Check for updates

Received: 19 February 2020 | Revised: 15 May 2020 | Accepted: 25 May 2020

DOI: 10.1002/ijc.33171

CANCER GENETICS AND EPIGENETICS

IJC International Journal of Cancer

Association of genomic variants at the human leukocyte antigen locus with cervical cancer risk, HPV status and gene expression levels

Dhanya Ramachandran[1] | Peter Schürmann[1] | Qianqian Mao[1] | Yingying Wang[1] | Lisa-Marie Bretschneider[1] | Lisa-Marie Speith[1] | Fabienne Hülse[1] | Julia Enßen[1] | Kristine Bousset[1] | Matthias Jentschke[1] | Gerd Böhmer[2] | Hans-Georg Strauß[3] | Christine Hirchenhain[4] | Monika Schmidmayr[5] | Johanna Tarbiat[6] | Ingo Runnebaum[7] | Matthias Dürst[7] | Alexander Hein[8] | Martin Koch[8] | Matthias Ruebner[8] | Arif Ekici[9] | Matthias W. Beckmann[8] | Peter A. Fasching[8] | Alexander Luyten[10,11] | Karl-Ulrich Petry[11] | Peter Hillemanns[1] | Thilo Dörk[1]

[1]Department of Gynaecology, Comprehensive Cancer Center, Hannover Medical School, Hannover, Germany

[2]IZD Hannover, Hannover, Germany

[3]Department of Gynaecology, University Clinics, Martin-Luther University, Halle-Wittenberg, Germany

[4]Department of Gynaecology, Clinics Carl Gustav Carus, University of Dresden, Dresden, Germany

[5]Department of Gynaecology, Technische Universität München, Munich, Germany

[6]Martin-Luther Hospital, Charite University, Berlin, Germany

[7]Department of Gynecology, Jena University Hospital, Friedrich-Schiller-University Jena, Jena, Germany

[8]Department of Gynecology and Obstetrics, Erlangen University Hospital, Comprehensive Cancer Center Erlangen-EMN, Friedrich Alexander University of Erlangen–Nuremberg (FAU), Erlangen, Germany

[9]Institute of Human Genetics, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany

[10]Dysplasia Unit, Department of Gynecology and Obstetrics, Mare Klinikum, Kronshagen, Germany

[11]Department of Gynecology, Wolfsburg Hospital, Wolfsburg, Germany

Correspondence
Thilo Dörk, Hannover Medical School, Gynaecology Research Unit, Carl-Neuberg-Street 1, 30625 Hannover, Germany.
Email: doerk.thilo@mh-hannover.de

Funding information
Bruno and Helene Jöster Foundation

Abstract

The human leukocyte antigen (*HLA*) locus on chromosome 6 has been reported to be associated with cervical cancer. We investigated two independent single-nucleotide polymorphisms in a large case-control series of cervical dysplasia and carcinoma that has been newly established by the German Cervigen Consortium, comprising a total of 2481 cases and 1556 healthy females. We find significant associations for both variants, rs9272117 at *HLA-DQA1* and rs2844511 at *MICA* and *HCP5*, with cervical

Abbreviations: cDNA, complementary DNA; CI, confidence interval; CIN, cervical intraepithelial neoplasia; CIS, carcinoma in situ; EDTA, ethylenediaminetetraacetic Acid; eQTL, expression quantitative trait loci; GWAS, genome-wide association study; HLA, human leukocyte antigen; HPV, human papillomavirus; HWE, Hardy-Weinberg equilibrium; ICC, invasive cervical cancer; LSIL/HSIL, low/high grade intraepithelial neoplasia; MAF, minor allele frequency; MHC, major histocompatibility complex; NK, natural killer cells; OR, odds ratio; qRT-PCR, quantitative real-time PCR; SNP, single nucleotide polymorphism.

[†]Karl-Ulrich Petry has deceased on April 20, 2020

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2020 The Authors. *International Journal of Cancer* published by John Wiley & Sons Ltd on behalf of UICC.

disease. Both variants showed evidence of association with invasive cervical cancer (rs9272117: OR 0.89, 95% CI 0.79-0.99, P = .036; rs2844511: OR 1.17, 95% CI 1.04-1.31, P = .008) and with high-grade dysplasia (rs9272117: OR 0.78, 95% CI 0.70-0.87, $P = 7.1 \times 10^{-6}$; rs2844511: OR 1.13, 95% CI 1.01-1.26, P = .035), as well as in a combined analysis of both groups (rs9272117: OR 0.83, 95% CI 0.75-0.91, $P = 6.9 \times 10^{-5}$; rs2844511: OR 1.14, 95% CI 1.04-1.26, P = .005). Variant rs2844511, but not rs9272117, also showed modest evidence of association with low-grade dysplasia (OR 1.26, 95% CI 1.04-1.54, P = .019). In case-only analyses, rs2844511 tended to predict HPV status (P = .044) and rs9272117 tended to associate with HPV16 (P = .022). RNA studies in cervical samples showed a significant correlation in the transcript levels of *MICA*, *HCP5* and *HLA-DQA1*, suggesting extensive co-regulation. All three genes were upregulated in HPV16-positive samples. In stratified analyses, rs9272117 was associated with *HLA-DQA1* levels, specifically in HPV-positive samples, while rs2844511 was associated with *MICA* and *HCP5* levels. The risk allele of rs2844511 was required for correlations between *MICA* or *HCP5* with *HLA-DQA1*. Altogether, our results support 6p21.32-33 as the first consistent cervical cancer susceptibility locus and provide evidence for a link between genetic risk variants, HPV16 status and transcript levels of *HLA-DQA1*, *HCP5* and *MICA*, which may contribute to tumor immune evasion.

KEYWORDS

association study, cervical malignancy, eQTL, HPV infection, single nucleotide polymorphism

1 | INTRODUCTION

Cervical cancer is the fourth most prevalent form of cancer in females worldwide and is the third leading cause of cancer death among women of the age group 15 to 44 years in Germany, highlighting the importance of early screening as well as preventive human papillomavirus (HPV) vaccination.[1,2] Among the high-risk HPV subtypes, 16 and 18 are traditionally regarded to be the major triggers of cervical cancer in Europeans.[3] However, a growing body of evidence suggests that the oncogenic potential of HPV 31, 33, 52 and 58 is at least similar to HPV 18.[4] HPV infection of the cervical epithelium may result in cervical intraepithelial neoplasia 1 (CIN1, or low-grade intraepithelial lesions LSIL).[5] This infection is known to resolve and those infected with HPV most likely will never develop cervical cancer.[6] However, some women with persistent infection of high-risk HPV go on to develop cervical cancer and the molecular drivers of this progression are incompletely understood. Age, multiple pregnancies, environmental factors such as smoking and use of oral contraceptives can contribute to increasing the risk of developing cervical cancer. Importantly, there is an increased familial relative risk for cervical cancer,[7,8] and heritability was found to contribute 27% to 36% toward the disease risk.[9,10] Hereditary genetic factors may affect host-pathogen interactions, impact the severity of infection and contribute to the development of tumors.[9,11]

The human leukocyte antigen (HLA) locus controls several host-pathogen interactions and has been proposed as a cervical cancer susceptibility locus in different populations.[12-16] A cervical cancer genome-wide association study (GWAS) in the Swedish population implicated rs9272143, located at the *HLA-DQA1* gene and rs2516448, located downstream of the MHC class I polypeptide-related sequence A (MICA), with increased susceptibility to cervical cancer.[12,13] It has remained largely unknown which genes or functional consequences underlie this association.

What's new?

The HLA locus, on chromosome 6, appears to contribute to cervical cancer risk, according to some studies. Genome wide association studies have uncovered two candidate variants on that chromosome: rs2844511, at the *HLA-DQ1* gene, and rs9272117, located in an intron of lncRNA *HCP5*, downstream of the *MICA* gene. Here, the authors tested the contribution of the two variants through genotyping of 2500 cervical cancer cases and 1500 healthy controls. Both variants showed an association with cervical cancer and with high-grade dysplasia, and they modulated the expression levels of nearby genes in cervical cells, providing evidence for the first cervical cancer susceptibility locus.

In the present study, we investigated the two variants rs2844511 and rs9272117 through direct genotyping. These are unlinked variants located about 1.2 MBp apart from each other. The variant rs9272117 is located at an enhancer region within *HLA-DQA1* (extended transcript *HLA-DQA1-204*, ENST00000422863.1 in Ensembl), whereas variant rs2844511 is located in an intron of lncRNA *HCP5* (extended transcript *HCP5-201*, ENST00000414046.2 in Ensembl), downstream of *MICA*. These variants have been associated with a wide range of immune diseases,[17-19] and the variant rs2844511 had been implicated in a smaller non-European multi-ethnic case-control study on cervical cancer.[20] As rs2844511 is in linkage disequilibrium with rs2516448 (r^2 = 1), and rs9272117 strongly correlates with rs9272143 (r^2 = 0.9), they could be directly used for candidate replication of the former GWAS results. We performed a genetic association study for these two variants in a large case-control series for cervical cancer and dysplasias that had been newly established by the German Cervigen consortium, and we additionally tested the hypothesis that these variants may regulate the RNA transcript levels of adjacent genes in nonmalignant cervical tissue.

2 | MATERIALS AND METHODS

2.1 | Patients

In total, 4037 samples from the German Cervigen Study were used for the present case-control analysis. This included 1042 cases with invasive cervical cancer and 1405 cases with cervical dysplasia (1045 CIN3, 227 CIN2, 133 CIN1) recruited from nine German hospitals in Hannover, Wolfsburg, Jena, Erlangen, Dresden, Halle, Munich, Berlin and Bad Münder. For comparison, a total of 1556 healthy female controls were recruited from the two centers in Hannover and Erlangen (1040 blood donors from Hannover, 516 blood donors from Erlangen). A detailed list of samples per center, stratified by histology, is provided in Table S1. Median age at diagnosis was 44 years (range 17-94 years) for patients with invasive cervical cancer and 31 years (range 16-79 years) for patients with cervical dysplasia, compared to a median age at recruitment of 32 years (range 18-86 years) for healthy female controls. Histology data were available for 2413 patients (98.6%). HPV status had been documented for 1356 patients (55.4%). Across the severity groups, HPV positivity rates were 54% for CIN1, 81% for CIN2, 88% for CIN3 and 97% for invasive cervical cancer.

Our study was approved by the Ethics committee of Hannover Medical School (Vote No. 441) and the patient samples obtained and data used were in accordance with German medical council regulations. Informed consent was obtained from all participants, 5 mL peripheral venous EDTA blood samples were taken for genomic DNA extraction, and methanol-fixed cervical tissue smears were obtained from a smaller cohort of healthy participants.

In our cohort of 261 cervical tissue smear samples, 73 were found to be HPV positive and 188 were HPV negative. Further stratification by HPV subtype shows that 31 samples were HPV16 positive, 8 samples were HPV18 positive and 34 samples were infected by other strains of HPV. Of the 188 women with HPV-negative status, 169 had no sign for a cytologically or histologically detectable lesion (r^2 = 0.39). This "$HPV^{-}Lesion^{-}$" subgroup was analyzed separately in comparison with the "$HPV^{+}Lesion^{+}$" subgroup that only contained HPV positive samples with a documented lesion (CIN1-3 or PAPIII+, n = 53).

2.2 | SNP genotyping

Genomic DNA was extracted from peripheral white blood cells using standard phenol-chloroform extraction. Fluidigm SNPtype assays were used for SNP genotyping for the three variants that were initially chosen for this analysis: rs2844511, rs9272117 and rs9272146 (Fluidigm SNPtype assay ID: GTA0087749, GTA0215524 and GTA0215526). All three variants are pyrimidine transitions with MAFs between 25% and 49% (Table S2). The allele-specific probes were labeled with FAM or HEX dyes. Two samples without template served as negative controls.

Call rates were 98.8% for rs2844511, 98.4% for rs9272117 and 98.3% for rs9272146. However, our subsequent quality check identified a deviation from Hardy-Weinberg equilibrium for rs9272146, and this variant was therefore excluded from further analysis. As rs9272117 is in strong linkage disequilibrium with rs9272146, this signal at *HLA-DQA1* was still represented in our study.

2.3 | Statistical analysis

Goodness-of-fit chi-square tests were used to determine whether the candidate SNPs were in Hardy-Weinberg equilibrium (HWE). For the two variants fulfilling HWE, we carried out logistic regression analyses to calculate odds ratios (ORs), *P* values and 95% confidence intervals (CIs) under an additive model, with case-control status as the outcome and variant genotype as the predictor variables, using the STATA12 software package. Odds ratios are given relative to the common homozygous (CC) genotypes for both variants. We restricted all analyses to study participants with questionnaire-based European ancestry. We further performed stratified analyses for the invasive, high-grade dysplasia or low-grade dysplasia case groups in comparison with all controls. For this purpose, we grouped CIN1 patients together with CIN2 cases at age < 30 years ($CIN2_{<30}$), and CIN2 cases at age ≥ 30 years ($CIN2_{\geq 30}$) together with CIN3 patients. A combined analysis was performed for CIN2 at age ≥ 30 together with CIN3 and with invasive cases. We also stratified by HPV status and performed case-only logistic regression analyses with overall HPV status, HPV16 or HPV18 status as the outcome and variant genotype as the predictor variables. Two-sided *P* values below .01 were considered significant in the main analyses and two-sided *P* values below .005 were considered significant in the subgroup analyses.

2.4 | Transcript analysis

Total RNA was extracted from methanol-fixed cervical tissue smear samples of 261 healthy females who underwent routine HPV testing

at Hannover Medical School using guanidinium-phenol-chloroform based extraction with Trizol reagent (peqGOLD TriFast). In parallel, genomic DNA was extracted from these samples via magnetic beads using the M24 SP robot (Abbott) for SNP genotyping. A total of 1ug RNA per sample was reverse transcribed into cDNA using the ProtoScript II First Strand cDNA Synthesis Kit (New England BioLabs, Ipswich, Massachusetts). Fluidigm DeltaGene assays were obtained for the genes *MICA*, *HLA-DQA1*, *HCP5*, *B2M* and *RPL13A* (Assay ID GEP00097324, GEP00097327, GEP00097325, GEP00055772 and GEP00091292). The cDNA was first preamplified using pooled DeltaGene assays and the Pre-amp Master Mix (Fluidigm, South San Francisco, California). After purification, 2X SsoFast EvaGreen Supermix with low ROX (BioRad, Hercules, California) was used as the binding dye for determining gene amplification curves and multiple qPCR reactions were carried out in 48 × 48 GE integrated fluidic circuit (IFC) plates on a BioMark HD real-time PCR instrument (Fluidigm). qBASE+ (Biogazelle, Gent, Belgium)[21,22] and GeNorm[23] were used to check the stability of housekeeping genes and calculate relative gene quantities taking *B2M* and *RPL13A* as housekeeping controls.

Outliers were identified and excluded from each dataset with the ROUT method (1% false discovery rate) on Graphpad Prism version 8.4.2, before performing any further statistical analysis. Pearson correlation coefficients were then calculated for pairwise combinations of calculated relative gene quantities using GraphPad Prism version 8.4.2. Association of HPV status with gene expression was investigated and stratified analysis was performed for high-risk subtypes 16 and 18. To investigate whether rs2844511 and rs9272117 are eQTLs for *MICA*, *HLA-DQA1* or *HCP5*, relative gene quantities were further analyzed for association with the genotypes of the two candidate variants in the corresponding genomic DNA samples and examined under the allelic, dominant and recessive models of inheritance for each SNP-gene pair, as well as after HPV status based stratification. Student's *t* test was performed to compare two groups, whereas ANOVA was performed to compare three or more groups. A *P* value < .05 was considered noteworthy, A selected region of *MICA* containing two coding variants, rs1140700 and rs1063635, was amplified from cDNA of 55 samples using primers spanning an exon-exon junction (forward primer: 5′-TCAGACCTTGGCCATGAACG-3′; reverse primer: 5′-GTGCTGTGATTCCCGCTGTG-3′) and sequenced using BigDye Sanger sequencing on a SeqStudio Genetic Analyzer (Applied Biosystems, Foster City, California). The sequences were subsequently visualized using FinchTV 1.4.0 (Geospiza, Inc., Seattle, Washington; http://www.geospiza.com).

TABLE 2 Association of rs2844511 and rs9272117 with HPV status (case-only analysis)

	rs2844511		rs9272117	
HPV	OR (95% CI)	P_{reg}*	OR (95% CI)	P_{reg}*
HPV16+ve	0.84 (0.65-1.09)	.189	0.74 (0.58-0.96)	.022
HPV18+ve	1.05 (0.78-1.42)	.743	1.05 (0.78-1.40)	.766
HPV total	0.79 (0.63-0.99)	.044	0.95 (0.76-1.19)	.643

Note: Stratified regression analysis of rs2844511 and rs9272117 with different HPV subtypes. Results are shown from a case-only analysis of patients with known HPV status. HPV status is shown overall and for the two main defined subgroups HPV16 and HPV18. Overall HPV status, HPV16 status or HPV18 status served as the outcome and variant genotype as the predictor variable, respectively.

Abbreviations: CI, confidence interval; OR, odds ratio with negative HPV status as the reference; P_{reg}, *P* value after logistic regression analysis restricted to cases with known HPV status.

TABLE 1 Association of rs2844511 and rs9272117 with case-control status

	rs2844511				rs9272117			
Stratum	n_{cases}	$n_{controls}$	OR (95% CI)	P_{reg}	n_{cases}	$n_{controls}$	OR (95% CI)	P_{reg}
CIN1 + $CIN2_{< 30}$	239	1436	1.26 (1.04-1.54)	.018	239	1431	1.09 (0.91-1.32)	.360
HPV−ve	65	1436	1.48 (1.04-2.12)	.029	65	1431	1.25 (0.88-1.75)	.211
HPV+ve	131	1436	1.18 (0.91-1.52)	.211	131	1431	1.00 (0.78-1.28)	1.000
$CIN2_{\geq 30}$ + CIN3	1151	1436	1.13 (1.01-1.26)	.034	1149	1431	0.78 (0.70-0.87)	7.1×10^{-6}
HPV−ve	93	1436	1.44 (1.07-1.94)	.016	93	1431	0.79 (0.59-1.05)	.101
HPV+ve	575	1436	1.12 (0.98-1.29)	.097	575	1431	0.82 (0.72-0.94)	.004
Invasive	1039	1436	1.17 (1.04-1.31)	.008	1038	1431	0.89 (0.79-0.99)	.036
SC	655	1436	1.19 (1.05-1.36)	.009	653	1431	0.84 (0.74-0.95)	.008
AC	180	1436	1.14 (0.91-1.42)	.244	180	1431	1.08 (0.87-1.34)	.473
$CIN2_{\geq 30}$ + CIN3 + Invasive	2190	1436	1.14 (1.04-1.26)	.005	2187	1431	0.83 (0.75-0.91)	6.9×10^{-5}

Note: Stratified analysis of rs2844511 and rs9272117 in the Cervigen cohort, *P* values were generated after stratified logistic regression analyses restricted to the disease subtype. Cervical intraepithelial neoplasia was differentiated into low-risk (CIN1 + $CIN2_{<30}$) and high-risk ($CIN2_{\geq 30}$ + CIN3) groups. Both groups were further stratified by documented HPV status into HPV-negative (HPV−ve) or HPV-positive (HPV+ve) subgroups. Invasive cervical cancer was further stratified into squamous epithelial cell carcinoma (SC) or adenocarcinoma (AC). This group was not stratified by HPV status as the number of HPV-negative invasive cancers was small. High-risk dysplasia ($CIN2_{\geq 30}$ + CIN3) and invasive cancer were further combined for joint analysis.

Abbreviations: CI, confidence interval; n_{cases}, number of successfully genotyped cases per stratum; $n_{controls}$, number of successfully genotyped controls; OR, odds ratio for minor allele with healthy controls as reference group; P_{reg}, *P* value from logistic regression analysis of cases within the stratum compared to all controls.

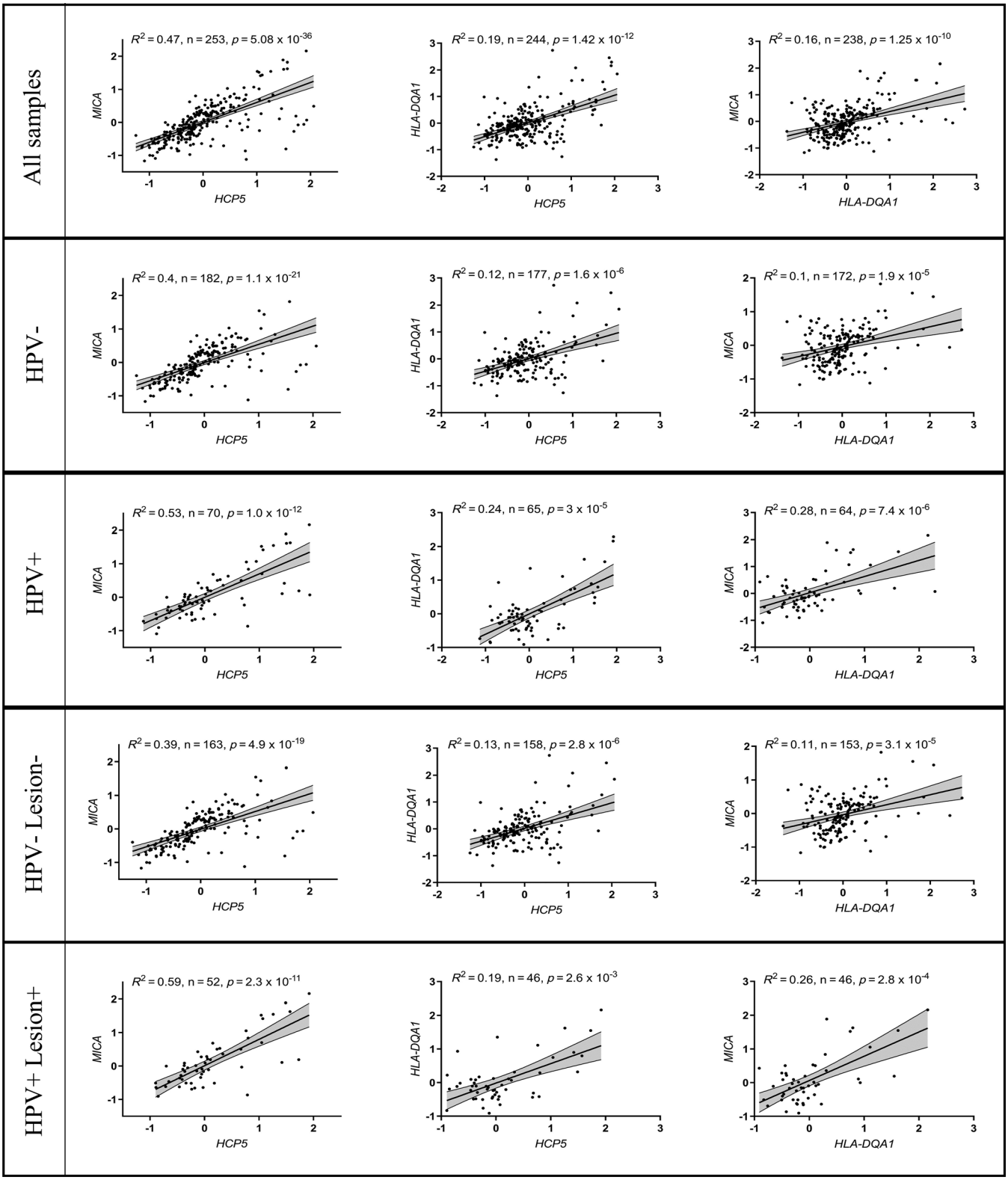

FIGURE 1 Correlation of transcript levels between candidate genes. Pearson correlation coefficients and *P* values for regression were generated to compare the transcript abundance for each gene pair in the cDNA cohort. *From left to right*: *MICA* and *HCP5*; *HLA-DQA1* and *HCP5*; and *MICA* and *HLA-DQA1*. *Panels from top to bottom*: All samples, HPV negative samples, HPV positive samples, HPV negative samples with no lesion, HPV positive samples with documented lesion

3 | RESULTS

The two variants rs2844511 at the *MICA* locus and rs9272117 at the *HLA-DQA1* locus were successfully genotyped in the Cervigen case-control series with call rates of 97.5% and 97.3%, respectively, and fairly good clustering (Figure S1). A third genotyped variant, rs9272146, was found to be out of Hardy-Weinberg equilibrium and hence excluded from further analysis. After excluding females of known non-European descent, we obtained genotypes for rs2844511 from 3936 individuals (2500 cases and 1436 controls) and for rs9272117 from 3928 individuals (2497 cases and 1431 controls).

Both variants showed significant associations with cervical disease status (Table 1). When cases were stratified by disease severity, there was evidence of association with invasive cervical cancer for both variants (rs9272117: OR 0.89, 95% CI 0.79-0.99, P = .036; rs2844511: OR 1.17, 95% CI 1.04-1.31, P = .008). This association was significant for squamous cervical cancer but was not detected for adenocarcinomas (Table 1). There was also evidence for association with high-grade dysplasia (HSIL) for both variants when $CIN2_{\geq 30}$ and CIN3 were combined (rs9272117: OR 0.78, 95% CI 0.70-0.87, $P = 7.1 \times 10^{-6}$; rs2844511: OR 1.13, 95% CI 1.01-1.26, P = .035). Variant rs2844511, but not rs9272117, also showed modest evidence of association with low-grade dysplasia (OR 1.26, 95% CI 1.04-1.54, P = .019). In a combined analysis of $CIN2_{\geq 30}$, CIN3 and invasive cervical cancer, we found significant association for both variants (rs9272117: OR 0.83, 95% CI 0.75-0.91, $P = 6.9 \times 10^{-5}$; rs2844511: OR 1.14, 95% CI 1.04-1.26, P = .005).

From the results in Table 1, rs2844511 appeared to associate more strongly with risk of HPV-negative dysplasia in both, the low-grade and high-grade dysplasia groups, when compared to all controls (Table 1). In contrast, rs9272117 tended to have a stronger impact on HPV-positive high-grade dysplasia (Table 1). We performed case-only analyses, using HPV negative status as the baseline, and found borderline evidence that rs2844511 may be associated with overall HPV status (P = .044) and rs9272117 may specifically associate with HPV16 status (P = .022; Table 2). The minor alleles of these variants predicted HPV below detection. However, none of these results remained significant after correction for multiple testing and we had insufficient power to further stratify by histology groups.

To obtain more insight into the potential regulatory function of these two variants, we tested the expression of the three adjacent candidate genes at this locus (*MICA*, *HLA-DQA1* and *HCP5*) after normalization to two housekeeping genes (*B2M* and *RPL13A*) by qRT-PCR from cervical tissue smear samples of 261 female controls. Interestingly, the transcript levels of all three genes were positively correlated in our cohort (R^2 = 0.47, $P = 5.1 \times 10^{-38}$ for *MICA* and *HCP5*; R^2 = 0.19, $P = 1.4 \times 10^{-12}$ for *HLA-DQA1* and *HCP5*; R^2 = 0.16,

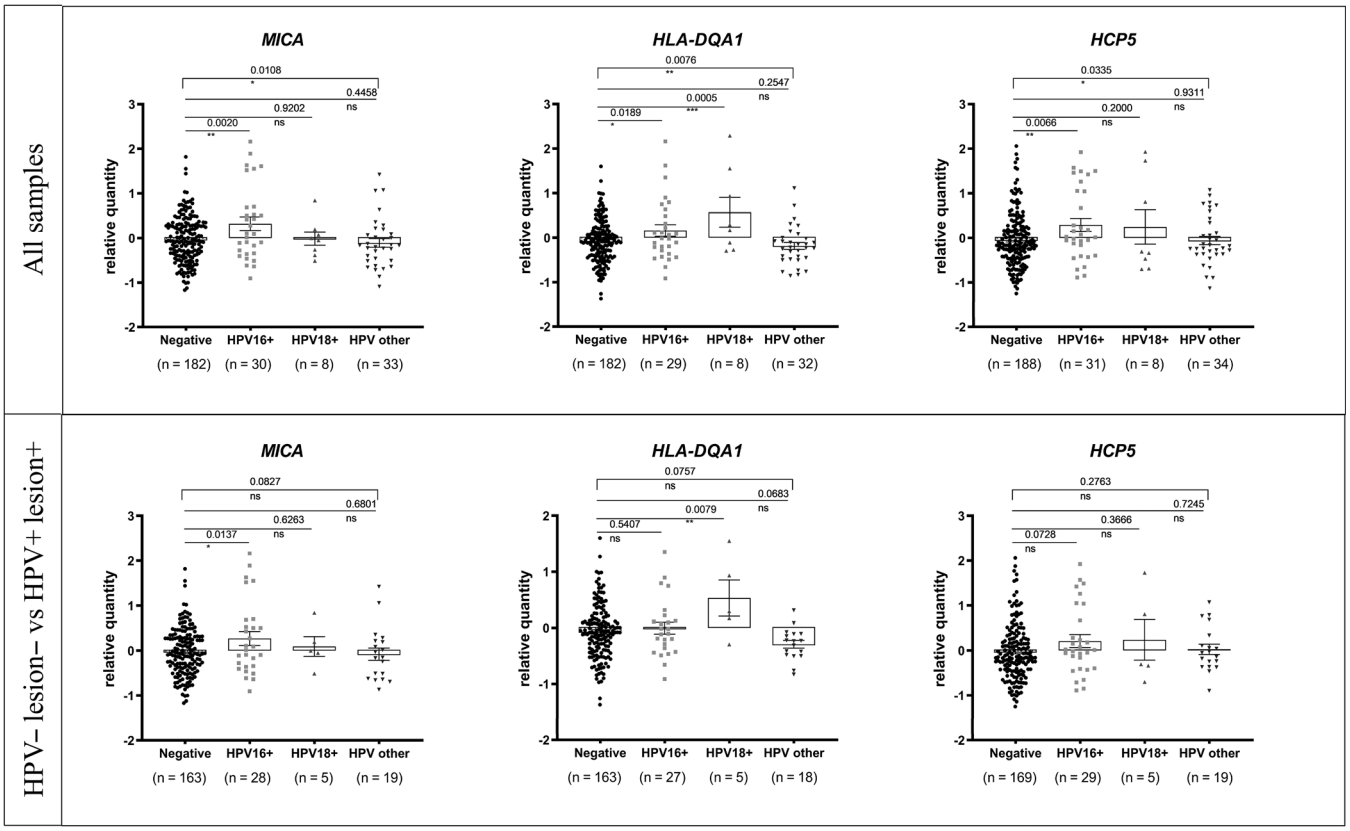

FIGURE 2 Association of transcript levels in dependence of HPV status. Comparison of candidate transcript levels for *MICA*, *HLA-DQA1* or *HCP* in HPV negative vs positive cervical samples (top panel) and in HPV negative cervical samples without lesion vs HPV positive cervical samples with documented lesion (bottom panel), each upon stratification of HPV status into subtypes 16, 18 and others. Asterisks indicate $P \leq .05$ (*) or $P \leq .01$ (**) after an ANOVA or t test comparison between groups. ns, $P > .05$

$P = 1.3 \times 10^{-10}$ for *HLA-DQA1* and *MICA*), indicating co-regulation of genes at that locus (Figure 1). This correlation was observed independently of the HPV status but correlation coefficients appeared higher in HPV positive samples (Figure 1). In a multivariate regression analysis, the mRNA levels of *MICA* and *HLA-DQA1* were independently correlated with the levels of the *HCP5* noncoding RNA.

When we stratified the cervical samples by HPV status, we found that the expression of *HCP5* tended to be higher in HPV-positive samples as compared to samples without detectable HPV infection ($P = .06$) (Figure S2). Upon stratification of HPV status into high-risk HPV subtypes, all three genes, *MICA*, *HLA-DQA1* and *HCP5* had a higher transcript level in HPV16-positive cervical samples as

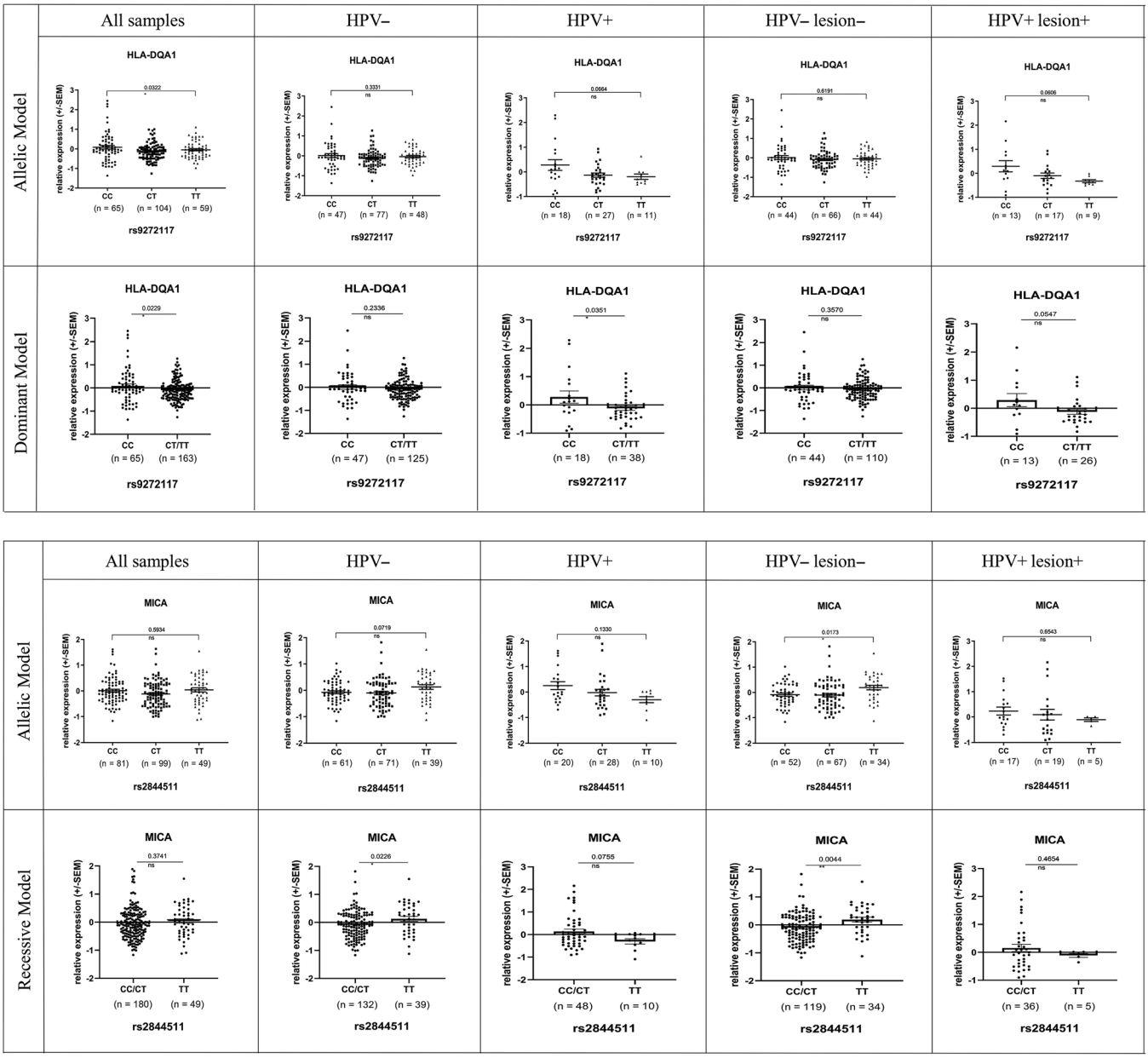

FIGURE 3 Association of transcript levels with genotypes for eQTL analysis. Transcript levels of *HLA-DQA1*, *MICA* and *HCP5* were associated with the corresponding genotypes of rs2844511 and rs9272117 under different models of inheritance: Allelic model, where each allele contributes to the transcript level; Dominant model, where the minor allele is dominant and both heterozygous and minor allele homozygous genotypes have similar effects on the transcript level; or Recessive model, where the minor allele is recessive and the heterozygous and major allele homozygous have the similar effect on transcript levels. Asterisks indicate $P \leq .05$ (*) or $P \leq .01$ (**) after an ANOVA or *t* test comparison between groups. ns, $P > .05$. First panel: Transcript level association of *HLA-DQA1* with genotype of rs9272117 under the allelic and dominant model of inheritance in all samples and samples stratified by the presence of HPV and/or lesions. Second panel: Transcript level association of *MICA* with genotype of rs2844511 under the allelic and recessive model of inheritance in all samples and samples stratified by the presence of HPV and/or lesions. Third panel (next page): Transcript level association of *HCP5* with genotype of rs2844511 under the allelic and recessive model of inheritance in all samples and samples stratified by the presence of HPV and/or lesions

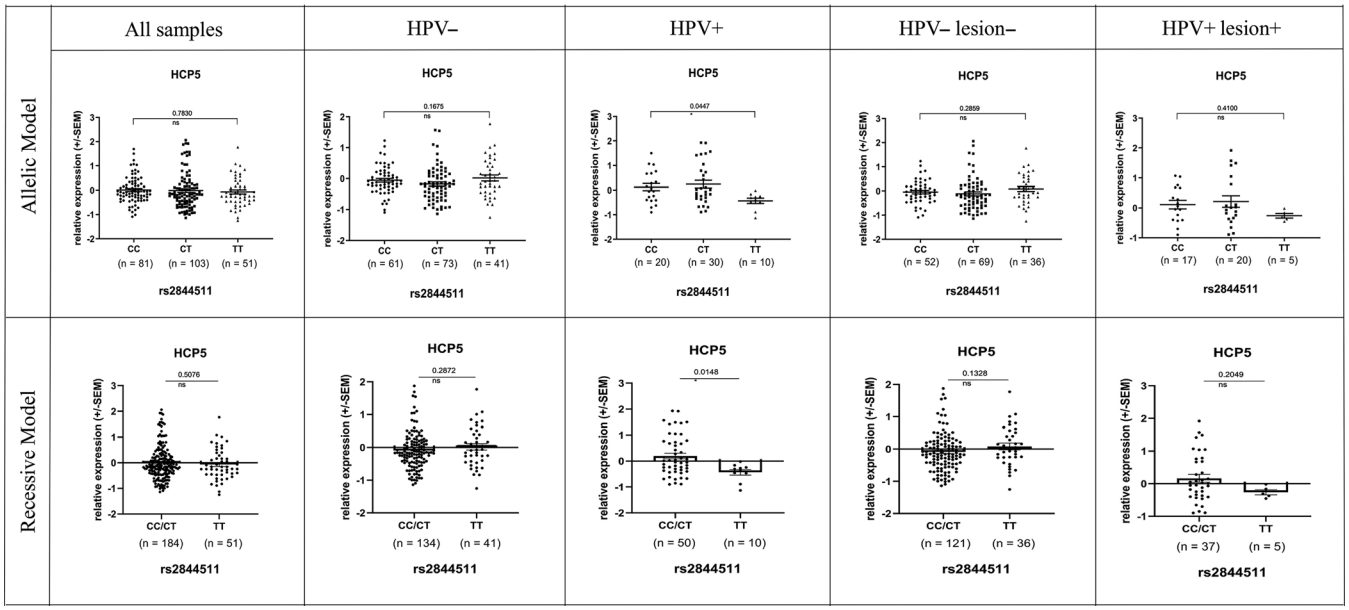

FIGURE 3 (Continued)

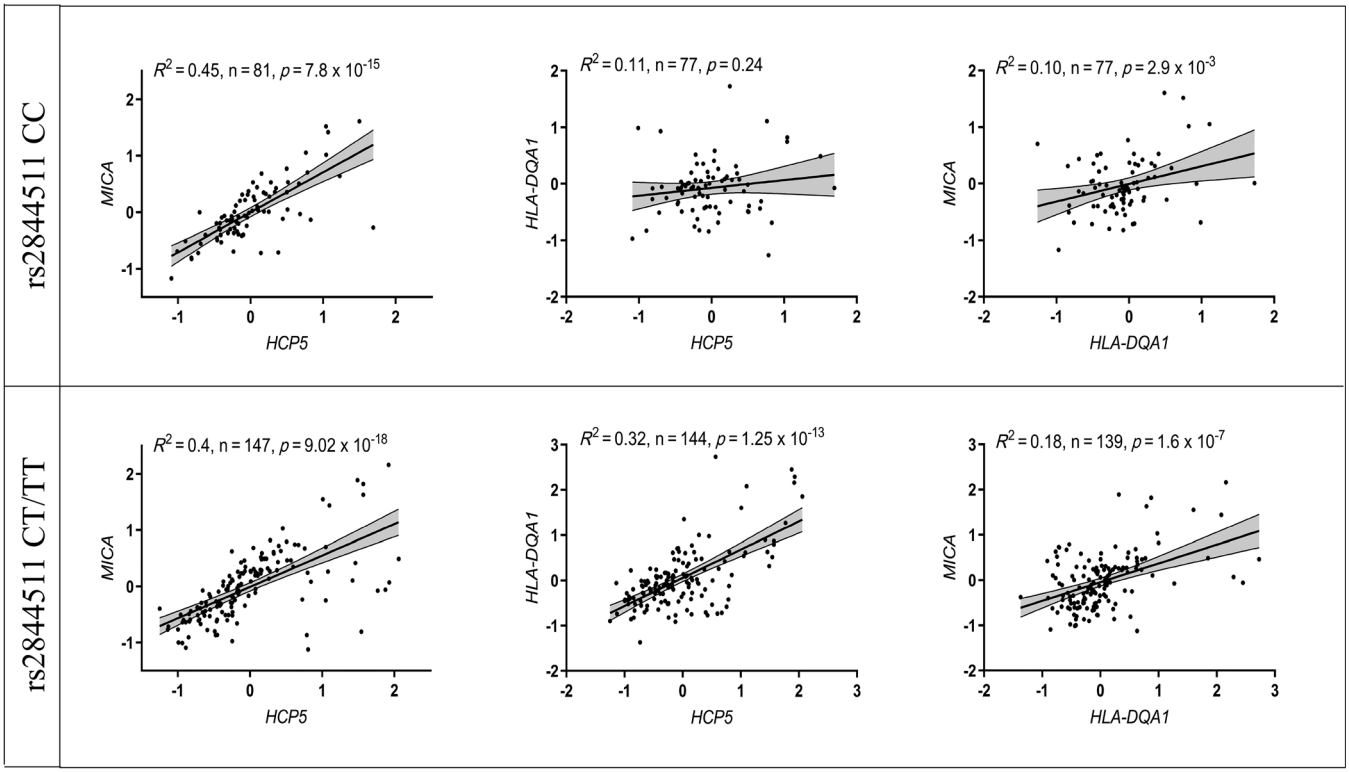

FIGURE 4 Correlation of transcript levels by rs2844511 genotypes. Pearson correlation coefficients and *P* values for regression were generated to compare the transcript abundance for each gene pair in the cDNA cohort. In contrast to Figure 1, these analyses were performed after stratification by rs2844511 genotype (dominant model). *From left to right*: *MICA* and *HCP5*; *HLA-DQA1*; and *HCP5* and *MICA* and *HLA-DQA1*. *Top panel*: rs2844511 common homozygotes; *bottom panel*: rs2844511 carriers of the rare allele (heterozygotes and rare homozygotes combined)

compared to HPV-negative samples (*P* = .002 for *MICA*, *P* = .019 for *HLA-DQA1* and *P* = .007 for *HCP5*; Figure 2). *HLA-DQA1* also tended to be higher in HPV18 positive samples (*P* = .0005) but numbers were small.

We then aimed to investigate how the expression levels of these genes was affected by the genotype for the two risk-associated variants rs2844511 and rs9272117. A cDNA sequencing of selected regions in *MICA* with an exonic variant suggested allelic imbalances in

some samples (Figure S3), but the lack of linkage disequilibrium did not allow us to correlate these with rs2844511 or rs9272117 and determine the direction of effect. We, therefore, stratified our qPCR results by rs2844511 and rs9272117 genotypes (Figure 3, Figures S4-S7). In our eQTL analysis, rs9272117 was associated with the levels of *HLA-DQA1* (P = .03 under an allelic model, P = .02 under a dominant model; Figure 3A). The association of rs9272117 with *HLA-DQA1* was restricted to HPV-positive samples (P = .04, dominant model), with the rare allele T showing decreased *HLA-DQA1* levels. In HPV-positive samples, the second variant rs2844511 tended to associate with *HCP5* under the allelic model (P = .04) and recessive model (P = .01), with the rare allele T showing decreased *HCP5* levels (Figure 3B). The presence of the rare allele also appeared to show decreased *MICA* levels in a dose-dependent manner in HPV positive samples, but this observation did not reach statistical significance. In HPV negative samples, rs2844511 was associated with *MICA* under the allelic model (P = .02) and the recessive model (P = .004; Figure 3C), with the rare allele T showing increased *MICA* levels. We finally tested whether the risk genotypes affected the strong correlation identified between *MICA*, *HLA-DQA1* and *HCP5* at the transcript level. No significant effect was observed when the correlation analyses were stratified by rs9272117 genotypes (data not shown). In an analysis stratified by rs2844511 genotype, the correlation between *MICA* and *HCP5* remained stable for any genotype of rs2844511 (Figure 4, left panel), whereas the correlation between *HLA-DQA1* and *MICA* became weak and the correlation between *HLA-DQA1* and *HCP5* was lost in common homozygotes (Figure 4, middle and right panel), indicating that the rs2844511 risk locus impacts on the co-regulation of these distant genes.

4 | DISCUSSION

The familial relative risk for cervical cancer is incompletely understood. Previous studies for cervical cancer so far have implicated variants at the *HLA* locus near or within *HLA-DQA1*, *HLA-DRB1/2*, *HLA-G*.[10,20,24-28] Polymorphisms in *MICA* have been reported to contribute towards the pattern of association of the *HLA* region with cervical cancer susceptibility.[29-32] A study in a Swedish cervical cancer cohort showed that a frameshift mutation A5.1. in *MICA* decreases the expression of *MICA* on the tumor cell surface.[29,31] Conversely, a prospective study found that low levels of soluble *MICA* are correlated with a higher chance of disease relapse in cervical adenocarcinoma patients.[33]

We aimed to validate the reported association of *HLA* with cervical cancer risk using the two variants rs2844511 and rs9272117 which are strongly correlated with previously reported GWAS hits but are located about 1.2 Mb apart and not linked to each other. We genotyped these variants in a newly established and large case-control study of the German Cervigen Consortium, recruiting a total of 2481 cases and 1556 healthy female controls from the German population. Importantly, both variants rs9272117 (*HLA-DQA1*) and rs2844511 (*MICA/ HCP5)* were associated with invasive cervical cancer, with high-grade dysplasia and, in case of rs2844511, with low-grade dysplasia. These data, obtained in a candidate approach, strongly corroborate the two previously reported GWAS signals as true genetic associations in an independent case-control series. We also found evidence of an association with HPV status for both variants, suggesting that they could serve as genetic markers for susceptibility to reproductive infection. However, the latter association did not survive correction for multiple testing, and more data will be needed to support these observations.

We selected the genes *HLA-DQA1*, *MICA* and *HCP5* for our gene expression study in cervical tissue smear samples due to their proximity to the SNPs under investigation. We identified strong correlations between their mRNA levels, specifically between *HLA-DQA1* or *MICA* with *HCP5*. The regulatory lncRNA named HLA complex P5 (*HCP5*) is known to affect immune response in various infectious diseases and cancers. *HCP5* acts as a genomic anchor point for binding transcription factors, enhancers and chromatin remodeling enzymes in the regulation of transcription and chromatin folding.[34] A part of the *HCP5* RNA sequence was found to be 99% identical to the papillomavirus minor structural protein-interacting protein (PMSP), which is a peptide interacting with viral capsid proteins involved in viral assembly.[34] Higher expression of *HCP5* is seen in cervical carcinoma, where it sequesters miRNA-15a and thereby increases cell proliferation.[35]

Variants within this gene may indirectly impact viral replicative infection, possibly via downstream effects such as miRNA inhibition.[34,36] In line with this suggestion, the levels of *HCP5* were significantly higher in HPV-positive cervical samples as compared to samples without detectable HPV infection, and all three tested genes (*MICA*, *HLA-DQA1* and *HCP5*) had higher transcript levels in HPV16-positive cervical samples as compared to HPV-negative samples. These data suggest a significant co-regulation of genes across the 6p21.32-33 genetic risk region in primary cervical tissue, and the level of upregulation correlates with HPV infection. Interestingly, the risk allele of rs2844511 was required for the tight correlation of *MICA* or *HCP5* with *HLA-DQA1* at the transcript level in uninfected cells, while the common (protective) homozygous genotype uncoupled this co-regulation. The two loci are over 1 MBp apart, and further research is needed to reveal the potential mechanisms at the molecular level.

With these correlations, the association of genotypes with levels of the single genes *MICA*, *HLA-DQA1* and *HCP5* were interdependent and influenced by HPV status. While the associations with *HCP5* and *HLA-DQA1* were observed in HPV positive samples and need to be followed in larger series, there was evidence for rs2844511 being an eQTL for *MICA* in noninfected cervical epithelium. *MICA*, the "major histocompatibility complex class I polypeptide-related sequence A gene", encodes a cancer cell surface ligand that is recognized by the NKG2D receptor, which is expressed on natural killer (NK) and CD8+ T cells. Recognition of the ligand then initiates a protective cytolytic immune response. Little surface expression of MICA is seen in healthy cells; however, most epithelial tumor cells express MICA. There are multiple modes through which cervical cancer cells evolve to evade this detection, including the secretion of MICA in a soluble form to block NKG2D receptors or the aberrant expression of NKG2D

receptors on their own surface.[30,31,37-39] It is possible that one mechanism by which rs2844511 could modify cervical cancer risk is the regulation of *MICA* expression and subsequent immune-surveillance of tumor cells. From the results of our study, the risk allele of rs2844511 associated with increased *MICA* levels in HPV negative (normal) cervical epithelium, but this was reversed in HPV infected cells. rs2844511 has recently been documented as a strongly significant pQTL for MICA in blood.[40] There, the risk allele (T) also was associated with increased MICA levels. The evidence that the risk allele of rs2844511, or a variant in close LD, increases *MICA* levels and triggers co-regulation of genes across the *HLA* locus in a manner that is codirectional with but independent of HPV, in uninfected cervical epithelium, may help to explain why we find an increased carrier risk already for low-grade dysplasia in our genetic case-control study and why the risk allele may predict low HPV.

Compared to rs2844511, our findings were different for rs9272117 where the protective allele, or a variant in close LD, appears to antagonize HPV by limiting *HLA-DQA1* levels in HPV positive tissues, may predict low HPV16 and manifests its significant effect on risk only in the groups of high-grade dysplasia and invasive cancer. HLA class II molecule DQA1 is a membrane-anchored protein expressed on the surface of antigen-presenting cells. Past studies have identified protective and risk alleles for cervical dysplasia and cancer within the *HLA-DQA1* gene.[10,13,32,41] Polymorphisms in the alpha chain (DQA1) or beta chain (DQB1) may affect antigen-recognition via an alteration of the binding affinity and thereby influence the immune response, which can lead to persistent lesions.[13,41] Levels of HLA-DQA1 protein are reported to be upregulated in esophageal squamous cell carcinoma (ESCC), another HPV-related cancer and were associated with poor outcomes.[42] Based on this previous evidence and our findings of higher *HLA-DQA1* levels in HPV16 positive cervical samples, such mechanisms will be worthy of investigation in cervical cancer as well.

GWAS results, when confirmed in independent studies, can eventually uncover novel therapeutic targets or serve as useful biomarkers to determine more precisely the risk of progression in females with a diagnosed dysplasia and/or HPV infection. Our observed effects on gene expression levels point to complex genetic architecture and regulatory network at this gene-dense region and should inform further fine-mapping and functional studies to identify the set of credible causal variants and to understand the mechanism of the risk variants in cervical dysplasia and carcinoma. The present data verify 6p21.32-33 as the first consistently replicated genetic susceptibility locus for cervical cancer, although larger case-control studies will likely lead to additional genomic risk loci in the future.

ACKNOWLEDGEMENTS

We cordially thank all the patients, volunteers and clinicians who participated in sampling and documentation. We specifically thank Ivonne Bünting for her technical assistance, Eda Celik for her support in documentation, Anja Thümmler, Tina Ruider, Florian Müller and the whole German Colposcopy Study Group for their support in patient recruitment, and Juliane Hoyer for her support in DNA extraction. We are grateful to Thomas Schulz and Michael Heuser for helpful discussion. Our study was supported by the Bruno and Helene Jöster Foundation. Dhanya Ramachandran was supported by the Hannover Biomedical Research School (HBRS) and the MD/PhD program "Molecular Medicine". Qianqian Mao was supported by a research stipend from the Shanghai New Medical Union Hospital, and Yingying Wang was supported by a research stipend from the Xuzhou Medical University.

CONFLICT OF INTEREST

P. A. F. received grants from Novartis and Biontech and discloses personal fees from Novartis, Roche, Pfizer, Celgene, Daiichi-Sankyo and Lilly during the conduct of our study. This funding had no direct connection to the study presented here and did not influence our study design and results. The other authors have no disclosures and declare there are no conflicts of interest in the writing, preparing or finalizing of this manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of our study are available from the corresponding author upon reasonable request.

ORCID

Peter A. Fasching https://orcid.org/0000-0003-4885-8471
Karl-Ulrich Petry https://orcid.org/0000-0002-5875-9840
Thilo Dörk https://orcid.org/0000-0002-9458-0282

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
2. Bruni L, Albero G, Serrano B, et al. Human papillomavirus and related diseases report—World. ICO (Institut Català d'Oncologia) Information Centre on HPV and Cancer. 2019;307. https://hpvcentre.net/statistics/reports/XWX.pdf.
3. Burd E. Human papillomavirus and cervical cancer. *Clin Microbiol Rev*. 2003;16:1-17.
4. Bonde J, Bottari F, Parvu V, et al. Bayesian analysis of baseline risk of CIN2 and ≥CIN3 by HPV genotype in a European referral cohort. *Int J Cancer*. 2019;145:1033-1041.
5. Bava SV, Thulasidasan AKT, Sreekanth CN, Anto RJ. Cervical cancer: a comprehensive approach towards extermination. *Ann Med*. 2016;48:149-161.
6. Graham SV. The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review. *Clin Sci*. 2017;131:2201-2221.
7. Ahlbom A, Lichtenstein P, Malmström H, Feychting M, Hemminki K, Pedersen NL. Cancer in twins: genetic and nongenetic familial risk factors. *J Natl Cancer Inst*. 1997;89:287-293.
8. Hemminki K, Vaittinen P. Familial risks in in situ cancers from the family-cancer database. *Cancer Epidemiol Biomarkers Prev*. 1998;7:865-868.
9. Magnusson PKE, Lichtenstein P, Gyllensten UB. Heritability of cervical tumours. *Int J Cancer*. 2000;88:698-701.
10. Leo PJ, Madeleine MM, Wang S, et al. Defining the genetic susceptibility to cervical neoplasia—a genome-wide association study. *PLoS Genet*. 2017;13:1-20.

11. Chen D, Gyllensten U. Lessons and implications from association studies and post-GWAS analyses of cervical cancer. *Trends Genet*. 2015;31:41-54.
12. Chen D, Juko-Pecirep I, Hammer J, et al. Genome-wide association study of susceptibility loci for cervical cancer. *J Natl Cancer Inst*. 2013;105:624-633.
13. Chen D, Enroth S, Liu H, et al. Pooled analysis of genome-wide association studies of cervical intraepithelial neoplasia 3 (CIN3) identifies a new susceptibility locus. *Oncotarget*. 2016;7:42216-42224.
14. Bao X, Hanson AL, Madeleine MM, et al. HLA and KIR associations of cervical neoplasia. *J Infect Dis*. 2018;218:2006-2015.
15. Safaeian M, Johnson LG, Yu K, et al. Human leukocyte antigen class I and II alleles and cervical adenocarcinoma. *Front Oncol*. 2014;4:119.
16. Spelten B, Grußendorf-Conen EI, Rübben A. Human leukocyte antigen class II alleles and natural history of HPV 2/27/57-induced common warts. *Arch Dermatol Res*. 2004;296:105-111.
17. Remmers EF, Cosan F, Kirino Y, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. *Nat Genet*. 2010;42:698-702.
18. The International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, et al. Risk alleles for multiple sclerosis identified by a Genomewide study. *N Engl J Med*. 2007;357:851-862.
19. Reveille JD, Sims AM, Danoy P, et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat Genet*. 2010;42:123-127.
20. McKay J, Tenet V, Franceschi S, et al. Immuno-related polymorphisms and cervical cancer risk: the IARC multicentric case-control study. *PLoS One*. 2017;12:1-13.
21. Hellemans J, Mortier G, de Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol*. 2008;8:R19.
22. Mestdagh P, Van Vlierberghe P, De Weer A, et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol*. 2009;10:R64.
23. Vandesompele J, de Preter K, Pattyn F, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3:research0034.1.
24. Li X, Huang K, Zhang Q, et al. Genome-wide association study identifies four SNPs associated with response to platinum-based neoadjuvant chemotherapy for cervical cancer. *Sci Rep*. 2017;7:41103.
25. Wang S, Sun H, Jia Y, et al. Association of 42 SNPs with genetic risk for cervical cancer: an extensive meta-analysis. *BMC Med Genet*. 2015;16:1-10.
26. Masuda T, Low SK, Akiyama M, et al. GWAS of five gynecologic diseases and cross-trait analysis in Japanese. *Eur J Hum Genet*. 2019;28:95-107.
27. Yang YC, Chang TY, Chen TC, Lin WS, Lin CL, Lee YJ. Replication of results from a cervical cancer genome-wide association study in Taiwanese women. *Sci Rep*. 2018;8:8-12.
28. Burk RD, Chen Z, Saller C, et al. Integrated genomic and molecular characterization of cervical cancer. *Nature*. 2017;543:378-384.
29. Chen D, Hammer J, Lindquist D, Idahl A, Gyllensten U. A variant upstream of HLA-DRB1 and multiple variants in MICA influence susceptibility to cervical cancer in a Swedish population. *Cancer Med*. 2014;3:190-198.
30. Chen D, Gyllensten U. MICA polymorphism: biology and importance in cancer. *Carcinogenesis*. 2014;35:2633-2642.
31. Chen D, Gyllensten U. A cis-eQTL of HLA-DRB1 and a frameshift mutation of MICA contribute to the pattern of association of HLA alleles with cervical cancer. *Cancer Med*. 2014;3:445-452.
32. Chen D, Gyllensten U. Systematic investigation of contribution of genetic variation in the HLA-DP region to cervical cancer susceptibility. *Carcinogenesis*. 2014;35:1765-1769.
33. Samuels S, Ferns DM, Meijer D, et al. High levels of soluble MICA are significantly related to increased disease-free and disease-specific survival in patients with cervical adenocarcinoma. *Tissue Antigens*. 2015;85:476-483.
34. Kulski JK. Long noncoding RNA HCP5, a hybrid HLA class I endogenous retroviral gene: structure, expression, and disease associations. *Cell*. 2019;8:480.
35. Yu Y, Shen HM, Fang DM, Meng QJ, Xin YH. LncRNA HCP5 promotes the development of cervical cancer by regulating MACC1 via suppression of microRNA-15a. *Eur Rev Med Pharmacol Sci*. 2018;22:4812-4819.
36. Farquhar C, Mbori-Ngacha D, Wamalwa D, Harris J, John-Stewart G. Illness during pregnancy and bacterial vaginosis are associated with in-utero HIV-1 transmission. *AIDS*. 2010;24:153-157.
37. Du C, Bevers J, Cook R, et al. MICA immune complex formed with alpha 3 domain-specific antibody activates human NK cells in a Fc-dependent manner. *J Immunother Cancer*. 2019;7:1-13.
38. Gutiérrez-Hoya A, Zerecero-Carreón O, Valle-Mendiola A, et al. Cervical cancer cells express markers associated with Immunosurveillance. *J Immunol Res*. 2019;2019:1242979.
39. Weiss-Steider B, Soto-Cruz I, Martinez-Campos CA, Mendoza-Rincon JF. Expression of MICA, MICB and NKG2D in human leukemic myelomonocytic and cervical cancer cells. *J Exp Clin Cancer Res*. 2011;30:37.
40. Suhre K, Arnold M, Bhagwat AM, et al. Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat Commun*. 2017;8:14357.
41. Helland Å, Olsen AO, Gjøen K, et al. An increased risk of cervical intra-epithelial neoplasia grade II-III among human papillomavirus positive patients with the HLA-DQA1*0102- DQB1*0602 haplotype: a population-based case-control study of Norwegian women. *Int J Cancer*. 1998;76:19-24.
42. Shen FF, Pan Y, Li JZ, et al. High expression of HLA-DQA1 predicts poor outcome in patients with esophageal squamous cell carcinoma in northern China. *Medicine (Baltimore)*. 2019;98:e14454.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Ramachandran D, Schürmann P, Mao Q, et al. Association of genomic variants at the human leukocyte antigen locus with cervical cancer risk, HPV status and gene expression levels . *Int. J. Cancer*. 2020;147:2458–2468. https://doi.org/10.1002/ijc.33171