Identifying the HLA DRB1-DQB1 molecules and predicting epitopes associated with high-risk HPV infection clearance and redetection

Luisa Del Río-Ospina1,2,7, Milena Camargo1,3,4,7, Sara C. Soto-De León1,7, Ricardo Sánchez5, Darwin A. Moreno-Pérez2,6, Manuel E. Patarroyo1,5 & Manuel A. Patarroyo1,6✉

Several determining factors are involved in HPV infection outcomes; human leukocyte antigen (HLA) polymorphisms have been described as related factors. This study has ascertained the effect of genetic variation on HLA-DRB1 and DQB1 genes on HPV-16/-18/-31/-33/-45 and -58 clearance and redetection in Colombian women. PCR and qPCR were used for viral identification and the Illumina MiSeq system was used for HLA-typing of cervical samples (n = 276). Survival models were adjusted for identifying alleles/haplotypes related to HPV clearance/redetection; L1/L2 protein-epitope binding to MHC-II molecules was also predicted. Significant associations suggested effects favouring or hampering clearance/redetection events depending on the viral type involved in infection, e.g. just DRB1*12:01:01G favoured HPV-16 (coeff: 4.8) and HPV-45 clearance (coeff: 12.65) whilst HPV-18 (coeff: 2E-15), HPV-31 (coeff: 8E-17) and HPV-58 hindered elimination (coeff: 1E-14). An effect was only observed for some alleles when configured as haplotypes, e.g. DRB1*04:07:01G (having the greatest frequency in the target population) was associated with DQB1*02:01:1G or *03:02:03. Epitope prediction identified 23 clearance-related peptides and 29 were redetection-related; eight might have been related to HPV-16/-18 and -58 persistence and one to HPV-18 elimination. HLA allele/haplotype relationship with the course of HPV infection (clearance/redetection) depended on the infecting HPV type, in line with the specific viral epitopes displayed.

Human papilloma virus (HPV) is the most common sexually-transmitted viral infection, having around 291 million infections annually1–3. The causal relationship between persistent high-risk types of human papillomavirus (HR-HPV) infections and the development of cervical lesions progressing to cervical cancer (CC) has been extensively demonstrated4,5. Many infections (around 90%) are eliminated during an average period of 2 years; however, some of them are latent as non-productive infection is limited to the epithelium’s base layer and they do not become detected6,7. A complex interaction between viral and host factors is responsible for CC’s clinical course and development8,9. Host immunological and genetic factors play an essential role in HPV infection outcome10,11. Human leukocyte antigen (HLA) system alleles and haplotypes have been reported as being responsible for antigen presentation, recognition of infected cells and HPV elimination12,13.

Some HLA alleles and haplotypes have been described as being associated with CC, such as DRB1*07:02, DRB1*13:01, DQA1*01:03 and DQB1*06:0314–16. However, most reports contain inconsistencies given the differences in the populations being analysed, the typing techniques, the outcomes considered and the methodological design11,15,17–19. Few longitudinal studies have evaluated these molecules’ association with HPV’s natural history.

1Molecular Biology and Immunology Department, Fundación Instituto de Inmunología de Colombia, Carrera 50 # 26-20, Bogotá, Colombia. 2Clinical Research Group, Instituto Nacional de Cancerología, Calle 1 # 9-85, Bogotá, Colombia. 3PhD Programme in Biomedical and Biological Sciences, Universidad del Rosario, Carrera 24 # 63C-69, Bogotá, Colombia. 4Animal Science Faculty, Universidad de Ciencias Aplicadas y Ambientales (U.D.C.A), Calle 222 # 55-37, Bogotá, Colombia. 5Faculty of Medicine, Universidad Nacional de Colombia, Carrera 45 # 26-85, Bogotá, Colombia. 6School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 # 63C-69, Bogotá, Colombia. 7These authors contributed equally: Luisa Del Río-Ospina, Milena Camargo and Sara C. Soto-De León. ✉e-mail: mapatarr.fidic@gmail.com
of infection (considering them as being generic infections) and just HPV-16 and -18 have been considered at type-specific level19–22. This study was aimed at identifying HLA-DRB1 and DQB1 alleles related to the clearance and redetection of the 6 HPV types having the greatest distribution in Colombia (HR-HPV-16, -18, -31, -33, -45 and -58) in a cohort of Colombian women using next generation sequencing (NGS) for HLA typing and quantitative PCR assay for viral detection. L1 and L2 protein peptides fitting into alleles were analysed for predicting which of them might have been related to infection events regarding each viral type. Such information is relevant to understanding specific infections' natural history and the genetic factors modulating them. The results should prove useful in identifying immunological biomarkers enabling establishing HPV infection susceptibility and its clinical course.

Results
Selecting the study population. Two hundred and seventy-six women were considered eligible for this study; 12 were excluded as complete information was missing (lack of information regarding HPV or HLA typing). Cohort follow-up lasted 32.3 months; 206 women were visited four times and 59 women a fifth time during the follow-up period. Table 1 describes the target population and its sociodemographic characteristics. Regarding type-specific detection, HPV-16 had the greatest prevalence, followed by HPV-18; however, variations regarding specific type distribution were found during follow-up (shown in Supplementary Fig. S1).

Table 1. Baseline sociodemographic characteristics, risk factors and cervical findings regarding the study population. IQR: interquartile range, 95%CI: 95% confidence interval, STD: sexually transmitted disease, ASCUS: atypical squamous cells of undetermined significance, LSIL: low-grade squamous intraepithelial lesion.

| Characteristic                          | n    | %    | 95%CI       |
|----------------------------------------|------|------|-------------|
| Age, years (median, IQR)              | 43 (34-50) |
| Origin (n = 266)                      |      |      |             |
| Chaparral                              | 7    | 2.63 | 0.70 4.57   |
| Bogotá                                 | 82   | 30.83| 25.24 36.41 |
| Girardot                               | 187  | 66.54| 60.83 72.25 |
| Ethnicity (n = 266)                    |      |      |             |
| Afro-descendant                        | 5    | 1.88 | 0.24 3.52   |
| Mestizo                                | 259  | 97.37| 95.43 99.30 |
| Indigenous                             | 2    | 0.75 | -0.29 1.80  |
| Smoker (n = 266)                       |      |      |             |
| No                                     | 200  | 75.19| 69.61 80.04 |
| Yes                                    | 66   | 24.81| 19.96 30.39 |
| Age on first intercourse, years (n = 266) |      |      |             |
| ≤ 18 years                             | 154  | 57.89| 51.84 63.72 |
| > 18 years                             | 112  | 42.11| 36.28 48.16 |
| Lifetime amount of sexual partners (n = 260) |      |      |             |
| 1                                      | 116  | 44.62| 38.47 50.88 |
| 2                                      | 88   | 33.85| 28.12 39.95 |
| ≥ 3                                    | 56   | 21.54| 16.70 27.04 |
| Contraceptive method (n = 248)         |      |      |             |
| None                                   | 96   | 38.71| 32.61 44.81 |
| Barrier                                | 42   | 16.94| 12.24 21.64 |
| Surgery                                | 81   | 32.66| 26.78 38.54 |
| Hormonal                               | 29   | 11.69| 7.67 15.72  |
| Births (n = 257)                       |      |      |             |
| None                                   | 10   | 3.89 | 1.51 6.27   |
| 1-2                                    | 127  | 49.42| 43.26 55.57 |
| 3-4                                    | 104  | 40.47| 34.43 46.51 |
| ≥ 4                                    | 16   | 6.23 | 3.25 9.20   |
| Abortions (n = 200)                    |      |      |             |
| None                                   | 104  | 52.00| 45.02 58.98 |
| 1                                      | 72   | 36.00| 29.29 42.71 |
| ≥ 2                                    | 24   | 12.00| 7.46 16.54  |
| STD (n = 256)                          |      |      |             |
| No                                     | 205  | 80.08| 75.15 85.00 |
| Yes                                    | 51   | 19.92| 15.00 24.85 |
| Cytological findings (n = 264)         |      |      |             |
| Negative                               | 241  | 91.29| 87.86 94.71 |
| ASCUS                                  | 10   | 3.79 | 1.47 6.11   |
| LSIL                                   | 13   | 4.92 | 2.30 7.55   |

of infection (considering them as being generic infections) and just HPV-16 and -18 have been considered at type-specific level19–22. This study was aimed at identifying HLA-DRB1 and DQB1 alleles related to the clearance and redetection of the 6 HPV types having the greatest distribution in Colombia (HR-HPV-16, -18, -31, -33, -45 and -58) in a cohort of Colombian women using next generation sequencing (NGS) for HLA typing and quantitative PCR assay for viral detection. L1 and L2 protein peptides fitting into alleles were analysed for predicting which of them might have been related to infection events regarding each viral type. Such information is relevant to understanding specific infections' natural history and the genetic factors modulating them. The results should prove useful in identifying immunological biomarkers enabling establishing HPV infection susceptibility and its clinical course.
redetection of positive cases at the end of follow-up for HPV-16 (6.95%) and HPV-31 (9.64%) compared to a higher non-redetection percentage for HPV-58 (66.8%) and HPV-33 (51.82%) at the end of follow-up (Fig. 1). The highest clearance rate was observed for HPV-33 (6.99) whilst HPV-18 (6.27) and HPV-31 (6.31) had the highest redetection rates (Table 2).
Identifying HLA alleles and haplotypes related to HR-HPV infection’s clinical course.

Calculating multicollinearity between the variables (evaluated by variance inflation factor (VIF) and tolerance) revealed no collinearity between the variables included in the model. Cox multivariate and log-normal models were thus adjusted for identifying alleles and haplotypes related to infection clearance/persistence (Supplementary Tables S5 to S8) and redetection (Supplementary Tables S9 to S11) for each HPV type evaluated.

Alleles/haplotypes having associations favouring clearance/redetection were found (considered in the model as having greater and earlier probability of occurrence) as well as hindering associations (considered as lower probability and later occurrence).

Sixty-three DRB1 and nine DQB1 allele associations with HR-HPV infection clearance were found (Supplementary Tables S5 and S6) of which 47 for DRB1 (represented by 20 alleles) and 6 for DQB1 (represented by 3 alleles) continued being significant after p values (p ≤ 0.001) had been corrected (Fig. 2 and Supplementary Table S12). Fifteen associations were identified favouring clearance (greater probability or earlier occurrence) and 38 hindering it (lower probability or later occurrence) (Fig. 2).

The HPV-33 model could not be adjusted for infection redetection due to the few (n = 9) events found for this viral type in the study population (Supplementary Table S9); 88 associations were found for the remaining viral types (Supplementary Tables S9 and S13). Forty-five associations continued being significant for DRB1 (23 alleles) and five for DQB1 (3 alleles) after the p values had been corrected (p ≤ 0.001), 17 of these associations favouring infection redetection and 33 hindering it (Fig. 2 and Supplementary Table S13).

Two hundred and ninety-five associations were identified for 144 DRB1-DQB1 haplotypes (Supplementary Tables S7, S8 and S14); 233 associations (represented in 88 haplotypes) continued being significant (Fig. 3 and Supplementary Table S14) following Bonferroni correction (p ≤ 0.001). Twenty-four of these were positively associated with the probability of infection clearance whilst 63 had a negative effect, i.e. favouring HPV infection persistence (Fig. 3). Regarding redetection, 264 associations were found, 197 of them (represented by 94 haplotypes) continued being significant following p value correction (p ≤ 0.001) (Fig. 3 and Supplementary Table S15), 64 favouring redetection and 133 hindering it.

Some alleles and haplotypes were found to be related to a single HR-HPV type and others were associated with more than just one HR-HPV type. Concerning the latter, allele/haplotype associations were found which were
consistent amongst the different HPV types (i.e. in the same sense, favouring or hindering a particular event), e.g. DRB1*04:11:01 favouring HPV-31, -45 and -58 redetection. Different associations were found depending on HR-HPV type (favouring and hindering a particular event), e.g. DRB1*03:02:01 was associated with favouring HPV-31 and -33 clearance, whilst HPV-16, -45 and -58 hindered clearance (Fig. 2).

The association found for some haplotypes agreed with that found for the alleles constituting them (e.g. DRB1*15:02:01G-DQB1*06:01:01G) (Figs. 2 and 3). Haplotypes having significant associations were found where DRB1 and DQB1 constituting them had no independent associations (e.g. DRB1*04:05:01G-DQB1*02:01:01G, for HPV-33, -45 and -58 for clearance) as was modification of the effect when alleles were associated and not associated independently (e.g. DRB1*11:04:01G-DQB1*02:01:01G for redetection) (Figs. 2 and 3).

Allele/haplotype associations were found having the same sense as the event (clearance and/or redetection) for the same HPV type, i.e. an allele/haplotype was associated with hindering clearance of an HPV type and also associated with a greater probability of redetection of the same viral type. For example, DRB1*11:02:01 was associated with a lower probability of HPV-31 clearance and earlier redetection of the same viral type. DRB1*03:02:01-DQB1*04:02:01 was associated with earlier HPV-31 clearance and its later redetection (Figs. 2 and 3).

Clearance and redetection event-related peptides. It was predicted that 23 L1 and L2 capsid protein peptides from different viral types had strong binding to different HLA-II molecules (Table 3 and Supplementary Tables S16 to S19). Some peptides could have been related to the same infection outcome amongst viral types (peptide binding to DRB1*01:03, DRB1*07:11 and DRB1*15:03) and DRB1*13:04 (HPV-18, -31 and -58). Others could have explained the differences regarding an event due to being specific for each virus (peptides for DRB1*04:08, DRB1*07:11, DRB1*13:03 and DRB1*15:03 all from HPV-18) (Table 3 and Supplementary Tables S16 to S19).

Interestingly, some peptides could have favoured or hindered a clearance event depending on the allele presenting it, e.g. HPV-16 L1-QFPLGRKFLLQAGLKAKPKF presented by DRB1*04:08 and DRB1*15:03, HPV-18 L1-GHYIILFLRNVNVFPIFLQM presented by DRB1*04:08, DRB1*13:03 and DRB1*15:03 and HPV-18 L2-TTSFAFFKYSPITSSASSYS presented by DRB1*04:08, DRB1*07:11 and DRB1*15:03.

Twenty-nine L1 and L2 capsid protein peptides were found to be related to redetection. Some peptides could have been related to the same event amongst viral types (e.g. those presented by DRB1*01:03 (HPV-18 and HPV-45), DRB1*13:04, DRB1*15:02 (HPV-31 and HPV-58), whilst peptides presented by the other alleles could explain the differences found regarding events (Table 4 and Supplementary Tables S16 to S19).
Clearance/Persistence

| Event          | Allele | HR-HPV    | Outcome          | Peptides per protein | %Rank |
|----------------|--------|-----------|-------------------|----------------------|-------|
| **DRB1*01:03** | 31     | Earlier occurrence | L1-VTRTNYYHAGSARLTIVQH | 0.08 |
|                |        |           |                   | L1-FPLGRKFLLOQAGYRRAPFKK | 0.60 |
|                |        |           |                   | L1-FLSADNYTHSNSNAILE | 1.30 |
|                |        |           |                   | L2-DPDFLSQLHLHPALTSRBN | 3.50 |
| **DRB1*04:08** | 45     | Greater probability | L1-YVSTRNFLHVAGSRLTVHG | 0.15 |
|                |        |           |                   | L1-QPFLGRKFLVQAGLRPRPTI | 3.00 |
|                |        |           |                   | L1-TAEVMSYINTSNSILEWNN | 3.00 |
|                |        |           |                   | L2-SDFMDIRLHRPALTSRSGT | 1.10 |
| **DRB1*07:11** | 16     | Lower probability | L1-TGFMAGDFTTVLQANKEVPL | 4.00 |
|                |        |           |                   | L1-QPFLGRKFLLOQAGLAKPKF | 4.00 |
|                |        |           |                   | L2-EIPMTDFSTVSTPNVTSST | 1.20 |
|                |        |           |                   | L2-MLKRRRLTVPFSVDLSAA | 1.60 |
|                |        |           |                   | L2-AETGGHFTLSSLSTTSHNYE | 4.50 |
| **DRB1*13:03** | 18     | Greater probability | L1-GHYIILLFRNVNVPFIQLM | 3.00 |
|                |        |           |                   | L2-TSFSAFFKYSPTISSASSYS | 0.17 |
| **DRB1*13:04** | 16     | Lower probability | L1-QUQVYVDAPVFTPTKLITY | 3.00 |
|                |        |           |                   | L2-FFSFAMKSYTISASSY | 4.50 |
|                | 45     | Greater probability | L2-QUYRVSSTRFITPSSLTVF | 4.00 |
|                | 33     | Greater probability | L2-FFSFAMKSYTISASSY | 4.50 |
|                | 45     | Greater probability | L2-FFSFAMKSYTISASSY | 4.50 |
|                | 58     | Earlier occurrence | — | — |
| **DRB1*15:03** | 16     | Greater probability | L1-GHYIILLFRNVNVPFIQLM | 3.00 |
|                |        |           |                   | L1-QPFLGRKFLQAGLAKPKF | 3.00 |
|                |        |           |                   | L1-FYLRREQMIVRHILNRAGAV | 4.00 |
|                |        |           |                   | L2-MLHPSYTLRRRKKRFYPF | 0.40 |
|                | 31     | Lower probability | L2-VLYPPLPYIPKKKRRVIFYF | 1.70 |
|                |        |           |                   | L2-YLHPSTMKLRRKRRVSYFYF | 0.40 |
|                | 58     | Later occurrence | L2-MLHPSYTLRRRKKRFYPF | 0.40 |
| **DRB1*13:04** | 18     | Lower probability | L1-QFPLGRKFLQAGLAKPKF | 3.00 |
|                |        |           |                   | L1-FYLRREQMIVRHILNRAGAV | 4.00 |
|                | 31     | Lower probability | L2-VLYPPLPYIPKKKRRVIFYF | 1.70 |
|                |        |           |                   | L2-YLHPSTMKLRRKRRVSYFYF | 0.40 |
|                | 58     | Later occurrence | L2-MLHPSYTLRRRKKRFYPF | 0.40 |
| **DRB1*15:03** | 16     | Greater probability | L1-GHYIILLFRNVNVPFIQLM | 3.00 |
|                |        |           |                   | L1-FYLRREQMIVRHILNRAGAV | 4.00 |
|                | 18     | Lower probability | L1-GHYIILLFRNVNVPFIQLM | 3.00 |
|                |        |           |                   | L1-FYLRREQMIVRHILNRAGAV | 4.00 |
|                |        |           |                   | L2-MLHPSYTLRRRKKRFYPF | 0.40 |

Table 3. HR-HPV L1 and L2 proteins’ MHC-II binding epitopes related to clearance/persistence events. *The peptide core for binding to an allele is shown in bold.

Some viral peptides could have favoured or hindered a redetection event depending on the allele presenting them, e.g. L1-GHYIILLFRNVNVPFIQLM favouring a greater probability of HPV-18 redetection when presented by DRB1*04:02, DRB1*13:03 and DRB1*15:02, like HPV-16 L1-YVARTNYYHAGTSRLAVG and L1-TLEDTRYFVTSQAIACQKHT when presented by DQB1*03:02 and DQB1*05:02.

This study’s findings regarding infection clearance correlated with those of a prior cohort study carried out previously27,28. However, redetection events could not be differentiated between new infection and/or latent infection due to study design29.

Particularly interesting was the observation that responses to clearance and redetection were related when some MHC class II peptides were presented. For example, HPV-16 L1-TGFMAGDFTTVLQANKEVPL presented by DRB1*04:08 and HPV-18 L1-GHYIILLFRNVNVPFIQLM presented by DRB1*13:03 were related to a lower probability of clearance and greater probability of redetection. The same occurred when HPV-16 L2-AETGGHFTLSSLSTTSHNYE and HPV-18 L2-YLWPLYFIPPKRRKRPVF and HPV-31 L2-MLKRRRLTVPFSVDLSAA were presented by DRB1*04:02 and L2-SDFMDHRLHRPALTSRSGT could have been related to a later HPV-58 viral type occurrence. Contrarily, HPV-18 L2-SDFMDHRLHRPALTSRSGT could have been related to a later HPV-58 viral type occurrence. Contrarily, HPV-18 L2-SDFMDHRLHRPALTSRSGT could have been related to a later HPV-58 viral type occurrence.

Table 2. HR-HPV L1 and L2 proteins’ MHC-II binding epitopes related to clearance/persistence events. *The peptide core for binding to an allele is shown in bold.
Significant associations between HLA class II alleles/haplotypes and outcomes regarding infection (clearance/persistence, redetection) were found to be positive (greater probability or earlier occurrence of an event) or negative (lower probability or later occurrence of an event) (Figs. 2 and 3). The alleles/haplotypes favourably associated with viral clearance and hindering redetection could have been related to a lower risk of CC (given the lower risk of infection and viral persistence), whilst associations hindering clearance and favouring redetection

| Event | Allele | HR-HPV | Outcome | Peptides per protein | %Rank |
|-------|--------|--------|---------|------------------------|--------|
| DRBI*01:03 | 16 | Greater probability | L1-YLREQM4FLRHLFNRAAGV | 4.50 |
| | 18 | Lower probability | L2-SDFMD1IRLHPRALT5RGT | 1.40 |
| | 45 | Later occurrence | L2-SDFMD1IRLHPRAL5RGT | 1.10 |
| DRBI*03:02 | 18 | Lower probability | — | — |
| | 31 | Later occurrence | — | — |
| | 45 | Earlier occurrence | L2-QIGGRVHYD6ISPIAAT6EE | 4.50 |
| | 58 | Later occurrence | — | — |
| DRBI*04:02 | 18 | Greater probability | L1-GHYIILFLRNVYNVFPLQM, L1-IFYYRHGSFILT5VGNPYPFVR, L2-SFAFFKYST1ISSASYN7V, L2-SDFMD2IRLHPRALT5RGT | 4.00, 4.50, 1.70, 4.00 |
| | 58 | Later occurrence | L1-FPLGRKFLQ5GLKAKPKRL | 4.00 |
| DRBI*04:08 | 16 | Greater probability | L1-TGFGAMDTFLTLQAN7KSEVPL, L2-AETGGHFTL6SSSTISHNYE | 4.00, 4.50 |
| | 31 | Later occurrence | — | — |
| | 45 | Later occurrence | — | — |
| | 58 | Later occurrence | — | — |
| DRBI*10:01 | 18 | Lower probability | L2-TTSFAF4FKYST1ISSASSY | 0.25 |
| | 45 | Earlier occurrence | L1-KV5AYQYR5FRVALDP7NKF, L1-RLLTVGNPYPFRP5GAGNK, L1-LTAEV55YHSMSSILEWN | 1.40, 3.00, 4.50 |
| DRBI*12:01 | 18 | Greater probability | — | — |
| | 31 | Earlier occurrence | — | — |
| | 45 | Later occurrence | — | — |
| | 58 | Later occurrence | L1-QFPLGRKFLQ5GLKAKPKRL | 0.70 |
| DRBI*13:03 | 18 | Greater probability | L1-GHYIILFLRNVYNVFPLQM | 4.50 |
| | 45 | Later occurrence | — | — |
| | 58 | Later occurrence | — | — |
| DRBI*13:04 | 18 | Earlier occurrence | L1-YWPLLYYF4PKK4RVRPYFF | 1.70 |
| | 31 | Earlier occurrence | L1-YLP5Y5YMLE4RR4RVSSF | 0.40 |
| | 45 | Later occurrence | — | — |
| | 58 | Earlier occurrence | L2-FMLHP5YMLRRRKR4PPFY | 0.40 |
| DRBI*15:02 | 18 | Greater probability | L1-CGHYIILFLRNVYNVFPLQM, L2-STTSFAF4FKYST1ISSASSY, L2-5GAVRHY5FYHDSPIAP5SPEY | 1.10, 4.00, 4.00 |
| | 31 | Later occurrence | L2-SIF5VDG5DFLYHP5YMLRR | 3.50 |
| | 58 | Later occurrence | L2-TIV5V5GAD5FMLHP5YMLRR | 2.00 |
| DQB1*03:02 | 16 | Greater probability | L1-YV5ARTNIYHYHAGT5RSLAV | 1.20 |
| | 45 | Earlier occurrence | L1-5YSRT5FY5HAG55RSLVTYG, L1-TTSLYD5YRF5VQ5SAV5TQCQK | 0.50, 1.60 |
| | 58 | Earlier occurrence | L1-Y5SRT5Y5YAG5ASSR5LLAV | 0.70 |
| DQB1*05:02 | 16 | Earlier occurrence | L1-YV5ARTNIYHYHAGT5RSLAV | 1.20 |
| | 18 | Greater probability | L1-R5SF5YHAG55FRSLTV5GNYP, L1-TTSLYD5YRF5VQ5SAV5TQCQK | 0.90, 2.00 |
| DQB1*06:01 | 18 | Greater probability | — | — |
| | 31 | Later occurrence | L1-Q5FLGRKF5L5Q5G5YRARPKF | 0.70 |
| | 58 | Later occurrence | L1-Q5FLGRKF5L5Q5G5KAKPKRL | 0.70 |
| DQB1*04:02 | 18 | Greater probability | L2-TTSFAF4FKYST1ISSASSY | 1.80 |
| | 45 | Later occurrence | L2-SDFMD2IRLHPRAL5RGT | 4.50 |

Table 4. HR-HPV L1 and L2 proteins’ MHC-II binding epitopes related to redetection events. The peptide core for binding to an allele is shown in bold.
could have been related to a greater risk of CC (greater risk of infection and persistence). Previous cohort results (regarding just infection by HPV-16 and -18) would seem to support such inferences and the results presented here (Supplementary Tables S12 to S15)19-22. An immune response against HPV plays an essential role in determining such infections’ clinical course and the natural history of CC. Women having alleles/haplotypes negatively associated with clearance in this study would probably have had a lower viral peptide presentation for activating an immune response, thus favouring persistence and thereby increasing the risk of developing CC20,23. Immunoinformatics led to identifying viral peptides which could be considered factors favoured viral persistence in a host since they were related to a lower probability/later occurrence of a clearance event and the greater probability/earlier occurrence of redetection (Tables 3 and 4).

Alleles/haplotypes favouring redetection (greater probability or earlier occurrence), could have favoured HPV replication due to lower antigen presentation capability favouring the replication of such latent infections in a host, so that only when the amount of copies exceeded the detection threshold could they have been identified and diagnosed20. Epitope prediction thus contributes to identifying key host factors involved in the response to infection and could therefore be considered for designing therapeutic tools for HPV infection control (Tables 3 and 4).

Specific associations with HPV type were found for most HLA-DRB1 and DQB1 alleles and haplotypes identified in the study population, highlighting differences in the relationship with the clinical course of such infection in a host (Figs. 2 and 3). Some of these associations have been reported for other populations in which generic HPV has been considered or just HPV-16 or -18 (Supplementary Tables S12 and S14). Some alleles (DRB1*13:01 and DQB1*06:03) and haplotypes (DRB1*03:01:01G-DQB1*02:01:01G and DRB1*13:01:01G-DQB1*06:03:01G) previously reported as associated were not significant for this study’s population11,14-36. Other associations (e.g. DRB1*01:03, DRB1*04:02, DRB1*04:10, DRB1*04:11, DRB1*08:04, DRB1*11:14, DRB1*13:03 and DRB1*13:05 alleles and DRB1*03:02-DQB1*04:02:01G and DRB1*04:02:01-DQB1*03:02:01G haplotypes) were not consistent with that reported previously or constituted new findings. This was especially true regarding redetection, this being the first study to consider this outcome (Figs. 2 and 3; Supplementary Tables S12 to S15)19. The foregoing could be explained by the greater variability obtained for HLA molecules when using the NGS technique, including viral types different to HPV-16 and -18, each population’s particular characteristics, interaction with other host and viral factors11 and differences in studies’ methodological designs and the outcomes considered (i.e. pre-neoplastic lesions and CC)11,15,17-19.

A joint effect was observed in this population which could be considered as the average of each allele’s individual effect or as a secondary effect regarding possible allele interaction21,22. The foregoing considers that some haplotypes were found having consistent effects with that found for the alleles constituting them, haplotypes with alleles not associated independently and haplotypes formed by an associated allele and another non-associated allele.

Future studies should consider additional variables such as changes regarding the amount of sexual partners through follow-up and changes in viral load, thereby broadening understanding of outcomes and supporting the conclusions, mainly regarding this study’s redetection events. Although biological assays are required to support the bioinformatics findings described here, this has been the first study predicting peptides which could favour or hinder viral persistence (in viral types different to HPV-16 and -18) and demonstrating their potential usefulness as therapeutic anti-HPV peptide vaccines.

This study has thus reported, for the first time, alleles and haplotypes (typed by NGS) associated with clearance/persistence and redetection events, as well as L1 and L2 epitopes from the six most frequently occurring HPV types which are responsible for around 85% of CC cases. The information obtained through this study provides relevant knowledge for understanding the genetic component in the immune response against HR-HPV types and the natural history of CC. The associations described here enable constructing the bases for future studies aimed at evaluating the impact and effectiveness of anti-HPV vaccines and treatments (current and future), considering each population’s genetic particularities.

**Materials and Methods**

**Study population.** The study’s target population consisted of a cohort of women who had been participating in a Fundación Instituto de Inmunología de Colombia (FIDIC) multicentre study evaluating 6 HR-HPV types’ persistence, clearance and reinfection19. The women had been attending hospitals in the Colombian cities of Bogotá, Girardot and Chaparral between January 2007 and March 2010; they voluntarily participated in the study and declared that they had not changed their place of residence in at least two years after the start of the study. Detailed information regarding the study population and the procedures related to the 6 HR-HPV types’ detection and quantification have been described in previous publications21-23.

Women having had at least three follow-ups (6 ± 3-month intervals) and real-time PCR viral identification results were included in the study. Women whose cervical samples did not have the minimum amount of DNA and/or required quality for HLA-DRB1 and DQB1 typing were excluded from the analysis (Supplementary Fig. S2A). All the methods were performed in accordance with the Helsinki declaration and the Colombian Ministry of Health and Social Protection guidelines, as approved by the Universidad Nacional de Colombia’s Faculty of Medicine’s Research Ethics Committee (resolution 004-067-18). All the women who agreed to participate in the study signed an informed consent form.

**DNA extraction.** A QIAamp 96 DNA Blood Kit (QIAGEN, Hilden, Germany) was used for extracting genomic DNA (gDNA) from the 276 aliquots containing the cervical samples (stored in FIDIC’s DNA bank), following the manufacturer’s recommendations and specifications. The cervical samples’ gDNA quality for HLA typing was determined as >10 ng/μL concentration and ≥1.8/2.0 absorbance ratio at 260/280 nm.

**Detecting and quantifying human papillomavirus.** Conventional PCR was used for detecting six HR-HPV types (HR-HPV-16, -18, -31, -33, -45 and -58) and qPCR was used for quantifying the viral load,
following previously described procedures23,37,38. The human β-globin gene was amplified by PCR for evaluating DNA integrity, using GH20/PC04 primers. This was followed by using two sets of generic primers targeting a region of the L1 gene for the generic detection of HPV (GP5+/GP6+ and MY09/MY11) (Supplementary Table S20)23,38.

Samples proving positive by at least one generic primer set were typed using primers targeting a region located on the E5, E6 and E7 genes; these were specific for HR-HPV-16, -18, -31, -33, -45 or -58 (Supplementary Table S20). All amplification products were visualised on 2% agarose gels. Synthetic genes from early HPV-18, -31, -45 and -58 regions of interest and samples proving positive and confirmed by Sanger sequencing for HPV-16 and -33 were used as positive controls. DNA-free water was used as negative control23,38.

Real-time PCR detection was used for determining the amount of viral copies and the CFX96 Touch qPCR detection system was used for analysis; the primers and TaqMan probes used here are described in Supplementary Table S21. 1:10 (1011–1010) serial dilutions were obtained from plasmid DNA (known concentration) from each viral type and the HMBS gene for making the calibration curve. The human HMBS gene was used for validating DNA integrity and determining the amount of viral copies per cell. The samples were analysed for the aforementioned six HR-HPV types, involving absolute (total HPV copies) and relative quantification (HPV copies/cell = HPV copies/(HMBS/2 copies)) of each type’s load. Six dilutions of plasmid DNA were included for each type and included as controls for HMBS in each analysis along with a negative NTC control (no template control)23,38.

**HLA typing.** Illumina MiSeq (San Diego, CA, USA; Histogenetics, Ossining, NY, USA) was used for typing HLA-DRB1 and DQB1 allele and haplotype frequencies. Persistence was considered as being the detection of the same type of HPV in at least two consecutive visits, whilst clearance was determined as no viral detection in two consecutive samples after a positive detection when analysing infection by each of the 6 HPV types (-16, -18, -31, -33, -45 and -58). Redetection was taken as being positive for the same viral type following its prior non-detection (regardless of infection clearance) (Supplementary Fig. S2B)23,42,43. As some women had had more than one positive HPV-related event during follow-up, specific HR-HPV type infections were taken as the unit of analysis (i.e. not the women).

Cox-proportional hazard and accelerated failure time (AFT) models were adjusted when the assumption of proportionality was not met to evaluate allele and HLA-DRB1 and DQB1 relationship with type-specific infection (clearance/redetection) outcome44. Multicollinearity between the variables included in the models was also evaluated by VIF and tolerance.

Every AFT models’ goodness of fit was evaluated in line with Akaike information criterion (AIC) and Bayesian information criterion (BIC), selecting the model having the best fit (lowest AIC and BIC values)45. Hazard ratios (HR) and time ratios (TR) were reported, depending on the model used46,47. Independent models were run for each HPV type. TR values less than 1 denoted an earlier occurrence of an event, whilst values greater than 1 indicated a later occurrence of such event46,47.

A p < 0.200 value in univariate analysis was taken when selecting the independent variables for adjusting the models and the change in the crude estimator when added to the models. Origin, age, the amount of sexual partners, a background of abortion, coinfection (i.e. infection by 2 or more HR-HPV types) and absolute viral load (categorised as low: ≤9.99E+5, medium: from 1.00E+6 to 9.99E+9 and high: ≥1.00E+10) were taken as independent variables.

The Bonferroni method was used for correcting each model’s raw p values, considering the multiple alleles and haplotypes identified in the study population and that there was no a priori hypothesis concerning the associated alleles/haplotypes23,48. Two-tailed tests were used for hypotheses testing (0.05 significance) and STATA14 software was used for the aforementioned analysis.

**Predicting epitopes for HLA-II alleles.** The Immune Epitope Database (IEDB) was used for predicting binding peptides (20 aa long) for every HLA-II allele49, using the Technical University of Denmark’s Systems Biology Department’s Center for Biological Sequence Analysis’ NetMHCIipan 3.2 server prediction method50. Peptides predicted to have a strong binding threshold (<5% Rank) were sought manually by aligning each HR-HPV type’s complete L1 or L2 capsid protein reference sequences (GenBank codes: HPV16-L1: AANA05539.1; HPV18-L1: AGG40789.1; HPV31-L1: AEI60956.1; HPV33-L1: AEI11168.1; HPV45-L1: ABB99855.1; HPV58-L1: BBA20221.1; HPV16-L2: AFS44645.1; HPV18-L2: AAP20660.1; HPV31-L2: AIG59270.1; HPV33-L2: AMY16574.1; HPV45-L2: ALV85694.1; HPV58-L2: AMY16537.1). This was aimed at finding similar or different regions and determining possible explanations for the associations found between HLA-II and HR-HPV. Peptide sequences coinciding with the secretion signal sequence were not considered for analysis.

**Data availability**

The datasets produced and/or analysed during this study are available from the corresponding author on reasonable request.

Received: 14 November 2019; Accepted: 25 March 2020; Published online: 29 April 2020
42. Jaisamrarn, U. et al. Natural history of progression of HPV infection to cervical lesion or clearance: analysis of the control arm of the large, randomised PATRICIA study. *PLoS One.* 8, e79260, https://doi.org/10.1371/journal.pone.0079260 (2013).
43. Cuschieri, K. et al. Influence of HPV type on prognosis in patients diagnosed with invasive cervical cancer. *Int J Cancer* 135, 2721–2726, https://doi.org/10.1002/ijc.28902 (2014).
44. Grambsch, P. M. & Therneau, T. M. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika.* 81, 515–526, https://doi.org/10.1093/biomet/81.3.515 (1994).
45. Saikia, K. & Pratim, M. A Review on Accelerated Failure Time Models International Journal of Statistics and Systems 12, 311–322 (2017).
46. Bradburn, M. J., Clark, T. G., Love, S. B. & Altman, D. G. Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *Br J Cancer* 89, 431–436, https://doi.org/10.1038/sj.bjc.6601119 (2003).
47. Bradburn, M. J., Clark, T. G., Love, S. B. & Altman, D. G. Survival analysis Part III: multivariate data analysis—choosing a model and assessing its adequacy and fit. *Br J Cancer* 89, 605–611, https://doi.org/10.1038/sj.bjc.6601120 (2003).
48. Armstrong, R. A. When to use the Bonferroni correction. *Ophthalmic Physiol Opt* 34, 502–508, https://doi.org/10.1111/opo.12131 (2014).
49. Wang, P. et al. Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinformatics. 11, 568, https://doi.org/10.1186/1471-2105-11-568 [pii] (2010).
50. Nielsen, M. et al. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol* 4, e1000107, https://doi.org/10.1371/journal.pcbi.1000107 (2008).

Acknowledgements
This work was financed by Colombia’s General Royalties System, regarding a project [project BPIN-233, special agreement 021] entitled, “Developing a management plan for controlling cervical cancer in the Amazonas department,” which was approved by the Amazonas Department Local Government and Colombia’s the Science, Technology and Innovation Department (COLCIENCIAS).” Neither of the aforementioned entities participated in developing the study. We would like to thank the Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS) for financing MC as part of the National Programme for Promoting Researcher Training – Call 617 (PhD study in Colombia). We would like to thank Indira María Báez Murcia for providing technical support and Jason Garry for translating the manuscript.

Author contributions
Conceived and designed the study, analysed and interpreted the data and prepared the manuscript: L.D.R.O. and M.C; Designed the study, collected and interpreted data: S.C.S.D.L. Analysed the data and have interpreted the results: R.S. Analysed the data, have interpreted the results and drafted the manuscript: D.A.M.P. Conceived and designed the study and revised the manuscript: M.E.P. and M.A.P. All authors have reviewed and approved the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-64268-x.
Correspondence and requests for materials should be addressed to M.A.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2020