frontiers
in Oncology

Human Papilloma Viruses and Breast Cancer

James S. Lawson[1], Wendy K. Glenn[1], Daria Salyakina[2], Warick Delprado[3],
Rosemary Clay[3], Annika Antonsson[4], Benjamin Heng[1], Shingo Miyauchi[1], Dinh D. Tran[1],
Christopher C. Ngan[1], Louise Lutze-Mann[1] and Noel J. Whitaker[1]*

[1] School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, NSW, Australia, [2] Center for
Computational Science, University of Miami, Miami, FL, USA, [3] Douglass Hanly Moir Pathology, Macquarie Park, NSW,
Australia, [4] QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

Purpose: Human papillomaviruses (HPV) may have a role in some breast cancers. The purpose of this study is to fill important gaps in the evidence. These gaps are: (i) confirmation of the presence of high risk for cancer HPVs in breast cancers, (ii) evidence of HPV infections in benign breast tissues prior to the development of HPV-positive breast cancer in the same patients, (iii) evidence that HPVs are biologically active and not harmless passengers in breast cancer.

Methods: RNA-seq data from The Cancer Genome Atlas (TCGA) was used to identify HPV RNA sequences in breast cancers. We also conducted a retrospective cohort study based on polymerase chain reaction (PCR) analyses to identify HPVs in archival specimens from Australian women with benign breast biopsies who later developed breast cancer. To assess whether HPVs in breast cancer were biologically active, the expression of the oncogenic protein HPV E7 was assessed by immunohistochemistry (IHC).

Results: Thirty (3.5%) low-risk and 20 (2.3%) high-risk HPV types were identified in 855 breast cancers from the TCGA database. The high risk types were HPV 18 (48%), HPV 113 (24%), HPV 16 (10%), HPV 52 (10%). Data from the PCR cohort study indicated that HPV type 18 was the most common type identified in breast cancer specimens (55% of 40 breast cancer specimens) followed by HPV 16 (13%). The same HPV type was identified in both the benign and subsequent breast cancer in 15 patients. HPV E7 proteins were identified in 72% of benign breast specimens and 59% of invasive breast cancer specimens.

Conclusion: There were four observations of particular interest: (i) confirmation by both NGS and PCR of the presence of high-risk HPV gene sequences in breast cancers, (ii) a correlation between high-risk HPV in benign breast specimens and subsequent HPV-positive breast cancer in the same patient, (iii) HPVs in breast cancer are likely to be biologically active (as shown by transcription of HPV DNA to RNA plus the expression of HPV E7 proteins), (iv) HPV oncogenic influences may occur early in the development of breast cancer.

Keywords: breast cancer, benign breast, human papilloma virus, HPV E7 proteins, retrospective cohort study

INTRODUCTION

High risk for cancer human papilloma viruses (HPVs) have been identified in breast cancers in 30 studies conducted in 17 countries and 4 continents (1, 2). HPV type 33 is the most prevalent (14%) but is mainly confined to Asian women, followed by HPV 18 (7%) and 16 (7%) (1, 2). In 10 case control studies, the prevalence of HPVs in breast cancers was significantly higher than in controls OR = 3.60 (1). Women with HPV-associated cervical neoplasia can later develop HPV-associated breast cancer (3, 4). However, several groups have not identified HPVs in breast cancer (5). All of the above studies have been based on polymerase chain reaction techniques (PCR).

High-risk HPVs have been identified in breast cancers by Next Generation Sequencing – NGS (also known as massive parallel sequencing)[1]. These studies have been based on The Cancer Genome Atlas (TCGA) database of 855 breast cancers. They demonstrate that high-risk HPVs are present in <2% of breast cancers and at very low viral loads. Recently, 1.3% of the TCGA specimens have been shown to be contaminated with HPV type 18 with the source probably HPV-positive HeLa laboratory cells (6). This contamination does not appear to affect the TCGA breast cancer specimens, which we have studied and reported in this current article (Cantalupo, 2015, personal communication). However, minor contamination cannot be excluded.

It is difficult to reconcile the data based on TCGA and analyzed by NGS, with the data based on Australian breast cancer specimens and analyzed by PCR. The TCGA data indicate a 2% prevalence of high-risk HPV-positive breast cancers compared to 60% of Australian breast cancers in this current study. However, of this 60% only approximately 40% had observable protein expression. Other studies based on PCR, such as the Ohba et al. study of breast cancer in Singapore residents (7), and prior studies of Australian breast cancers, have identified high-risk HPVs in from 25 to 100% of breast cancers (8, 9). The most likely explanation is the use of different viral identification techniques, namely, NGS as compared to PCR. NGS is a relatively new technique. Positive identification of HPV and other viruses is dependent on the recombination of fragmented gene sequences known as "reads." There are no established conventions concerning the best length of the recombined "reads." Therefore conclusions are in part, matters of opinion. While this may explain some of the different outcomes of NGS as compared to PCR, it is clear that the HPV viral load in breast cancer is extremely low as compared to cervical cancer.

While these data indicate that HPVs may have a possible role in some breast cancers, they may not have a major causal role. There are two reasons for this view: (i) in immunocompromised patients (due to either human immunodeficiency infections or post organ transplantation therapy), there is no increased prevalence of breast cancer, which is in contrast to the two- to sixfold increased prevalence of HPV associated cervical and head and neck cancer in these patients (10), and (ii) the HPV viral load in breast cancer is extremely low (11, 12).

However, it is important to determine whether high-risk HPVs have a causal role in even a small proportion of breast cancers. In addition to the epidemiologically based evidence outlined above, there is substantial but incomplete evidence, which indicates a potential role for HPVs in breast cancer. This evidence includes: (i) HPVs immortalize and transform human mammary epithelial cells (13), (ii) HPVs are located in breast cancer cell nuclei (8), (iii) HPV E6 oncoprotein has been identified in breast cancers (8), (iv) high-risk HPVs have been identified in breast cancer cell cultures (8), (v) HPV-positive breast cancer is more common in young as compared to older women, an observation which is compatible with sexual transmission of HPV among sexually active younger women (14, 15), (vi) approximately 16% of normal Australian women are serologically HPV type 16 positive (16), and (vii) HPV-positive koilocytes have been identified in breast cancers (17).

This evidence is incomplete. There is a need to determine (i) if high-risk HPVs are present in normal and benign breast tissues prior to the development of HPV-positive breast cancer and (ii) whether these HPVs are biologically active. These form the aims of this study.

MATERIALS AND METHODS

Research Approaches

The research approach to the first aim – identification of HPVs – was to use a range of different methods for the identification of HPVs. These methods included Next Generation Sequencing (NGS), *in situ* and standard PCR, and immunohistochemistry. The approach to the second aim – identification of HPVs in benign non-malignant breast tissues prior to breast cancer – was conducted through a retrospective cohort study. The approach to the third aim – to determine the biological activity of HPVs in breast cancer – was to study HPV RNA sequences in breast cancers as indicators of transcription activity plus a study of HPV associated oncogenic protein expression.

TCGA Next-Generation Sequencing Data and Bioinformatics Analysis

Whole transcriptome sequencing data for 855 (USA) Breast cancers were obtained through TCGA. Sequencing has been done using Illumina (Solexa, GAII) or SOLiD™ technology. A detailed description of the TCGA data can be found on the following TCGA websites: http://cancergenome.nih.gov/, https://tcga-data.nci.nih.gov/tcga/. Transcriptome data (BAM files) generated by TCGA were analyzed in an automated fashion on a computational cluster hosted by the High-Performance Computing core at the Center for Computational Science, University of Miami[2]. All non-viral sequences were subtracted

---

[1] http://larssonlab.org/tcga-viruses/report_BRCA.php

Abbreviations: APOBEC3B, apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B; CIN, cervical intraepithelial neoplasia; EBV, Epstein–Barr virus; ER, estrogen receptors; HER 2 (also known as ErbB-2), human epidermal growth factor receptor 2 protein; HPV, human papilloma virus; HPV E7, human papilloma virus E7 protein; IHC, immunohistochemistry; NGS, next generation sequencing; PCR, polymerase chain reaction; PR, progesterone receptors; *r*, correlation co-efficient; TCGA, the cancer genome atlas.

---

[2] http://ccs.miami.edu/

from NGS data as described in Salyakina and Tsinoremas (18). Reference genomes for all known HPV types obtained from PaVE database[3] were used for HPV screening. Even the single short read 48 nt or longer aligning to the viral reference was considered as successful detection, if following BLAST analysis versus the NCBI nucleotide collection confirmed sequence similarity with the target over 98%. Sequences aligning to multiple organisms, known artificial (vector) sequences, or low complexity sequences were considered false positive and removed.

The outcomes for HPV 18 were compared to the outcomes of the Larsson lab for the same TCGA breast cancer specimens.

APOBEC gene family mutations and expression data for TCGA cohort were available for 850 out of 855 analyzed samples from http://www.cbioportal.org. The Fisher exact test was employed to test possible association of HPV status with APOBEC genes expression or mutations.

Clinical data was available for 836 out of 855 patients from the TCGA portal.

These analyses were conducted by Daria Salyakina at the Center for Computational Science, University of Miami, Coral Gables, FL, USA.

Retrospective Cohort Study

We conducted a retrospective cohort study to identify HPV sequences in benign breast biopsy tissues in women who later developed HPV-positive breast cancer. This is a practical and low-cost method of achieving sound outcomes. If the HPV DNA was of the same type and sequence in both the prior benign breast tissues as in the subsequent breast cancer tissues from the same patient this potentially provides evidence of prior HPV breast infection. All the specimens were from women residing in Australia. We reviewed approximately 4,000 pathology reports of breast cancer and identified 41 patients who had benign breast biopsies 1–11 years prior to developing breast cancer. The pathology reports of associated archival formalin fixed specimens from these patients were identified and collected from an Australian pathology service (Douglas Hanly Moir Pathology). The average age of the patients with benign breast conditions was 50.5 years and subsequent cancer 56.1 years.

We used normal breast specimens from 21 patients who had cosmetic breast surgery as a comparative group. The average age of these donors was 35.7 years. These patients were much younger than the patients who developed breast cancer, and therefore are not an exact control group. None of these patients had developed breast cancer 10 years after the initial surgery.

Identification of HPV Gene Sequences by Polymerase Chain Reaction

In situ PCR, semi-nested PCR, and real-time PCR were used for the detection of HPV. All PCR products were sequenced to help identify any contamination. Although *in situ* PCR can produce false positive outcomes, use of this method can add to the validity of results based on semi-nested and real time PCR.

---

[3]http://pave.niaid.nih.gov/

The detailed methods are presented in Supplementary Materials and Methods.

Preparation of Genomic DNA
For standard and real-time PCR, the methods were as described in Steinau et al. (19). The primers used were β-actin 5′ *forward* (5′-CTTCTGCCGTTTTCCGTAGG-3′) and β-actin 3′ *reverse* (5′-TGGGATGGGGAGTCTGTTCA-3′).

Standard PCR
The primers for semi-nested PCR My11 (5′-GCACAGGGYCAYAAYAATGG-3′) to modified GP6 (5′-AATCATATTCCTCMMCATGTC-3′). The second round was Gp5+ (5′-TATTTGTTACTGTKGTWGATAC-3′) to Gp6+. These primers were degenerate for HPV16 and 18, but were also capable of bringing up types 3, 11, 12, 45, 58, 73, and 75.

Real-Time PCR
The HPV L1 gene in gDNA samples was amplified using a real-time PCR machine (Rotor Gene Q, QIAGEN).

Sequencing the PCR Products and Identification of HPV Types
The HPV PCR products from GP5 to Gp6 were sequenced to determine the HPV type. The HPV genotypes were identified by BLAST via the US National Center for Biotechnology Information.

In situ PCR
These analyses were conducted as described in Heng et al. (8). *In situ* PCR is less susceptible to contamination and has the important advantage of localizing the specific genetic material at the cellular level.

The identification by PCR of HPV in benign non-malignant breast or breast cancer specimens was considered as positive if two or more of the following outcomes in the same specimen were observed: (i) HPV DNA sequences identified by standard PCR and/or real time PCR, (ii) HPV positive *in situ* PCR.

Confirmatory PCR Studies
DNA extracts from six selected (for positive HPV) benign and breast cancer specimens were independently analyzed by the Antonsson group at the QIMR Berghofer Medical Research Institute, Brisbane, for the presence of HPV DNA using their previously published methods (20).

HPV Biological Activity
An assessment of HPV biological and oncogenic activity was conducted by (i) the identification of HPV RNA sequences, which are an indication of HPV transcription activity (ii) the identification of HPV E7 proteins, (iii) the high expression of p16 protein, (iv) high expression of ER, and (v) inhibition of p53 protein expression.

Antibodies that are specific for HPV E7 proteins, have recently been developed (21). These antibodies (Cervimax) were used in this study. The specificity of these antibodies has been demonstrated experimentally and by epidemiological studies (21).

The HPV E7 antibody reacts with a wide range of HPV types including high risk for cancer HPV 16 and 18. HPV E7 antibodies have not been previously used on breast tissues.

p16

p16 contributes to regulation of the cell cycle. There is substantial evidence that indicates that p16 expression is associated with HPV biological activity (22). High expression of p16 as a surrogate for HPV E7 activity is a sound indication of transcriptionally active HPV (23). In breast tissues, negative or low expression of p16 is present in normal ductal epithelial tissues with a progressive increase of p16 expression in benign and malignant breast lesions (24–26). The situation is complicated by Epstein–Barr virus (EBV), which blocks the expression of p16 (27).

p53

The inhibition of expression of the cell death related p53 protein by HPV E6 protein is a well-documented mechanism for HPV oncogenesis. p53 expression is inhibited in HPV 16-positive breast cancer (28).

Estrogen Recptor

Estrogens have an essential role in breast cancer and high ER expression is a feature of most breast cancers (29). Recently, high ER expression has been associated with HPV-positive breast cancers (30).

HER 2 (ErbB-2)

The HER 2 receptor co-operates with HPV E6 and E7 oncoproteins in breast oncogenesis (31).

Immunohistochemistry

Standard manual IHC methods were used with the omission of the antigen retrieval step for the identification of HPV E7 protein expression. The antibodies were anti-HPV E7 monoclonal "Cervimax" – Valdospan GmbH, Austria. Positive controls for the E7 antibody were the Hela (HPV18) cell line and cervical tissues that were positive by PCR and sequencing.

Immunohistochemistry for estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER 2) were previously performed as part of the standard assessment of breast cancer specimens. For benign breast specimens, the same assessments were conducted as part of this current project.

ER, PR, p16, and p53 proteins are expressed in cell nuclei. HER 2 protein is expressed in cell cytoplasm and membranes. HPV types 16 and 18 E6 protein is expressed in both cell nuclei and cytoplasm. Anatomical pathologists from Douglass Hanly Moir – Pathology conducted the routine assessment of ER, PR, and HER in breast cancer specimens. Two independent observers (James S. Lawson and Wendy K. Glenn) used the same methods for the assessment of ER, PR, and HER in the benign breast specimens. The staining of breast tissues during immunohistochemistry procedures is heterogenous with staining frequently confined to only part of a tissue section. The assessments were based on those tissue sections with positive staining cells. The outcomes were assessed by the intensity of staining on a scale of 0 and 1–3. Positive (HPV-positive cervical

cancer specimen) and negative (antibody omitted) controls were used for each batch of specimens.

In breast cancer specimens, IHC staining for p16 appears to differ from the patterns commonly seen in cervical neoplasia. p16 expression in both non-invasive and invasive breast cancer specimens are most commonly seen in clusters of cancer cells with no staining in other parts of the cancer specimen (23). For this reason, we have adopted the same methods of assessment as outlined above for ER and PR but confined to assessments of percentages of positive and intensity of staining cancer cells in the clusters. We have adopted the same approach to assessment of p53, which also commonly stains in clusters in breast cancer specimens.

Statistics

Associations between HPV identification and various breast specimens were tested for correlations using the SPSS statistical package. The tests were all two-sided, and statistical significance was defined as $p \leq 0.05$. The numbers in this study are small. Accordingly, it is necessary to be cautious despite some of the results reaching statistical significance. The specific statistical tests used are listed with the different results sections.

RESULTS

Identification of HPV Gene Sequences in the TCGA Breast Cancers Cohort

Whole transcriptome data identified 30 (3.5%) tumors with low-risk and 20 (2.3%) with high-risk HPV types in 855 breast cancers from the TCGA cohort (Table S1 in Supplementary Material). The high risk types that were identified in the TCGA cohort were HPV 18 (50%), HPV 113 (20%), HPV 16 (10%), HPV 52 (10%) plus one identification of HPVs 31 and 35 (5% each) (**Table 1**). Three specimens had two HPV types.

As outlined in the Section "Introduction" of this report, 1.3% of the TCGA specimens have been shown to be contaminated with HPV type 18. The source of contamination was probably

---

TABLE 1 | HPV types in TCGA breast cancers – next generation sequencing.

High-risk HPVs $n = 20$	
HPV 16 – 2 (10%)	HPV E1, E2, E6, E7, and L1 identified
HPV 18 – 10 (50%)	HPV E1, E6, and E7 identified
HPV 31 – 1 (5%)	HPV E7 identified
HPV 35 – 1 (5%)	HPV E7 identified
HPV 52 – 2 (10%)	HPV E2, E4, E5, and E7 identified
Low-risk HPVs $n = 30$	
HPV 4 – 3 (10%)	
HPV 17 – 3 (10%)	
HPV 22 – 2 (7%)	
HPV 38 – 2 (7%)	
HPV 49 – 3 (10%)	
HPV 65 – 2 (7%)	
HPV 113 – 4 (20%)	HPV L1, L2, E7 identified

HPVs 1, 4, 5, 10, 15, 17, 19, 22, 23, 24, 38, 49, 63, 65, 80, 93, 104, 110, 124, 132, 137 – one identification each.

Low and high risk types of HPV sequences were identified in 50 of 855 TCGA breast cancers. E5, E6, and E7 oncogenes were identified.

HPV-positive HeLa laboratory cells (6). This contamination does not appear to affect the TCGA breast cancer specimens that we have studied and report in this current article (Cantalupo, 2015, personal communication). However, minor contamination cannot be excluded.

When considered together, the HPV nucleotides that have been identified in the 50 breast cancer specimens range from position 109 to 6589 on the HPV 18 complete genome, which has approximately 7,800 nt.

The majority of breast tumors had extremely low levels of viral expression with one or two short reads identified resulting in 50–100 nt long unique HPV fragments. Only four of the tumors showed longer continuous sequences of 300 nt or more. All identified HPV sequences had high and specific homology to reference HPV genomes. Short reads were randomly spread over thousands of nucleotides of the viral genome with large gaps between them due to the low viral load. An example of sequence reads can be seen in **Figure 1**.

The outcomes for HPV type 18 were identical to those of the Larsson lab.

HPV E7 gene sequences were identified in each of the high-risk HPV groups of breast cancers. HPV E7 oncoproteins are involved in oncogenesis. HPV E6 and E7 protein gene sequences

by NGS in breast cancer are shown in **Figure 1**. The homology between the identified viral and reference sequences was greater than 95%. These HPV RNA sequences indicate that this virus has the capacity to express HPV E6 and E7 proteins, which interact with p53 and Rb to cause malignant transformation.

Associations between HPV and APOBEC could not be detected in the TCGA cohort with respect to either mutations or high/low APOBEC expression levels ($p > 0.05$).

Identification by PCR of HPV Gene Sequences in Australian Benign and Subsequent Breast Cancers

As shown in **Table 2** and Table S2 in Supplementary Material, fifteen (65%) of the HPV-positive specimens were of the same type in both the benign and subsequent breast cancer (p = 0.001).

Consistent variations in the HPV sequences of both the benign and subsequent IDC breast cancers in the same patients is suggestive that the same HPV infection is involved in both sets of breast tissues. HPV 18 sequences were identified in five of seven ductal carcinoma in situ (DCIS) breast cancer specimens. HPV 18 was identified in three of seven benign and subsequent DCIS

```
1. 137 - 186      HPV E6 gene sequences

Cancer     1      CTACAAGCTACCTGATCTGTGCACGGAACTGAACACTTCACTGCAAGACA    50
                  |||||||||||||||||||||||||||||||||||||||||||||||||||
Reference  137    CTACAAGCTACCTGATCTGTGCACGGAACTGAACACTTCACTGCAAGACA    186


2. 461 - 510      HPV E6 gene sequences

Cancer     51     ACACCTTAATGAAAAACGACGATTTCACAACATAGCTGGGCACTATAGAG    100
                  ||||||||||||||||||||||||||||     ||||||||||||||||||
Reference  461    ACACCTTAATGAAAAACGACGATTCCACAAAATAGCTGGGCACTATAGAG    510


3. 682 to 731     HPV E7 gene sequences

Cancer     101    AAGCGACTCAGAGGAAGAAAACGATGAAATAGATGGAGTTAATCATCAAC    150
                  ||||||||||||||||||||||||||||||||||||||||||||||||||
Reference  682    AAGCGACTCAGAGGAAGAAAACGATGAAATAGATGGAGTTAATCATCAAC    731


4. 815 to 867     HPV E7 gene sequences

Cancer     151    GTAGAAAGCTCAGCAGACGACCTTCGAGCATTCCAGCAGCTGTTTCTGAAAAC    203
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Reference  815    GTAGAAAGCTCAGCAGACGACCTTCGAGCATTCCAGCAGCTGTTTCTGAAAAC    867


5. 2280 to 2329   HPV E1 gene sequences

Cancer     201     AACAAATAGAGTTTATAACATTTTTAGGAGCCTTAAAATCATTTTTAAAA    250
                   ||||||||||||||||||||||||||||||||||||||||||||||||||
Reference  2280    AACAAATAGAGTTTATAACATTTTTAGGAGCCTTAAAATCATTTTTAAAA    2329


6. 2375 to 2425   HPV E1 gene sequences

Cancer     250    ACAGGAAAATCATATTTTGGAATGAGTTTTATACACTTTATACAAGGAGCA    300
                  |||||||||||||||||||||||||||||||||||||||||||||||||||
Reference  2375    ACAGGAAAATCATATTTTGGAATGAGTTTTATACACTTTATACAAGGAGCA    2425
```

FIGURE 1 | HPV type 18 E6 and E7 RNA gene sequence identified in an invasive breast cancer specimen by Next Generation Sequencing. Invasive breast cancer specimen 36 TCGAAR-A24P-01A-11R-A169-07 (see Table S1 in Supplementary Material) sequence compared to reference HPV 18 KC470228. E6 gene sequences reference 105-581, E7 590-907, E1 914-2887. The two point mutations at reference 484 and 490 have been observed in reference HPV 18 KC470209.1.

specimens. The PCR products that were detected and sequenced are from the L1 region of the HPV genome.

HPV type 18 was the most common type identified in breast cancer specimens (55% of 40 IDC and DCIS breast cancer specimens) followed by HPV 16 (13%) and uncommonly HPV high-risk types 45, 58, and 73. Low-risk HPV types 3, 78, and 124 plus three different new HPV types were also identified. There were mixed HPV types in both prior benign and later breast cancer specimens. HPV viruses were not identified in 7 of 40 (17.5%) of sets of both benign and subsequent breast cancer specimens. Sequences from one benign and three invasive breast cancer specimens contained novel HPV sequences whose identity was the same as that reported from a Sydney renal transplant patient (AF019978).

The details of these findings are shown in Table S2 in Supplementary Material.

HPV type 18 DNA sequences in benign and subsequent breast cancer in a typical patient are shown in **Figure 2**. HPV sequences in a breast DCIS specimen is shown in **Figure 3**. As demonstrated in **Figure 2**, there are identical sequence variations of a PCR product Gp5+ to Gp6+ in both the benign and later breast cancer. These variations differ from the reference sequences based on HeLa cervical cell cultures and suggest that HPV infection of benign breast tissues has later become the same HPV virus-positive invasive breast cancer. In other HPV 18-positive sets of benign and later breast cancers, the sequence patterns were identical in the benign breast, subsequent breast cancer, and reference HeLa sequences.

An HPV 18-positive mixed DCIS and invasive ductal carcinoma of the breast identified by *in situ* PCR is shown in **Figure 3**. This image is typical of other images of invasive and non-invasive breast cancers identified by *in situ* PCR.

The identification of different, new, and novel HPV types in both the benign and breast cancer specimens indicates that false positive outcomes due to contamination during PCR analyses are unlikely.

The difference in prevalence of HPV sequences between normal breast and benign breast is significant ($p = 0.001$) and between normal breast and breast cancer is significant ($p = 0.001$) (**Table 3**).

Confirmation of Data Based on PCR

DNA extracts from six HPV-positive benign and breast cancer (extracted and analyzed at the University of New South Wales laboratories) patients were independently analyzed at the Antonsson laboratories in Brisbane, QLD, Australia. HPV types 16 and 18 were identified in 2 of 6 specimens. The identification of different HPV types indicates that contamination is unlikely.

Indicators of HPV Biological Activity and Oncogenic Influences
Identification of RNA by Next-Generation Sequencing

The identification of DNA viruses in cancer tissues by NGS involves recognition of RNAs. DNA viruses must be biologically active to transcribe into RNAs. Therefore, the outcomes of NGS in this investigation provide evidence that the high-risk HPVs identified are biologically active. In addition, the identification of HPV E6 and E7 genes in these high-risk HPV sequences is an indication that they are capable of expressing HPV E6 and E7 oncoproteins.

Immunohistochemistry for HPV E7

The data are shown in **Table 4** and in detail in Table S2 in Supplementary Material. The prevalence of HPV E7-positive protein expression is similar in both HPV-positive and HPV-negative benign breast specimens. The prevalence of HPV E7-positive protein expression is significantly higher in HPV DNA sequence-positive as compared to HPV sequence-negative breast cancer specimens (Pearson chi-square $p = 0.007$).

In nine sets of specimens, there was strong HPV E7 expression in the benign breast specimens but little or no HPV E7 staining in the subsequent breast cancer that developed later in the same patient. This is illustrated in Patient A in **Figure 4** in which there is high HPV E7 protein expression in the cell nuclei of the benign breast specimen but no expression in the nuclei of the subsequent breast cancer specimens. This pattern was not observed in four sets of specimens, in which the HPV E7 protein expression was higher in the cancer than the prior benign breast cell nuclei. This different pattern of HPV E7 protein expression

TABLE 2 | High risk HPV sequences in benign breast and subsequent invasive ductal carcinoma (IDC) breast cancer specimens in the same patients.

HPV sequences	Same HPV type: benign/cancer	Nil HPV benign HPV- positive cancer	HPV-positive benign Nil HPV cancer
HPV 16		1	1
HPV 16/18	2		
HPV 18	13	3	3
HPV 58		1	1
Nil HPV	10		

Fifteen (65%) of the HPV-positive specimens were of the same type in both the benign and subsequent breast cancer (p = 0.001).

Patient 21

Benign ACCTGGGCAATATGATGCTACCAAATTTAAGCAGTATAGCTAACACGCCGTACCTGGGCAATATGAT

Cancer ACCTGGGCAATATGATGCTACCAAATTTAAGCAGTATAGCTAACACGCCGTGCCTGGGCAATATGAT

Ref: ACCTGGGCAATATGATGCTACCAAATTTAAGCAGTATAGCAGACATGTTGTACCTGGGCAATATGAT

FIGURE 2 | HPV type 18 DNA sequences in benign breast and subsequent breast cancer. There are identical sequence variations of a PCR product Gp5+ to Gp6+ in both the benign and later breast cancer in patient 21. (The sequence has been abbreviated to include the differences). This indicates it is the same virus and differs from the reference sequences based on HeLa cervical cell culture.
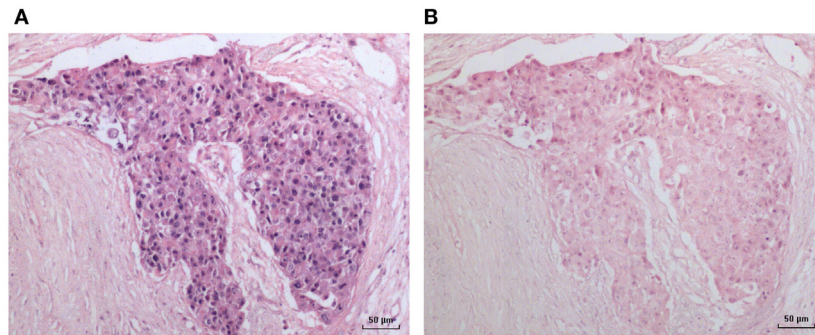
FIGURE 3 | HPV by *in situ* PCR in a breast ductal carcinoma *in situ* specimen. (A) HPV; **(B)** Negative (no primers) control.

TABLE 3 | Identification by PCR of high risk for cancer HPVs in normal breast (source cosmetic surgery), benign breast and breast cancer (same patients with prior benign breast and subsequent breast cancer).

Tissue type	High risk HPV sequences
Normal breast *n* = 21	6 (29%)
Benign breast *n* = 40	22 (55%)
Breast cancer *n* = 40	27 (66%)

The difference in prevalence of HPV sequences between normal breast and benign breast is significant (p = 0.001) and between normal breast and breast cancer is significant (p = 0.001).

TABLE 4 | HPV E7 protein expression in high risk HPV-positive and -negative benign breast and invasive ductal carcinoma (IDC) and ductal carcinoma *in situ* (DCIS) breast cancer.

HPV DNA sequences (PCR)	HPV E7 positive (immunohistochemistry)	HPV E7 negative (immunohistochemistry)
Benign breast		
HPV positive *n* = 20	13 (65%)	7 (35%)
HPV negative *n* = 11	8 (73%)	3 (27%)
IDC and dcis breast cancer		
HPV positive *n* = 21	16 (76%)	5 (24%)
HPV negative *n* = 9	2 (22%)	7 (78%)

The prevalence of HPV E7-positive protein expression is similar in both HPV-positive and HPV-negative benign breast specimens.
The prevalence of HPV E7-positive protein expression is significantly higher in HPV DNA sequence positive as compared to HPV-negative breast cancer specimens (Pearson chi-square p = 0.007).

is shown in Patient B in **Figure 4**. The implication is that the oncogenic influence of HPVs may be early in step by step onco-genesis in some patients. This will be considered in detail in the Section "Discussion." The positive control for HPV E7 protein expression – HPV type 18 cervical intraepithelial neoplasia (CIN 1), is shown in **Figure 5**.

There were no correlations between p16, p53, ER, PR, and HER2 protein expression in benign breast biopsy tissues and subsequent HPV-positive breast cancer (Table S2 in Supplementary Material). There were no correlations between the identification of HPVs and the presence of proliferative breast epithelial cells in benign breast biopsy tissues.

p16 proteins were highly expressed (2 or 3 on a scale of 0–3 staining intensity) in 12 benign and 11 invasive breast cancers in which high-risk HPVs had been identified (Table S2 in Supplementary Material). However, similarly high p16 expression was present in four benign and five invasive breast cancers in which HPV sequences were not identified.

Estrogen receptor protein expression
The expression of ER was very high in normal, benign, and subsequent breast cancer (Table S2 in Supplementary Material) whether or not HPV was identified. Therefore, HPV influences cannot be implied from these data.

HER2 protein expression
There was no correlation between the identification of HPV sequences and HER2 protein expression in either benign or later breast cancer.

Left or Right Breast Cancer Following Prior Benign Breast Conditions

In the same patients, 49% of the breast cancers developed in the same side as the benign breast condition and 51% developed in the opposite side breast. There was no association between HPV-positive breast cancers and whether the side of breast was the same as the prior HPV-positive benign condition. These observations suggest that if HPVs are oncogenic in breast cancer, the prior HPV infection and oncogenic influences may affect both breasts but only one develops breast cancer.

DISCUSSION

These investigations indicate (i) data based on the TCGA by Next-Generation Sequencing (NGS) confirms the identification of high-risk HPVs in approximately 2% of breast cancers, (ii) that high-risk HPV sequences of the same type identified by PCR are present in 37% of benign non-malignant breast and subsequent non-invasive and invasive breast cancers in the same Australian

Patient A. Benign breast at age 33 and subsequent breast cancer at age 44 years.
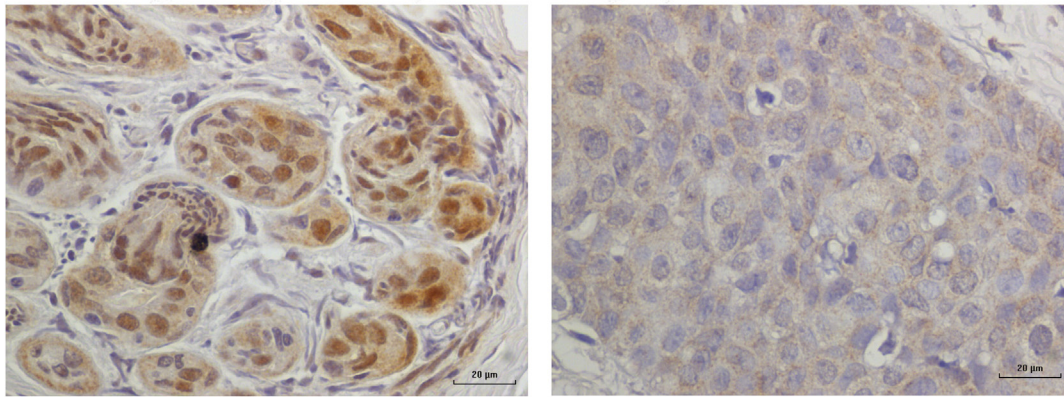
Patient B. Benign breast at age 53 and subsequent breast cancer at age 62 years.
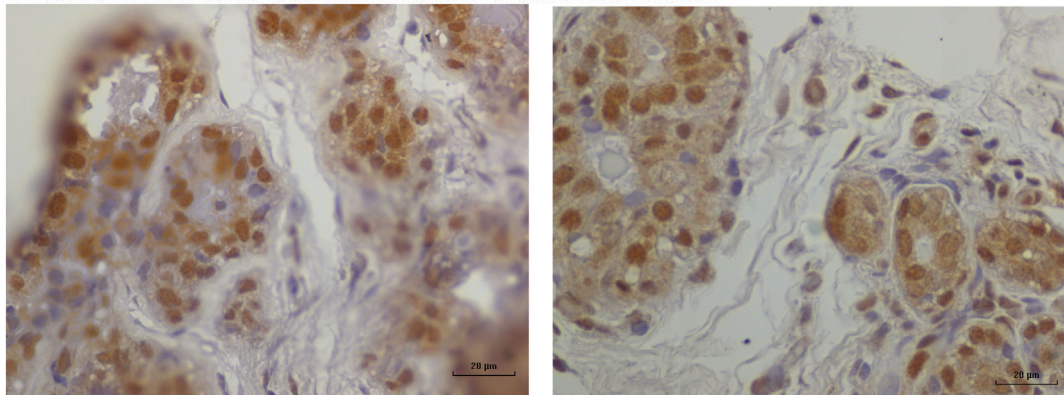
FIGURE 4 | HPV E7 protein expression in HPV 18-positive benign and subsequent HPV 18-positive breast cancer demonstrated by immunohistochemistry. Patient A. Benign breast specimen from a patient at age 33 years and subsequent invasive breast cancer in the same patient aged 44 years. HPV E7 protein brown staining is positive, blue staining is negative. There is strong HPV E7 protein expression in the nuclei of benign specimen. There is no HPV E7 protein expression in the subsequent invasive breast cancer specimen from the same patient. These observations are compatible with HPV having a role early in breast oncogenesis. Patient B. Benign breast specimen from a patient at age 53 and subsequent breast cancer in the same patient aged 62 years. There is strong HPV E7 protein staining in both the benign and subsequent breast cancer.
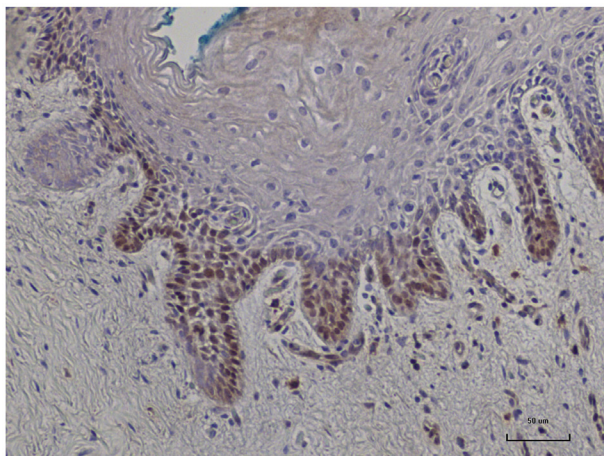

FIGURE 5 | HPV E7 protein expression identified in HPV type 18 positive cervical intraepithelial neoplasia (CIN 1). This is a positive control for Cervimax antibodies used in the immunohistochemistry analyses.

patients, and (iii) high-risk HPVs in breast cancer are biologically active as indicated by HPV E7 protein expression and the transcription of HPV DNA to RNA. In addition, the observation that HPV E7 protein may be highly expressed in the nuclei of benign breast cells but not expressed in the nuclei of subsequent breast cancer cells suggests HPV biological activity in breast cancer may be an early phenomena. This differs from cervical cancer, which usually requires the continuing presence of active HPVs for oncogenic progression.

In this study, HPV type 18 was the most frequently identified type. In each of the previous studies of HPV in Australian breast and prostate cancers, HPV type 18 has been the most common type (8, 9, 32). Types of high-risk HPVs vary in different populations. For example, HPV type 58 is common in Chinese breast cancers and HPV types 33 and 35 are the most common in Syrian breast cancers (33). HPV type 18 was the most common type identified by the Larsson Lab in the TCGA breast cancer data bank, which contains predominently US breast cancers. HPV type 16 was the most common type recently identified in breast cancers in Egyptian patients (34).

Validity of the Findings

The use of a range of techniques, namely, NGS, *in situ* PCR, nested PCR, real-time PCR, and immunohistochemistry, substantially adds to the validity to the outcomes. However, the identification of HPVs by NGS is much lower than their identification by PCR techniques. The high number of short "reads" by NGS can be regarded as an indication of possible false positives; however, the very high specificity and homology to HPV reference sequences suggests the outcomes are valid. The findings based on NGS confirm the extremely low HPV viral load in breast cancer.

The finding by both NGS and PCR of a variety of HPV types, suggests these HPV sequences were derived from patient samples and not contaminants. With respect to outcomes, based on NGS, there does not appear to be contamination (Cantalupo, 2015, personal communication). However, as previously indicated, minor contamination cannot be excluded. The identical identification by NGS of HPV 18 in the same TCGA breast cancers by the Larsson lab and in this current study adds to the validity of the NGS results. The outcomes of the investigations based on PCR are in agreement with the recent results of Ohba et al (7) who identified high-risk HPVs in 20 to 100% of breast cancers in different ethnic groups in Singapore. They used novel DNA chip technology for the identification of HPVs.

The HPV E7 gene sequences identified by NGS add validity to the outcomes based on immunohistochemistry.

Identification of High-Risk HPVs in Benign Non-Malignant Breast and Subsequent Breast Cancer

The much higher prevalence of HPV sequences in both the benign non-malignant breast and later breast cancer as compared with their prevalence in normal breast tissues from cosmetic surgery is striking. It is particularly noteworthy that HPV type 18 sequences with identical variations from the HeLa reference sequence were identified in benign breast tissues and 1–11 years later in breast cancer in the same patients. This suggests that it is probably the same virus in both the benign and subsequent breast cancer that developed later in the same patient. This finding adds to the Hill (35) criteria for causation.

HPV type 18 is the most common HPV type in breast cancer among Australian women (8, 9). HPV type 18 is also tropic to glandular cervical epithelial cells (36). The high (55%) proportion of HPV type 18 identified by PCR is very similar to the proportion of HPV type 18 identified in adenocarcinoma of the cervix (36). It is therefore of considerable interest that HPV 18 was by far the most common HPV type identified by NGS in the TCGA breast cancers.

Conflicting Data

It is difficult to reconcile the data based on TCGA and analyzed by NGS, with the data based on Australian breast cancer specimens and analyzed by PCR. The TCGA data indicates a 2% prevalence of high-risk HPV-positive breast cancers compared to 60% of Australian breast cancers in this current study. Other studies

based on PCR, such as the Ohba et al. study of breast cancer in Singapore residents (7), and prior studies of Australian breast cancers, have identified high-risk HPVs in from 25 to 100% of breast cancers (8, 9). The most likely explanation is the use of different viral identification techniques, namely, NGS as compared to PCR. An additional explanation is that only about 40% of the 60% of Australian breast cancers that were positive for high-risk HPV sequences had observable HPV protein expression. As the TCGA data is based on protein expression, this may explain part of those differences.

However, this does not explain the two- to six-fold higher prevalence of other HPV associated cancers, such as cervical cancer, in immunocompromised patients, as compared to immunocompromised patients with breast cancer (10). There are several possible explanations. One, HPVs may be present in breast cancers but are mostly not oncogenic, and are therefore unaffected by immunosuppression. Two, the oncogenic mechanisms of HPVs in breast cancer may differ from those in cervical cancer and may not be affected by immunosuppression. In our view, the second explanation is the most likely.

The mechanisms of HPV oncogenesis in cervical cancer are well established and include the influences of HPV E6 and E7 oncoproteins, which work in concert to disrupt cell-cycle regulation, inhibit apoptosis, and stimulate cell cycle progression by binding and inhibiting the p53 and p110RB tumor suppressor genes (28). In most instances, continued presence of HPV is required for cervical oncogenesis. In addition, HPV E5 and E6 act early in transformation (before integration) and are known to disrupt cytokeratin causing perinuclear cytoplasmic clearing and nuclear enlargement, which leads to the appearance of koilocytes (cells with halos surrounding the nuclei) (28).

There is experimental evidence that indicates there are different HPV oncogenic mechanisms in breast cancer as compared to cervical cancer. Much of this evidence has been developed by Ohba et al. in Singapore (7). They have demonstrated (i) that breast cancer cells with high ER protein are significantly correlated with high HPV infections (ii) that HPVs appear to activate APOBEC3B enzymes leading to genomic instability – abnormal APOBEC3B protein expression has been associated with increased breast cancer risk (37, 38). These oncogenic influences occur in the early stages of multi-step, multi-factorial breast carcinogenesis.

In this current study, we have observed there may be higher levels of HPV E7 protein expression in the nuclei of benign breast glandular cells as compared to HPV E7 expression in the nuclei of subsequent breast cancer cells in the same patients. This observation supports the hypothesis that HPVs may be involved in the early stages of breast oncogenesis. This may be the HPV "hit and run" phenomenon previously described by others (39).

Overall, this evidence offers a plausible explanation for the lack of influence of immunosuppression in HPV associated breast cancer.

Indicators of HPV Biological Activity and Oncogenic Influences

The outcomes of NGS in TCGA data in this investigation provide suggestive evidence that the high-risk HPVs identified

are biologically active. This is demonstrated by the recognition of HPV RNAs by NGS. This is because HPVs are DNA viruses that generate RNA transcripts for the expression of oncogenic proteins.

There was a significant correlation ($r = 0.603$, $p = 0.004$) between positive and negative HPV sequence identification and HPV E7 protein expression in normal (cosmetic surgery) breast tissues. The implication is that HPV E7 protein is a likely indicator of HPV biological activity in breast tissues. However, 10 years later, none of the women who had HPV-positive normal breast tissues had developed breast cancer. Therefore, the identification of HPV sequences and HPV E7 protein expression in benign breast tissues does not necessarily lead to subsequent breast cancer.

The inhibited expression of p53 protein in benign non-malignant breast and breast cancer observed in this study parallels the prior observation by Hennig et al (3). The inhibition of expression of the cell death related p53 protein by HPV E6 protein is a well-documented mechanism for HPV oncogenesis (28).

The role of p16 in breast cancer is not clear. In this current study, there were no significant differences in p16 protein expression as assessed by IHC between normal, benign, and malignant breast tissues. In addition, there were no significant correlations between high p16 protein expression and triple negative breast cancer as has been observed in other studies (40, 41). High p16 protein expression has been associated with HPV infections in breast, cervical, and other cancers (40, 41). However, in these and other studies of p16 expression in breast cancer, normal breast controls were not used, and therefore is not possible draw any conclusions from our current data.

HPV Transmission

The mode of transmission of HPV to the breast is not known. However, it is possible that HPVs are transmitted to the genital tract during sexual activities and later transmitted by white blood cells throughout the body, including the breasts. This hypothesis is based on the young age of women with HPV-positive breast cancer. These young women are sexually active and have a high incidence of HPV cervical infections (42). In addition, there have been repeated identification of high-risk HPVs in white blood cells (43).

Breast Cancer is Hormone Dependent

The presence of a hormone influenced virus in breast cancers is of special interest. The levels of ER expression in normal, benign, and breast cancer specimens in this study were very high. High ER expression in association with HPV-related breast cancer was also observed by Ohba (7). Estrogens synergize with high-risk HPV E6 and E7 oncogenes to cause human cervical cancer (44). In addition, the regulatory region of HPV 16 genomic DNA contains sequences that are responsive to glucocorticoid hormones. The HPV18 regulatory region contains one functional GRE sequence that interacts with the glucocorticoid receptor and confers hormonal activation to HPV18 promoter (45).

Benign Breast Disease as Precancerous

These data indicate the marked differences between normal breast and benign breast tissues. These differences include the significantly higher prevalence of HPV sequences and proliferation of epithelial cells in benign as compared to normal breast tissues. In addition, there is a significant correlation between benign and later breast cancer in the same patient for the identification of HPV gene sequences and abnormal biomarker expression, including HPV E7 protein expression. While it is well documented that "benign" breast hyperplasia (glandular cell proliferation) leads to a fourfold increased risk of breast cancer, the observations in this current study, suggest that non-proliferative benign breast disease may also be precancerous (46, 47).

Left or Right Breast Cancer Following Prior Benign Breast Conditions

The observation that there was no association between the side of HPV-positive benign breast conditions and the side of later HPV-positive breast cancer is of considerable interest. It may be that prior HPV infection and oncogenic influences may affect both breasts but only one develops breast cancer. However, there is no evidence from this current study to support this notion. The development of breast cancer in both breasts either at the same or at different times occurs in women with and without a genetic predisposition to develop breast cancer (48). The risk of bilateral breast cancer at different times is higher among young women (48). This is compatible with an HPV infectious cause of the cancer.

CONCLUSION

In this current study, there were four observations of particular interest: (i) confirmation by both NGS and a range of PCR techniques, of the presence of high-risk HPV gene sequences in breast cancers, (ii) there is a correlation between high-risk HPV gene sequences in benign breast biopsy specimens and subsequent HPV-positive breast cancer in the same patient, which is an indication of prior HPV infection before the development of HPV-positive breast cancer, (iii) that HPVs in breast cancer are likely to be biologically active (as shown by transcription of HPV DNA to RNA plus the expression of HPV E7 proteins), and (iv) that HPV oncogenic influences may occur early in the step by step development of breast cancer.

ETHICS STATEMENT

This project has formal ethics approval by the University of New South Wales Human Research Ethics Committee – number HREC HC11421. Participants gave written informed consent to participate in this study.

Ethics approvals for the follow up of women who donated normal breast tissues was given by the New South Wales Population and Health Services Research Ethics Committee – number 2009/12/203. This Ethics committee waived the need for consent. The reasons for waiving consent were (i) the specimens were

archival having been collected in 1999, 2000 and 2001, (ii) all specimens were "de-identified" to the research group and (iii) retrospective approaches to donors, all of whom had cosmetic surgery, may have caused unnecessary anxiety.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://journal.frontiersin.org/article/10.3389/fonc.2015.00277

REFERENCES

1. Li N, Bi X, Zhang Y, Zhao P, Zheng T, Dai M. Human papillomavirus infection and sporadic breast carcinoma risk: a meta-analysis. *Breast Cancer Res Treat* (2011) **126**:515–20. doi:10.1007/s10549-010-1128-0

2. Simões PW, Medeiros LR, Simões Pires PD, Edelweiss MI, Rosa DD, Silva FR, et al. Prevalence of human papillomavirus in breast cancer: a systematic review. *Int J Gynecol Cancer* (2012) **22**:343–7. doi:10.1097/IGC.0b013e31823c712e

3. Hennig EM, Suo Z, Thoresen S, Holm R, Kvinnsland S, Nesland JM. Human papillomavirus 16 in breast cancer of women treated for high grade cervical intraepithelial neoplasia (CIN III). *Breast Cancer Res Treat* (1999) **53**:121–35. doi:10.1023/A:1006162609420

4. Widschwendter A, Brunhuber T, Wiedemair A, Mueller-Holzner E, Marth C. Detection of human papillomavirus DNA in breast cancer of patients with cervical cancer history. *J Clin Virol* (2004) **31**:292–7. doi:10.1016/j.jcv.2004.06.009

5. Joshi D, Buehring GC. Are viruses associated with human breast cancer? Scrutinizing the molecular evidence. *Breast Cancer Res Treat* (2012) **135**:1–15. doi:10.1007/s10549-011-1921-4

6. Cantalupo PG, Katz JP, Pipas JM. HeLa nucleic acid contamination in the cancer genome atlas leads to the misidentification of human papillomavirus 18. *J Virol* (2015) **89**:4051–7. doi:10.1128/JVI.03365-14

7. Ohba K, Ichiyama K, Yajima M, Gemma N, Nikaido M, Wu Q, et al. In vivo and in vitro studies suggest a possible involvement of HPV infection in the early stage of breast carcinogenesis via APOBEC3B induction. *PLoS One* (2014) **9**:e97787. doi:10.1371/journal.pone.0097787

8. Heng B, Glenn WK, Ye Y, Tran B, Delprado W, Lutze-Mann L, et al. Human papilloma virus is associated with breast cancer. *Br J Cancer* (2009) **101**:1345–50. doi:10.1038/sj.bjc.6605282

9. Antonsson A, Spurr TP, Chen AC, Francis GD, McMillan NA, Saunders NA, et al. High prevalence of human papillomaviruses in fresh frozen breast cancer samples. *J Med Virol* (2011) **83**:2157–63. doi:10.1002/jmv.22223

10. Grulich AE, Vajdic CM. The epidemiology of cancers in human immunodeficiency virus infection and after organ transplantation. *Semin Oncol* (2015) **42**:247–57. doi:10.1053/j.seminoncol.2014.12.029

11. Khan NA, Castillo A, Koriyama C, Kijima Y, Umekita Y, Ohi Y. Human papillomavirus detected in female breast carcinomas in Japan. *Br J Cancer* (2008) **99**:408–14. doi:10.1038/sj.bjc.6604502

12. Herrera-Romano L, Fernández-Tamayo N, Gómez-Conde E, Reyes-Cardoso JM, Ortiz-Gutierrez F, Ceballos G, et al. Absence of human papillomavirus sequences in epithelial breast cancer in a Mexican female population. *Med Oncol* (2012) **29**:1515–7. doi:10.1007/s12032-011-0059-x

13. Dimri G, Band H, Band V. Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res* (2005) **7**:171–9. doi:10.1186/bcr973

14. Lawson JS, Glenn WK, Whitaker NJ. Breast cancer, human papilloma virus and sexual activities. *Br J Cancer* (2008) **98**:510–1. doi:10.1038/sj.bjc.6604104

15. Akil N, Kassab A, Yasmeen A, Darnel AD, Bismar TA, Al Moustafa AE. Human breast cancer and sexual activities. *Br J Cancer* (2008) **98**:508–9. doi:10.1038/sj.bjc.6604103

16. Antonsson A, Green AC, Mallitt KA, O'Rourke PK, Pandeya N, Pawlita M, et al. Prevalence and stability of antibodies to 37 human papillomavirus types – a population-based longitudinal study. *Virology* (2010) **407**:26–32. doi:10.1016/j.virol.2010.07.046

17. Lawson JS, Glenn WK, Heng B, Ye Y, Tran B, Lutze-Mann L, et al. Koilocytes indicate a role for human papilloma virus in breast cancer. *Br J Cancer* (2009) **101**:1351–6. doi:10.1038/sj.bjc.6605328

18. Salyakina D, Sinoremas NF. Viral expression associated with gastrointestinal adenocarcinomas in TCGA high-throughput sequencing data. *Hum Genomics* (2013) **7**:23. doi:10.1186/1479-7364-7-23

19. Steinau M, Patel SS, Unger ER. Efficient DNA extraction for HPV genotyping in formalinfixed, paraffin-embedded tissues. *J Mol Diagn* (2011) **13**:377–81. doi:10.1016/j.jmoldx.2011.03.007

20. Antonsson A, Nancarrow DJ, Brown IS, Green AC, Drew PA, Watson DI, et al. High-risk human papillomavirus in esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev* (2010) **19**:2080–7. doi:10.1158/1055-9965.EPI-10-0033

21. Faoro V, Barbazza R, Bonin S, Brunetti D, Sulfaro S, Stanta G. Detection of HPV E7 oncoviral protein in cervical lesions by a new antibody. *Appl Immunohistochem Mol Morphol* (2013) **21**:341–50. doi:10.1097/PAI.0b013e318269bf5d

22. Romagosa C, Simonetti S, Lopez-Vicente L, Mazo A, Lleonart ME, Castellvi J, et al. p16Ink4a overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene* (2011) **30**:2087–97. doi:10.1038/onc.2010.614

23. Lewis JS Jr, Chernock RD, Ma XJ, Flanagan JJ, Luo Y, Gao G, et al. Partial p16 staining in oropharyngeal squamous cell carcinoma: extent and pattern correlate with human papillomavirus RNA status. *Mod Pathol* (2012) **25**:1212–20. doi:10.1038/modpathol.2012.79

24. Subhawong AP, Subhawong T, Nassar H, Kouprina N, Begum S, Vang R, et al. Most basal-like breast carcinomas demonstrate the same Rb-/p16+ immunophenotype as the HPV-related poorly differentiated squamous cell carcinomas which they resemble morphologically. *Am J Surg Pathol* (2009) **233**:163–75. doi:10.1097/PAS.0b013e31817f9790

25. Kerlikowske K, Molinaro AM, Gauthier ML, Berman HK, Waldman F, Bennington J, et al. Biomarker expression and risk of subsequent tumors after initial ductal carcinoma in situ diagnosis. *J Natl Cancer Inst* (2010) **102**:627–37. doi:10.1093/jnci/djq101

26. Thomas J, Primeaux T. Is p16 immunohistochemistry a more cost-effective method for identification of human papilloma virus-associated head and neck

squamous cell carcinoma? *Ann Diagn Pathol* (2012) **16**:91–9. doi:10.1016/j.anndiagpath.2011.09.002

27. Ohtani N, Brennan P, Gaubatz S, Sanij E, Hertzog P, Wolvetang E, et al. Epstein–Barr virus LMP1 blocks p16INK4a-RB pathway by promoting nuclear export of E2F4/5. *J Cell Biol* (2003) **162**:173–83. doi:10.1083/jcb.200302085

28. Zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* (2002) **2**:342–50. doi:10.1038/nrc798

29. Krishnamurthy S, Sneige N. Molecular and biologic markers of premalignant lesions of human breast. *Adv Anat Pathol* (2002) **9**:185–97. doi:10.1097/00125480-200205000-00002

30. Woods Ignatoski KM, Dziubinski ML, Ammerman C, Ethier SP. Cooperative interactions of HER-2 and HPV-16 oncoproteins in the malignant transformation of human mammary epithelial cells. *Neoplasia* (2005) **7**:788–98. doi:10.1593/neo.05106

31. Yasmeen A, Bismar TA, Dekhil H, Ricciardi R, Kassab A, Gambacorti-Passerini C, et al. ErbB-2 receptor cooperates with E6/E7 oncoproteins of HPV type 16 in breast tumorigenesis. *Cell Cycle* (2007) **6**:2939–43. doi:10.4161/cc.6.23.4949

32. Whitaker NJ, Glenn WK, Sahrudin A, Orde MM, Delprado W, Lawson JS. Human papillomavirus and Epstein–Barr virus in prostate cancer: koilocytes indicate potential oncogenic influences of human papillomavirus in prostate cancer. *Prostate* (2013) **73**:236–41. doi:10.1002/pros.22562

33. Al Moustafa AE, Ghabreau L, Akil N, Rastam S, Alachkar A, Yasmeen A. High risk HPVs and human carcinomas in the Syrian population. *Front Oncol* (2014) **4**:68. doi:10.3389/front.2014.00068

34. El-Shinawi M, Mohamed HT, Abdel-Fattah HH, Ibrahim SAA, El-Halawany MS, Nouh MA, et al. Inflammatory and non-inflammatory breast cancer: a potential role for detection of multiple viral DNAs in disease progression. *Ann Surg Oncol* (2015). doi:10.1245/s10434-015-4888-2

35. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* (1965) **58**:295–330.

36. Clifford G, Franceschi S. Members of the human papillomavirus type 18 family (alpha-7 species) share a common association with adenocarcinoma of the cervix. *Int J Cancer* (2008) **122**:1684–5.

37. Xuan D, Li G, Cai Q, Deming-Halverson S, Shrubsole MJ, Shu XO, et al. APOBEC3 deletion polymorphism is associated with breast cancer risk among women of European ancestry. *Carcinogenesis* (2013) **34**:2240–3. doi:10.1093/carcin/bgt185

38. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* (2013) **21**:366–70. doi:10.1038/nature11881

39. Niller HH, Wolf H, Minarovits J. Viral hit and run-oncogenesis: genetic and epigenetic scenarios. *Cancer Lett* (2011) **305**:200–17. doi:10.1016/j.canlet.2010.08.007

40. Peurala E, Koivunen P, Haapasaari KM, Bloigu R, Jukkola-Vuorinen A. The prognostic significance and value of cyclin D1, CDK4 and p16 in human breast cancer. *Breast Cancer Res* (2013) **15**:R5. doi:10.1186/bcr3376

41. Shan M, Zhang X, Liu X, Qin Y, Liu T, Liu Y, et al. p16 and p53 play distinct roles in different subtypes of breast cancer. *PLoS One* (2013) **8**(10):e76408. doi:10.1371/journal.pone.0076408

42. Garland SM, Brotherton JM, Condon JR, McIntyre PB, Stevens MP, Smith DW, et al. Human papillomavirus prevalence among indigenous and non-indigenous Australian women prior to a national HPV vaccination program. *BMC Med* (2011) **9**:104. doi:10.1186/1741-7015-9-104

43. Foresta C, Bertoldo A, Garolla A, Pizzol D, Mason S, Lenzi A, et al. Human papillomavirus proteins are found in peripheral blood and semen Cd20+ and Cd56+ cells during Hpv-16 semen infection. *BMC Infect Dis* (2013) **13**:593. doi:10.1186/1471-2334-13-593

44. Brake T, Lambert PF. Estrogen contributes to the onset, persistence and malignant progression of cervical cancer in a human papilloma virus-transgenic mouse model. *PNAS* (2005) **102**:2490–5. doi:10.1073/pnas.0409883102

45. Bromberg-White JL, Meyers C. Comparison of the basal and glucocorticoid-inducible activities of the upstream regulatory regions of HPV18 and HPV31 in multiple epithelial cell lines. *Virology* (2003) **306**:197–202. doi:10.1016/S0042-6822(02)00041-7

46. Hartmann LC, Radisky DC, Frost MH, Santen RJ, Vierkant RA, Benetti LL, et al. Understanding the premalignant potential of atypical hyperplasia through its natural history: a longitudinal cohort study. *Cancer Prev Res (Phila)* (2014) **7**:211–7. doi:10.1158/1940-6207.CAPR-13-0222

47. Dyrstad SW, Yan Y, Fowler AM, Colditz GA. Breast cancer risk associated with benign breast disease: systematic review and meta-analysis. *Breast Cancer Res Treat* (2015) **149**:569–75. doi:10.1007/s10549-014-3254-6

48. Reiner AS, John EM, Brooks JD, Lynch CF, Bernstein L, Mellemkjær L, et al. Risk of asynchronous contralateral breast cancer in noncarriers of BRCA1 and BRCA2 mutations with a family history of breast cancer: a report from the Women's Environmental Cancer and Radiation Epidemiology Study. *J Clin Oncol* (2013) **31**:433–9. doi:10.1200/JCO.2012.43.2013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.