLA-DRB1 cis*-eQTL (rs9272143) and *DRB1\*1301-DQA1\*0103-DQB1\*0603* are both located in the MHC class II region and are in complete LD with each other ( $D' = 1$ ), resulting in three haplotypes between these two loci (Table 4). *DRB1\*1301-DQA1\*0103-DQB1\*0603* only occurs together with the protective allele C of rs9272143, which has recently been shown to be associated with increased expression of *HLA-DRB1* [22, 24]. The haplotype *C-DRB1\*1301-DQA1\*0103-DQB1\*0603* showed the strongest protection against cervical cancer, with carriers having 59% reduced risk as compared with those carrying the most common haplotype *T-others* (OR = 0.41, 95% CI = 0.32–0.52,  $P = 6.2 \times 10^{-13}$ ). This suggests that the class II HLA molecules enclosed by this haplotype have better specificity of binding HPV antigens and are also expressed at a higher level as compared to other alleles, and carriers may therefore benefit from a more efficient presentation of peptides from the causative agent of cervical cancer, HPV.

Some limitation in this study should be noted. First, most of the cervical cancer cases included are CIS and only 4.5% are invasive squamous cell carcinoma. We therefore have higher power to detect variants associated with CIS and only limited power to identify associations with invasive carcinoma. Our results apply to CIS, which represent the vast majority of cases, and progression to invasive cancer may require additional, as yet unknown, risk factors. However, CIS and invasive cancer share the same main etiological factor, namely persistent infection by oncogenic types of HPV, indicating that the genetic susceptibility loci identified for CIS will have a similar

effect on the two cancer stages. Second, our study did not have information on HPV type. Hence, we are not able to assess if the effect is specific for certain HPV types.

This study was motivated by the variable pattern of association of specific HLA alleles with cervical cancer reported in the literature. Using PCA allowed us to minimize the problem of population stratification. We have found that the associations of cervical cancer risk with *B\*0702* and *DRB1\*1501-DQB1\*0602* are attributable to the joint effects of a *HLA-DRB1 cis*-eQTL and a frame-shift mutation of *MICA*. The mechanism suggested may also explain similar inconsistent results seen between populations for other HLA-associated diseases. On the other hand, the protective haplotype *DRB1\*1301-DQA1\*0103-DQB1\*0603* has a direct effect on cervical cancer risk and is expressed at a higher level compared to other classic HLA alleles. Future studies are warranted to identify the specific viral epitopes presented by this haplotype.

## Acknowledgments

The authors thank all of the participants in this research and support staff members who made this study possible.

## Conflict of Interest

None declared.

## References

- Arbyn, M., X. Castellsagué, S. de Sanjosé, L. Bruni, M. Saraiya, F. Bray, et al. 2011. Worldwide burden of cervical cancer in 2008. *Ann. Oncol.* 22:2675–2686.
- Bosch, F. X., A. Lorincz, N. Muñoz, C. J. Meijer, and K. V. Shah. 2002. The causal relation between human papillomavirus and cervical cancer. *J. Clin. Pathol.* 55:244–265.
- Magnusson, P. K., P. Sparén, and U. B. Gyllensten. 1999. Genetic link to cervical tumours. *Nature* 400:29–30.
- Paul, W. E., ed. 1999. *Fundamental immunology*. 4th ed. Lippincott-Raven, Philadelphia, PA.

5. Hildesheim, A., and S. S. Wang. 2002. Host and viral genetics and risk of cervical cancer: a review. *Virus Res.* 89:229–240.
6. Zoodmsa, M., I. M. Nolte, G. J. Te Meerman, E. G. De Vries, and A. G. Van der Zee. 2005. HLA genes and other candidate genes involved in susceptibility for (pre) neoplastic cervical disease. *Int. J. Oncol.* 26:769–784.
7. Ellis, J. R., P. J. Keating, J. Baird, E. F. Hounsell, D. V. Renouf, M. Rowe, et al. 1995. The association of an HPV16 oncogene variant with HLA-B7 has implications for vaccine design in cervical cancer. *Nat. Med.* 1:464–470.
8. Hildesheim, A., M. Schiffman, D. R. Scott, D. Marti, T. Kissner, M. E. Sherman, et al. 1998. Human leukocyte antigen class I/II alleles and development of human papillomavirus-related cervical neoplasia: results from a case–control study conducted in the United States. *Cancer Epidemiol. Biomarkers Prev.* 7:1035–1041.
9. Wang, S. S., C. M. Wheeler, A. Hildesheim, M. Schiffman, R. Herrero, M. C. Bratti, et al. 2001. Human leukocyte antigen class I and II alleles and risk of cervical neoplasia: results from a population-based study in Costa Rica. *J. Infect. Dis.* 184:1310–1314.
10. Montoya, L., I. Saiz, G. Rey, F. Vela, and N. Clerici-Larradet. 1998. Cervical carcinoma, human papillomavirus infection and HLA-associated risk factors in the Spanish population. *Eur. J. Immunogenet.* 25:329–337.
11. Lin, P., L. A. Koutsky, C. W. Critchlow, R. J. Apple, S. E. Hawes, J. P. Hughes, et al. 2001. HLA class II DR-DQ and increased risk of cervical cancer among Senegalese women. *Cancer Epidemiol. Biomarkers Prev.* 10:1037–1045.
12. Gregoire, L., W. D. Lawrence, D. Kukuruga, A. B. Eisenbrey, and W. D. Lancaster. 1994. Association between HLA-DQB1 alleles and risk for cervical cancer in African–American women. *Int. J. Cancer* 57:504–507.
13. Ferrera, A., A. Olivo, C. Alaez, W. J. Melchers, and C. Gorodezky. 1999. HLA DQA1 and DQB1 loci in Honduran women with cervical dysplasia and invasive cervical carcinoma and their relationship to human papillomavirus infection. *Hum. Biol.* 71:367–379.
14. Ades, S., A. Koushik, E. Duarte-Franco, N. Mansour, J. Arseneau, D. Provencher, et al. 2008. Selected class I and class II HLA alleles and haplotypes and risk of high-grade cervical intraepithelial neoplasia. *Int. J. Cancer* 122:2820–2826.
15. de Bakker, P. I., G. McVean, P. C. Sabeti, M. M. Miretti, T. Green, J. Marchini, et al. 2006. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* 38:1166–1172.
16. Dilthey, A. T., L. Moutsianas, S. Leslie, and G. McVean. 2011. HLA\*IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* 27:968–972.
17. Chen, D., I. Juko-Pecirep, J. Hammer, E. Ivansson, S. Enroth, I. Gustavsson, et al. 2013. Genome-wide association study of susceptibility loci for cervical cancer. *J. Natl. Cancer Inst.* 105:624–633.
18. Engelmark, M., A. Beskow, J. Magnusson, H. Erlich, and U. Gyllensten. 2004. Affected sib-pair analysis of the contribution of HLA class I and class II loci to development of cervical cancer. *Hum. Mol. Genet.* 13:1951–1958.
19. Price, A. L., N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38:904–909.
20. Horton, R., L. Wilming, V. Rand, R. C. Lovering, E. A. Bruford, V. K. Khodiyar, et al. 2004. Gene map of the extended human MHC. *Nat. Rev. Genet.* 5:889–899.
21. Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265.
22. Laurila, P. P., I. Surakka, A. P. Sarin, L. Yetukuri, T. Hyötyläinen, S. Söderlund, et al. 2013. Genomic, transcriptomic, and lipidomic profiling highlights the role of inflammation in individuals with low high-density lipoprotein cholesterol. *Arterioscler. Thromb. Vasc. Biol.* 33:847–857.
23. Catalán, V., J. Gómez-Ambrosi, A. Rodríguez, and G. Frühbeck. 2013. Adipose tissue immunity and cancer. *Front Physiol.* 4:275.
24. Lappalainen, T., M. Sammeth, M. R. Friedländer, P. A. 't Hoen, J. Monlong, M. A. Rivas, et al. 2013. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 501:506–511.
25. Coleman, N., H. D. Birley, A. M. Renton, N. F. Hanna, B. K. Ryaite, M. Byrne, et al. 1994. Immunological events in regressing genital warts. *Am. J. Clin. Pathol.* 102:768–774.
26. Morelli, A. E., G. Belardi, G. DiPaola, A. Paredes, and L. Fainboim. 1994. Cellular subsets and epithelial ICAM-1 and HLA-DR expression in human papillomavirus infection of the vulva. *Acta Derm. Venereol.* 74:45–50.
27. Bauer, S., V. Groh, J. Wu, A. Steinle, J. H. Phillips, L. L. Lanier, et al. 1999. Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* 285:727–729.
28. Groh, V., R. Rhinehart, H. Secrist, S. Bauer, K. H. Grabstein, and T. Spies. 1999. Broad tumor-associated expression and recognition by tumor-derived  $\gamma\delta$  T cells of MICA and MICB. *Proc. Natl. Acad. Sci. USA* 96:6879–6884.
29. Groh, V., R. Rhinehart, J. Randolph-Habecker, M. S. Topp, S. R. Riddell, and T. Spies. 2001. Costimulation of CD8  $\alpha\beta$  T cells by NKG2D via engagement by MIC induced on virus-infected cells. *Nat. Immunol.* 2:255–260.
30. Bahram, S. 2000. MIC genes: from genetics to biology. *Adv. Immunol.* 76:1–60.
31. Groh, V., S. Bahram, S. Bauer, A. Herman, M. Beauchamp, and T. Spies. 1996. Cell stress-regulated

- human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc. Natl. Acad. Sci. USA* 93:12445–12450.
32. Zwirner, N. W., M. A. Fernández-Viña, and P. Stastny. 1998. MICA, a new polymorphic HLA-related antigen, is expressed mainly by keratinocytes, endothelial cells, and monocytes. *Immunogenetics* 47:139–148.
  33. Das, H., V. Groh, C. Kuijl, M. Sugita, C. T. Morita, T. Spies, et al. 2001. MICA engagement by human V $\gamma$ 2V $\delta$ 2 T cells enhances their antigen-dependent effector function. *Immunity* 15:83–93.
  34. Tieng, V., C. Le Bouguéneq, L. du Merle, P. Bertheau, P. Desreumaux, A. Janin, et al. 2002. Binding of *Escherichia coli* adhesin AfaE to CD55 triggers cell-surface expression of the MHC class I-related molecule MICA. *Proc. Natl. Acad. Sci. USA* 99:2977–2982.
  35. Gasser, S., S. Orsulic, E. J. Brown, and D. H. Raulet. 2005. The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* 436:1186–1190.
  36. Groh, V., A. Bruhl, H. El-Gabalawy, J. L. Nelson, and T. Spies. 2003. Stimulation of T cell autoreactivity by anomalous expression of NKG2D and its MIC ligands in rheumatoid arthritis. *Proc. Natl. Acad. Sci. USA* 100:9452–9457.
  37. Girardi, M., D. E. Oppenheim, C. R. Steele, J. M. Lewis, E. Glusac, R. Filler, et al. 2001. Regulation of cutaneous malignancy by  $\gamma\delta$  T cells. *Science* 294:605–609.
  38. Diefenbach, A., and D. H. Raulet. 2001. Strategies for target cell recognition by natural killer cells. *Immunol. Rev.* 181:170–184.
  39. Suemizu, H., M. Radosavljevic, M. Kimura, S. Sadahiro, S. Yoshimura, S. Bahram, et al. 2002. A basolateral sorting motif in the MICA cytoplasmic tail. *Proc. Natl. Acad. Sci. USA* 99: 2971–2976.
  40. Ioannidis, J. P., E. E. Ntzani, and T. A. Trikalinos. 2004. ‘Racial’ differences in genetic effects for complex diseases. *Nat. Genet.* 36:1312–1318.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Linkage disequilibrium between HLA alleles and associated SNPs in individuals of CEU, YRI, JPT, and CHB.