An analysis of clustering of betapapillomavirus antibodies

K. A. Mallitt,1 P. O'Rourke,1 J. N. Bouwes Bavinck,2 D. Abeni,3 M. N. C. de Koning,4 M. C. W. Feltkamp,5 A. C. Green,1 W. G. V. Quint,4 K. M. Michael,6 M. Pawlita,6 H. Pfister,7 S. Weissenborn,7 T. Waterboer,6 R. E. Neale1 and the EPI-HPV-UV-CA Group

1Cancer and Population Studies Group, Queensland Institute of Medical Research, Brisbane, Australia
2Department of Dermatology, Leiden University Medical Center, Leiden, The Netherlands
3Istituto Dermopatico dell'Immacolata, IDI-IRCCS, Rome, Italy
4DDL Diagnostic Laboratory, Voorburg, The Netherlands
5Department of Medical Microbiology, Leiden University Medical Center, Leiden, The Netherlands
6German Cancer Research Center (DKFZ), Heidelberg, Germany
7Institute of Virology, University of Cologne, Cologne, Germany

Correspondence
R. E. Neale
rachel.neale@qimr.edu.au

Received 27 October 2009
Accepted 13 April 2010

Betapapillomaviruses (βPVs) may contribute to the aetiology of cutaneous squamous cell carcinoma. However, no high-risk types have yet been identified, possibly because the high frequency of co-infection prevents a straightforward analysis of the independent effects of individual viruses. This study aimed to determine whether specific virus types were more likely to co-occur than others, thereby reducing the number of parameters needed in statistical models. Antibody data were analysed from controls who participated in case–control studies in The Netherlands, Italy and Australia and from participants in the German Nutrition Survey. Cluster analysis and two ordination techniques were used to identify patterns. Evidence of clustering was found only according to the number of viruses to which antibodies were detected. The lack of clustering of specific viral types identified suggests that if there are βPV types that are independently related to skin carcinogenesis, they are unlikely to be identified using standard epidemiological methods.

INTRODUCTION

Papillomaviruses are a family of non-enveloped, double-stranded DNA viruses associated with benign and malignant lesions of cutaneous and mucosal epithelia (zur Hausen, 1999).

Human papillomaviruses (HPVs) belonging to the genus Betapapillomavirus are thought to cause skin cancers in patients suffering from the rare hereditary disorder epidermodysplasia verruciformis (EV) (Majewski & Jablonska, 1995). Evidence is emerging to suggest that they may also play a contributory role in people without EV. Viral DNA is found in a high proportion of actinic keratoses and cutaneous squamous cell carcinomas (SCCs) (Harwood et al., 2000), but they do not integrate into the host genome and are not present in every cell, making their role in tumour development uncertain. Case–control studies have been conducted to examine further the relationship between betapapillomaviruses (βPVs) and SCC, in which βPV DNA in plucked eyebrow hairs, and serum antibodies, have been used as markers of infection. Overall, these studies have found associations between the presence of viral DNA or antibodies and SCC (Feltkamp et al., 2003; Karagas et al., 2006; Struijk et al., 2003, 2006), although the data have been somewhat inconsistent. Importantly, no high-risk types have been identified. This may partly be due to the frequency of co-infection. Any association with one type could simply be due to co-infection with another type. Due to the number of types and the degree of collinearity between them, standard statistical modelling techniques are unable to determine the independent effects of individual virus types.

To contribute to our understanding of the role that βPVs may play in the development of skin cancer, we sought to determine whether antibodies to particular βPV types co-occurred within individual people. By reducing the number...
of parameters needing to be included in statistical models and ensuring their relative independence, specific viral clusters that might be associated with skin cancer may be detected in epidemiological studies.

**RESULTS**

The mean age of the European Community (EC) study population was 65 years and of the VERA population 53 years (see Methods). Both populations contained just over 40% men. When the prevalence of any βPV antibody was adjusted for age and sex by direct standardization to the European population (Table 1), prevalence in the three countries included in the EC-funded study did not differ greatly from each other and was approximately 50%. The three countries were therefore combined for all analyses and referred to as the EC population. In contrast, the overall prevalence of any βPV antibody in the VERA population (36%) was lower than the EC population prevalence, and this was true for most of the individual viruses tested.

Among those who tested positive for βPV antibodies in the EC population, 20% [95% confidence interval (CI), 17–23%] had one virus antibody detected, 14% (95% CI, 12–16%) had two to three detected and 23% (95% CI, 20–26%) had four or more detected. In the VERA population, 14% (95% CI 12–16) had one virus antibody detected, 10% (95% CI 8–12%) had two to three detected and 10% (95% CI 8–12%) had four or more.

Fig. 1 and Supplementary Fig. S1 (available in JGV Online) show dendrograms based on cluster analyses for EC and VERA data, respectively. In the EC population, there were two major clusters grouping the participants, whilst in the VERA data there were three. Callinski and Harabasz’s index (pseudo-F) (see Supplementary Table S1, available in JGV Online) confirmed these findings.

Generally, in both populations, membership of cluster 1 was driven by low overall median fluorescence intensity (MFI) values for all βPV types (see Supplementary Table S2, available in JGV Online). In the EC population, membership of cluster 2 was driven by high overall MFI values for all βPV types, but MFI values were particularly high for HPV-8, -38 and -49, and low for HPV-93. In the VERA study, MFI values for all βPV types were intermediate in cluster 2 and high in cluster 3, but cluster 3 was also notable for high HPV-38 and -49 values, and low MFI for HPV-93.

There were significant associations (P<0.001) between the number of βPVs to which antibodies were detected and cluster membership for the EC and VERA populations (Table 2). Clusters that associated with higher mean MFI values also contained people with seroreponses for a higher number of viruses.

The results of the principal components analysis (PCA) and principal coordinates analysis (PCoA) (data not shown) confirmed the findings of the cluster analysis. PCA found that approximately 60% of the variability in the serum data was driven by the magnitude of MFI values for both populations. PCoA found that the only detectable grouping of participants was driven by the number of viruses for which antibodies were detected.

### Table 1. Age- and sex-standardized prevalence of βPV antibodies in the EC and VERA studies

<table>
<thead>
<tr>
<th></th>
<th>Percentage prevalence (95% CI)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VERA (&gt;35 years)</td>
<td>EC population serum</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall</td>
<td>Australia</td>
<td>Italy</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>Any βPV</td>
<td>36 (32–40)</td>
<td>51 (43–60)</td>
<td>44 (32–56)</td>
<td>50 (35–66)</td>
</tr>
<tr>
<td>HPV-5</td>
<td>8 (6–10)</td>
<td>10 (5–14)</td>
<td>11 (3–19)</td>
<td>12 (2–21)</td>
</tr>
<tr>
<td>HPV-8</td>
<td>16 (12–19)</td>
<td>29 (22–37)</td>
<td>28 (17–39)</td>
<td>32 (19–46)</td>
</tr>
<tr>
<td>HPV-9</td>
<td>10 (8–12)</td>
<td>11 (8–14)</td>
<td>9 (5–13)</td>
<td>15 (7–22)</td>
</tr>
<tr>
<td>HPV-17</td>
<td>10 (8–12)</td>
<td>21 (15–26)</td>
<td>20 (11–28)</td>
<td>23 (12–34)</td>
</tr>
<tr>
<td>HPV-20</td>
<td>7 (5–9)</td>
<td>12 (8–15)</td>
<td>12 (6–19)</td>
<td>10 (4–17)</td>
</tr>
<tr>
<td>HPV-23</td>
<td>5 (3–6)</td>
<td>12 (7–17)</td>
<td>13 (5–21)</td>
<td>12 (5–18)</td>
</tr>
<tr>
<td>HPV-24</td>
<td>7 (5–8)</td>
<td>13 (9–17)</td>
<td>14 (9–18)</td>
<td>13 (6–20)</td>
</tr>
<tr>
<td>HPV-38</td>
<td>12 (9–14)</td>
<td>23 (17–29)</td>
<td>23 (14–32)</td>
<td>21 (11–32)</td>
</tr>
<tr>
<td>HPV-49</td>
<td>13 (11–16)</td>
<td>21 (16–27)</td>
<td>17 (12–23)</td>
<td>22 (11–32)</td>
</tr>
<tr>
<td>HPV-75</td>
<td>7 (6–9)</td>
<td>14 (9–19)</td>
<td>9 (5–13)</td>
<td>15 (5–24)</td>
</tr>
<tr>
<td>HPV-76</td>
<td>6 (4–8)</td>
<td>11 (7–15)</td>
<td>10 (5–14)</td>
<td>13 (6–21)</td>
</tr>
<tr>
<td>HPV-92</td>
<td>8 (6–10)</td>
<td>10 (7–13)</td>
<td>8 (4–13)</td>
<td>13 (6–19)</td>
</tr>
<tr>
<td>HPV-93</td>
<td>1 (0–2)</td>
<td>2 (1–3)</td>
<td>2 (0–3)</td>
<td>5 (0–10)</td>
</tr>
</tbody>
</table>
DISCUSSION

The frequency of βPV co-occurrence makes it difficult to assess any independent effects of specific viruses, as any effect attributed to one virus may be due to co-infection with others. We attempted to resolve this problem by determining whether antibodies to specific virus types were more likely to co-occur than others. This would enable a reduction in the number of parameters needing to be included in statistical models, and ensure their relative independence, overcoming the problems of collinearity.

Our results highlight the prevalence of antibody positivity. Using the specified cut-off, over half of the EC population and one-third of the VERA population had antibodies to at least one βPV and this prevalence was very robust to changes in the cut-off value used (±100 MFI). Moreover, a very high proportion of people had evidence of multiple infections. Of the EC and VERA participants who were antibody positive, 37% and 20% had antibodies to multiple virus types, respectively. These results are similar to those found in a US population (Karagas et al., 2006).

We conducted the same analyses described here for viral DNA detected in eyebrow hair follicles (for detection methods and description of overall prevalence, see de Koning et al., 2009). Unlike a study of viral clustering in cervical tissue of patients with grade 1 cervical intraepithelial neoplasia (Spinillo et al., 2009), our results did not reveal any clustering according to βPV type or number of viruses (data not shown). In contrast, there was evidence of clustering of participants according to the number of viruses to which antibodies were detected. In the EC population data, two distinct clusters formed – one exemplified by people who had infrequent detection of multiple antibody types (only 2% had antibodies to four or more types) and one containing people with a high frequency of multiple antibody detection (88% had antibodies to four or more types). The VERA data revealed three clusters, with high and low groups as described above, but also an intermediate group. In both populations, the 'high antibody frequency' cluster was also characterized by an excess of HPV-38 and -49 antibodies. Although this could be attributed to their relatively high frequency, HPV-8 was the most frequent and only contributed to the high antibody frequency cluster in the EC but not in the VERA data.

Multiple seropositivity may result from type-specific reactions to multiple infections and/or could be due to cross-reactive antibodies induced by an infection with only one or a few HPV types. Although the antigens employed in the antibody detection assay used here have been shown to display virtually all conformational and linear epitopes displayed by virus-like particles (Rizk et al., 2008), the assay was not able to distinguish these two possibilities. Cross-reactivity can only be appropriately investigated by means of absorption experiments or monospecific antisera, which are currently not available for most HPV types. However, if multiple seropositivity was due to cross-reactivity, it should be associated with the degree of relatedness of HPV types, and a detailed analysis of the VERA samples revealed that multiple seropositivity was not or was only weakly correlated with L1 (the major capsid protein) amino acid sequence relatedness (Michael et al., 2008). Thus, we consider the measured antibodies to be mainly type specific.

Assuming that the division of people into those with antibodies to many viruses and those with antibodies to

<table>
<thead>
<tr>
<th>No. viruses</th>
<th>EC population data</th>
<th>VERA population data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td>Cluster 2</td>
</tr>
<tr>
<td>0</td>
<td>343 (56)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>1–3</td>
<td>255 (42)</td>
<td>22 (11)</td>
</tr>
<tr>
<td>4+</td>
<td>11 (2)</td>
<td>175 (88)</td>
</tr>
<tr>
<td>Total</td>
<td>609</td>
<td>198</td>
</tr>
</tbody>
</table>
Clustering of betapapillomavirus antibodies

none or a few viruses is real, there are several possible explanations. One is that antibodies are a good marker of pathogenic skin infection, and variation in susceptibility to infection might therefore underpin the clustering. There are two main genes that are thought to determine such susceptibility, the EVER1 and EVER2 genes, which are mutated in patients with EV (Ramoz et al., 2002). A recent study showed that controls in a case–control study who carried a particular single-nucleotide polymorphism in the EVER2 gene were more likely to carry βPV antibodies and had a greater likelihood of having antibodies to multiple viruses (Patel et al., 2008). The finding that the polymorphic form of EVER2 is associated with SCC (Patel et al., 2008) suggests that it is not simply driving the antibody response to infection.

An alternative explanation is that the clusters are due to other differences between participants that are related to the presence and/or multiplicity of βPV antibodies. However, in the EC population no factors other than age and sex were shown to be associated with βPV antibodies (Michael et al., 2008; Waterboer et al., 2009) and there was no association found between cluster membership and age or sex, indicating that these demographic variables are not related to cluster formation.

The statistical methods used here are hypothesis generating and cannot be interpreted as a hypothesis test for the presence of βPV clusters. There are a variety of algorithms applicable for use when performing cluster or ordination analysis, depending on the form of the data. These different algorithms often produce variable results. However, in this case, the results were consistent across methodologies, providing strong evidence for a lack of multivariate structure in βPV DNA or antibody data.

No βPVs have been classified by the International Agency for Research on Cancer as carcinogenic to humans. HPV-5 and -8 have been classified as possibly carcinogenic in patients with EV (group 2A), with the remaining βPV types currently considered unclassifiable (group 3). This is partly because no study to date has shown an unequivocal association between cancer and any specific type among non-EV patients. Evidence is accumulating to suggest that the presence of antibodies to multiple βPV types may be associated with the aetiology of cutaneous SCC (Karagas et al., 2006; Struijk et al., 2006), although it may be that multiplicity is a marker of infection with a specific oncogenic type or group of βPVs. Our study suggests that people who are unaffected by SCC can be discriminated on the basis of the number of viruses to which antibodies can be detected, lending some support to the hypothesis that multiplicity of infection is causally associated with SCC. The lack of clustering of specific viral types identified here suggests that if there are specific βPV types that are independently aetiologically related to skin carcinogenesis, they are unlikely to be identified using standard epidemiological methods.

METHODS

Participants. The participants in this study were drawn from several different populations. We first analysed βPV data from only the control participants who took part in a multi-centre EC-funded case–control study investigating the relationship between βPV infection, UV radiation and skin cancer carried out in Leiden (The Netherlands), Rome (Italy) and Queensland (Australia). The methods of this study have been reported previously (de Koning et al., 2009; Waterboer et al., 2009). Briefly, in The Netherlands, 275 controls with no history of any type of skin cancer who were age- and sex-matched to a series of cases with cutaneous SCC were recruited at the ophthalmology outpatient clinic at the Leiden University Medical Center (Struijk et al., 2003). In Rome, 256 control participants were selected from the dermatology outpatient clinic at the Istituto Dermopatico dell’Immacolata. In Australia, 276 eligible controls were enlisted: 133 randomly selected from the Electoral Roll, a register with almost complete population coverage, 58 from skin cancer-free patients attending primary practice skin cancer clinics for routine self-referred screening and 85 from a series of 14 community groups. The Dutch, Italian and Australian populations are together referred to as the EC population in this study.

To validate the results of analyses on the EC data, we analysed serum antibody data from participants in the VERA study, who were a random subsample of people who had participated in the population-based German Nutrition Survey in the late 1980s (Michael et al., 2008). This analysis was restricted to the 998 people aged over 35 years at the time of blood collection.

All studies were approved by the relevant institutional ethics committee.

Serum antibody detection. The analysis of serum samples for HPV antibodies has been described in detail elsewhere (Waterboer et al., 2005). Briefly, sera were analysed simultaneously using multiplex serology for 15 βPV types (HPV5, -8, -9, -15, -17, -20, -23, -24, -36, -38, -49, -75, -76, -92 and -93). Glutathione S-transferase–βPV L1-tagged fusion proteins from cleared lysates were affinity-purified in situ through binding to glutathione–casein-coated fluorescence-labelled polystyrene beads. Each fusion protein was bound to a spectrally distinct bead set, and fusion protein-loaded bead sets were mixed. Sera were incubated with the mixed bead sets at a final dilution of 1:100, and bound antibodies were detected with biotinylated goat anti-human IgG (H+L) secondary antibody and streptavidin–R-phycocerythrin. A Luminex xMAP analyser was used to identify the internal colour of the individual beads and to quantify their reporter fluorescence, expressed as MFI of at least 100 beads per set per serum.

Statistical analysis. Serum MFI values were analysed as continuous variables for all multivariate analyses, but were categorized into binary presence/absence form at the cut-off point of MFI ≥ 200 to calculate βPV prevalence. Negative values were corrected to 0, and the variables were ‘natural log + 1’-transformed.

The standard European population was used to age- and sex-standardize the VERA and EC serum antibody prevalence data directly (overall and separately by country) to compare βPV antibody prevalence between these populations.

One classification technique (cluster analysis) was used to identify multivariate trends in the data and two ordination techniques, PCA and PCoA, were used to verify the results of the cluster analysis.

For the cluster analysis, we calculated pair-wise distances between participants using Euclidean distance for the continuous data.
Agglomerative hierarchical clustering with Ward’s linkage method was used. We created dendrograms to visualize cluster formation, and the number of meaningful clusters was determined using Callinskie and Harabasz’s index (pseudo-F). The relationship between clusters and MFI was assessed using means, and the relationship between cluster and the number of βPV infections was assessed with a $r^2$ statistic (Williams, 1976).

For the PCA, an association matrix was created between variables (βPV types), and then through an eigenanalysis of the matrix, new axes that represent the major sources of variability in the data were identified. Eigenvectors with an eigenvalue of >1 were retained in the analysis, and the analysis was based on correlation matrices. Pearson correlations were used (Tabachnick & Fidell, 1996).

PCoA is an ordination technique based on a distance matrix between participants, which was calculated using Euclidean distance. An association matrix was created between variables (βPV types), and then through an eigenanalysis of the matrix, new axes that represent the major sources of variability in the data were identified. Eigenvectors with an eigenvalue of >1 were retained in the analysis, and the analysis was based on correlation matrices. Pearson correlations were used (Tabachnick & Fidell, 1996).

REFERENCES


