Human Papillomavirus and the Development of Different Cancers

Ge Gao · David I. Smith

Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

Keywords

Cervical cancer · Integration site · Next-generation sequencing · Oropharyngeal squamous cell carcinoma

Abstract

Human papillomaviruses (HPV) are responsible for the development of almost all cervical cancers. HPV is also found in 85% of anal cancer and in 50% of penile, vulvar, and vaginal cancers, and they are increasingly found in a subset of head and neck cancers, i.e., oropharyngeal squamous cell carcinomas (OPSCC). The model for how HPV causes cancer is derived from several decades of study on cervical cancer, and it is just presumed that this model is not only completely valid for cervical cancer but for all other HPV-driven cancers as well. Next-generation sequencing (NGS) has now provided the necessary tools to characterize genomic alterations in cancer cells and can precisely determine the physical status of HPV in those cells as well. We discuss recent discoveries from different applications of NGS in both cervical cancer and OPSCCs, including whole-genome sequencing and mate-pair NGS. We also discuss what NGS studies have revealed about the different ways that HPV can be involved in cancer formation, specifically comparing cervical cancer and OPSCC.

© 2017 S. Karger AG, Basel

Human Papillomaviruses

Human papillomaviruses (HPV) are a large group of double-stranded DNA viruses whose infection usually causes benign epithelial lesions (warts). However, infection with certain subtypes of HPV, such as HPV16, HPV18, and HPV31, is associated with an increased risk of developing cervical cancer [Brescia et al., 1986; Liu et al., 2015] as well as several other anogenital cancers and now increasingly in one subset of head and neck cancers, oropharyngeal squamous cell carcinoma (OPSCC) [Jenson et al., 2001; Srodon et al., 2006; Hoots et al., 2009; Flaherty et al., 2014; Panwar et al., 2014].

HPV have an approximately 8-kb genome comprised of 6 early transcribed genes E1, E2, E4, E5, E6, and E7 encoding the nonstructural proteins, 2 late transcribed genes L1 and L2 encoding the structural viral capsid proteins, and a noncoding long control region. Once host cells are infected with HPV, a viral early promoter is activated and a polycistronic primary RNA containing all 6 early open reading frames is transcribed. The early proteins are essential for virus gene regulation, replication, and pathogenesis [Tommasino, 2014].

The exposure to HPV is very common. HPV is primarily transmitted via sexual contact, but it can also be transmitted by skin contact. In the cervix, HPV infect keratinocytes in the basal layer of the squamous epithelium and then replicate in the nucleus of infected cells in a differentiation-dependent manner [Hoffmann et al.,

2006]. There are over 120 HPV types that have been characterized, and they are designated as high risk or low risk depending on their ability to promote carcinogenesis. HPV that are only associated with benign lesions are known as low-risk HPV subtypes, while those that have been associated with an increased risk of developing cancer are classified as high-risk HPV subtypes. The key differences between the high- and low-risk HPV subtypes reside within 2 key early genes, E6 and E7. In high-risk HPV subtypes, E6 and E7 have been shown to function as oncoproteins. The high-risk HPV E6 protein can bind to the tumor suppressor p53 and stimulate its degradation by the ubiquitination-mediated pathway [Scheffner et al., 1990, 1993]. In contrast, E6 from a low-risk HPV subtype is less capable of inactivating p53. Similarly, the high-risk HPV E7 gene product acts by binding to the pRB tumor suppressor protein and promotes its degradation, while the low-risk HPV E7 gene product does so at a lower efficiency [Dyson et al., 1989; Giarrè et al., 2001]. Infection with a high-risk HPV subtype can thus simultaneously inactivate 2 key tumor suppressor pathways leading to further genomic instability which can eventually result in the development of cancer.

The E1 and E2 gene products are involved in viral DNA replication and the regulation of early transcription. The E2 gene also specifically functions as a transcriptional repressor of the polycistronic primary HPV early RNA [Tommasino, 2014]. HPV E5 also has some oncogenic properties and has been shown to cooperate with E6 and E7 to promote hyperproliferation of infected cells, facilitating their malignant progression [DiMaio and Mattoon, 2001].

HPV and Cancers

High-risk HPV subtypes are known to be associated with over 99% of cervical cancers, 85% of anal cancers, and 50% of penile, vulvar, and vaginal cancers. In the past several decades there has been an epidemic increase in OPSCCs that are HPV positive. Several decades ago, the frequency of HPV-positive OPSCCs was quite low, but now up to 85% of the cases in the United States and an even higher frequency in some European countries are HPV positive [Marur et al., 2010].

HPV and Cervical Cancer

Cervical cancer is the second most deadliest cancer in women and is responsible for about 270,000 deaths each

year [Kumar, 2016]. Over 99% of cervical cancers are caused by high-risk HPV infections. Sexually active women are repeatedly infected with HPV, but for the majority of women their immune system rapidly clears the virus. However, the subset of women whose immune system does not is the group that is at an increased risk of developing cancer. HPV infects the basal membrane of the cervical epithelium, and persistent HPV infection can ultimately lead to an increased episomal copy of HPV. However, HPV infection alone (in vitro) only immortalizes cells but does not transform them. Most invasive cervical cancers have HPV sequences integrated into the human genome; thus viral integration is believed to be the key event leading to carcinogenesis [Walboomers et al., 1999; zur Hausen, 2002]. In vitro studies have suggested that cells harboring integrated HPV possess a selective growth advantage compared to cells maintaining only episomal HPV genomes.

Most of our understanding on HPV's role in cancer development comes from several decades of study of different cervical cancer cell lines and a limited number of primary cancers. It was revealed that the site of integration in the HPV genome was frequently within the early E2 gene [Badaracco et al., 2002]. E2 functions as a transcriptional repressor of the oncogenes E6 and E7; thus the disruption of E2 leads to increased expression of these 2 critical oncoproteins. The increased expression of high-risk HPV E6 and E7 oncoproteins and the inactivation of 2 critical tumor suppressor pathways would lead to increased genome instability, and this would facilitate other genomic alterations which would ultimately lead to malignant progression. Viral integrations have been observed in 100% of HPV18-positive and over 80% of HPV16-positive cases in cervical cancer [Cullen et al., 1991; Pirami et al., 1997]. HPV integration is also found to be correlated with cervical intraepithelial neoplasia grades and overall disease progression [Vega-Peña et al., 2013].

While there initially appeared to be some specificity for where integrations occurred within the HPV genome, the sites of integrations in the human genome appeared much more random. Hence, the classic model for HPV-driven carcinogenesis was that long-term infection with HPV can eventually result from HPV integrating randomly somewhere into the human genome. This integration, since it would frequently inactivate the HPV E2 gene, would result in greatly increased expression of E6 and E7, which would cause generalized genomic instability, and the resulting subsequent changes eventually lead to the development of invasive cancer. Thus, the sites of HPV integration into the human genome seemed unimportant for the eventual cancer that develops.

In order to determine if there was any validity to this model, we decided to characterize a large number of HPV16 and HPV18 integrations in cervical cancer. The technology we utilized at the time (late 1990s) was a PCR-based technology called restriction site oligonucleotide PCR (RSO-PCR), in which primers complementary to different regions of the HPV genome are paired with primers designed around the recognition sequence of different restriction endonucleases (with random nucleotides at their ends). With a nested PCR approach and with different combinations of the primer sets, one can detect the specific amplified product if there are potential HPV integration events that have occurred in the human genome close enough to one of the restriction endonuclease sites chosen to make RSO primers with. Sanger sequencing is then performed to validate and localize the site of HPV integration both within the HPV and the human genome. We observed that 50% of the HPV16 integrations and 65% of the HPV18 integrations occurred within chromosomal regions known as common fragile sites (CFSs) [Ferber et al., 2003; Thorland et al., 2003]. CFSs are regions of profound genomic instability that are observed in all individuals [Glover et al., 1984]. They were originally observed as specific chromosomal regions which failed to compact prior to cell division in a statistically significant manner when cells were cultured in the presence of DNA replication inhibitors. These highly unstable regions also are frequently found to have alterations in many different cancers.

There are at least 89 CFS regions that have been described throughout the human genome. The 3 most frequently expressed CFSs in human lymphoblasts are FRA3B (3p14.2), FRA16D (16q23.2), and FRA6E (6q26), which are quite large regions of instability, as they span 4.25 [Becker et al., 2002], 3.0 [Krummel et al., 2002], and 2.0 Mb [Denison et al., 2003], respectively. All 3 regions are also hotspots for deletions and other alterations in a number of different solid tumors. Furthermore, each of these highly unstable regions completely spans genes that themselves are quite large. These 3 genes are *FHIT* (1.5 Mb at FRA3B), *WWOX* (1.3 Mb at FRA16D), and *PARK2* (1.1 Mb at FRA6E) and have been demonstrated to function as important tumor suppressors that play roles in the development of a number of different cancers [Mao et al., 1997; Sozzi et al., 1998; Imai et al., 2000; Kuroki et al., 2004; Watson et al., 2004]. Many of the other largest human genes (in terms of the size of the genomic region that encodes them) have also been found to be located within one of the highly unstable CFS regions including the first and second largest human genes *CNTNAP2* (2.3 Mb) and

DMD (2.1 Mb), respectively. Some of the other very interesting very large CFS genes include *LRP1B* (1.9 Mb), *RORA* (732 kb), *CSMD1* (2.1 Mb), *DAB1* (1.6 Mb), and *DLG2* (1.5 Mb). Each of these genes is found to be a target of alterations in different cancers but they have not yet been sufficiently functionally characterized to demonstrate whether they are also true tumor suppressor genes as well.

Our studies in cervical cancer showed that deletions and complex rearrangements frequently occurred in the cellular sequences targeted by the integrations that were clustered in FRA13C (13q22), FRA3B (3p14.2), and FRA17B (17q23) [Thorland et al., 2003]. Other groups also reported that chromosome bands that contain CFS regions are the preferred sites for HPV integrations in both cervix carcinoma-derived cell lines and clinical samples [Hidalgo et al., 2003; Dall et al., 2008; Matovina et al., 2009]. In addition, we also found a hotspot of HPV18 integration located within one of the CFSs at chromosomal band 8q24 near the *c-MYC* proto-oncogene [Ferber et al., 2003]. A number of other integrations were also found to occur within very large genes, and one of the large CFS genes identified was *LRP1B*. Deletions in *LRP1B* have been found in many different cancers suggesting that it could play a very important role in cancer development.

Different signaling pathways have been shown to be inactivated or activated by the overexpression of oncogenes including E6 and E7. In addition to the degradation of p53 and pRB, the E6 and E7 proteins have been found to activate PI3K/AKT, Wnt/b-catenine, as well as Notch signaling pathways, and all of these would cause increased genome instability, cell proliferation, and cell growth but with decreased apoptosis [Chen, 2015]. In addition, the CFS large genes frequently disrupted by the viral integration were also found to be involved in oncogenic signaling pathways. For example, *FHIT*, which is located in FRA3B and is one of the most frequently disrupted large CFS genes, regulates EMT by targeting the EGFR/Src/ERK/Slug signaling pathway and can also induce apoptosis via the death receptor signaling pathway [Deng et al., 2007; Joannes et al., 2014]. Decreased expression of *LRP1B* was also shown to promote cell migration via the RhoA/Cdc42 pathway and actin skeleton remodeling in renal cell cancer [Ni et al., 2013].

HPV and Oropharyngeal Squamous Cell Carcinoma

OPSCC is a subtype of head and neck cancer which is composed of cancers that arise from the base of the tongue, tonsils, the walls of the pharynx, and the soft palate [Licitra et al., 2002]. With the reduced incidence of

smoking in the United States over the past several decades, there has been a dramatic decrease in the incidence of most types of head and neck cancer with the notable exception of OPSCC which has actually been dramatically increasing. In the United States, from 1988 to 2004, the incidence of HPV-positive OPSCC has increased over 225%, while HPV-negative OPSCCs have dropped around 50% [Chaturvedi et al., 2011]. It was reported that there has been about a 5% increase in the annual percentage rate in the incidence of OPSCC in the United States in the recent years, and around 11,700 HPV-positive OPSCCs are diagnosed annually, particularly among middle-aged white men [Chaturvedi et al., 2011]. The dramatically increased incidence of HPV-associated OPSCCs has also been observed in some European countries, such as Sweden and the Netherlands [Näsman et al., 2009; Rietbergen et al., 2013]. The most likely reason for this dramatically increased incidence of HPV-infected OPSCC is due to changing sexual behaviors, which include a higher proportion of patients engaging in oral sex as well as having an increased number of sexual partners. The high-risk type HPV16 has been reported as the most prevalent subtype in OPSCC with a prevalence of 93.1%, followed by HPV33 at 1.4%, HPV18 at 1.3%, and HPV35 at 0.5%. The prevalence of all other genotypes combined is around 4.2% from recent studies from North America, Europe, and other countries [Kreimer et al., 2005].

OPSCCs are distinct from cervical cancers, as virtually all cervical tumors contain HPV sequences and HPV is presumably the only cause for cancer development. Most OPSCC patients also have drinking and smoking histories which also play important roles in carcinogenesis. There are at least 3 groups of individuals who develop OPSCCs. Individuals who (a) have a history of smoking and drinking but no HPV sequences present, (b) have a history of smoking and drinking and have HPV sequences in their tumors, and finally (c) have no history of smoking and drinking but whose tumors contain HPV sequences. There are also other risk factors for the development of OPSCC, including diets with lots of barbecued foods. To confound things even further, not all HPV-positive OPSCCs are the same, as some of these tumors contain HPV sequences but with very low expression of HPV-specific transcripts, considered as “latent” HPV infections. In contrast, others have very high expression of those transcripts and are considered as “active” HPV infections.

In general, HPV-positive OPSCC patients present with distinct features than those who are HPV negative, both biologically and clinically. HPV-positive tumors are

less differentiated and usually have a lower degree of aneuploidy than HPV-negative tumors [Mellin et al., 2003], and HPV-positive patients are found to have a better overall clinical outcome than those whose tumors are HPV negative [Gao et al., 2013; Deng et al., 2014]. Thus, most HPV-positive patients need a less intensive chemoradiotherapy regime administered, which is currently being clinically explored.

An important question is whether HPV plays a similar role in OPSCC as it does in cervical cancer. Will the risk factors of smoking and drinking augment HPV’s role in carcinogenesis, or vice versa? It has been shown that patients with HPV-positive tonsil cancers who have never been smokers have a better clinical outcome than those who have a smoking history. Other important questions include: what is the physical status of HPV in these tumors? Has HPV integrated into the genome in most OPSCCs, as it is found in cervical cancer, and is this associated with loss of episomal sequences of HPV? Also, is there any specificity for the sites of integration either within the HPV genome or within the human genome, similar to what is observed in cervical cancer when integration does occur?

Our laboratory first utilized RSO-PCR to attempt to identify any sites of HPV integration in a group of HPV-positive OPSCCs, but did not find any validated HPV integrations from 20 OPSCCs [Ukpo et al., 2007]. This technique, however, is critically dependent upon having suitable restriction endonuclease sites sufficiently adjacent to an HPV integration to amplify a unique PCR product, and it does not discover all HPV integration events in cervical cancer. There are other different techniques that have been utilized to either directly determine or infer the physical status of HPV in one specific tumor sample. For example, 2 groups have used real-time PCR to examine the ratio in the expression of E2/E6 to characterize the HPV physical status (under the assumption that all integrations would inactivate E2 gene expression), and they determined that the frequency of HPV integration in HPV-positive OPSCCs was 48 and 78%, respectively [Koskinen et al., 2003; Deng et al., 2013]. Another group performed *in situ* hybridization to evaluate the physical status of HPV in OPSCCs based on the detection of punctate nuclear signals, and they reported that 42% of OPSCCs have HPV integrated into the human genome [Mooren et al., 2013]. These types of assays are not at all accurate, nor can they determine where within the human or viral genome actual integrations occur. However, Olthof et al. [2014] reported that they detected 39% HPV integration from a larger collec-

tion of 75 HPV16-positive OPSCC samples using amplification of papillomavirus oncogene transcripts PCR (APOT-PCR) and detection of integrated papillomavirus sequences PCR (DIPS-PCR). Their results are more accurate as they determined the viral-human gene junction by Sanger sequencing and also found that half of the integration sites occurred near or within the CFS regions. By examining a large number of HNSCC cell lines, Olthof et al. [2015] also reported that HPV integration preferentially occurred within several CFS regions including FRAXC (Xq22.1), FRA3C (3q27), FRA9D (9q22.1), and FRA17B (17q23.1). These studies suggested that HPV integration preferentially occurs within CFS regions as has been observed in cervical cancers [Thorland et al., 2003].

HPV and Other Cancers

There are a number of other anogenital cancers that are frequently found to contain HPV sequences, including 85% of anal cancers and 50% of vaginal, vulvar, and penile cancers [Anorlu, 2008]. At present it is unclear what the HPV physical status in these different groups of HPV-positive cancers is. Perhaps, since they are all anogenital, they may have an identical mechanism for HPV-driven carcinogenesis as it is observed in cervical cancers. However, the role(s) that HPV play(s) in the development of these different cancers is not known.

Recent Next-Generation Sequencing Advances in HPV-Associated Cancers

The development of technologies utilizing massively parallel sequencing has resulted in a revolution in the way that genomes can be rapidly and comprehensively characterized. A single lane on an Illumina HiSeq 4000 can generate 400 million DNA sequence reads, and this is sufficient to analyze genomes, transcriptomes, and even methylomes. There are different sequencing strategies that can be employed to characterize genomic alterations and also the physical status of viruses within a cancer genome. For example, whole-genome sequencing (WGS) provides the most comprehensive characterization of the cancer genome, as it can discover the full range of genomic alterations, including nucleotide substitutions, indel structural variations, copy number alterations, and viral integrations. Unfortunately, this requires considerable sequencing, which leads to problems with both data analysis and storage, and it is still very costly. Exome sequencing, which involves analyzing only 2% of the entire ge-

nome, offers an alternative at a lower cost and has been an effective approach focusing on DNA sequencing of just the coding genes. This approach is suitable for mutation discovery in cancer samples and for both somatic and germline analysis; however, exome sequencing is less suited for characterizing the physical status of a virus, unless that virus has integrated next to one of the exons of a gene. Transcriptome sequencing, also called RNA-seq, is a sensitive and efficient approach to detect changes in gene expression profiles of both human and viral-specific genes and in addition can detect novel transcripts, changes in the expression of isoforms from different alternative splicing, as well as intra- and intergenic fusions. This strategy could prove useful but only if the HPV integration event led to the generation of a novel fusion transcript between human and HPV sequences.

WGS of the HPV18-positive HeLa cell line revealed several novel yet significant features [Adey et al., 2013]. It showed that the virus had integrated into the HPV18 integration hotspot which is located within 1 of the 2 CFS that flank the proto-oncogene *c-MYC*. The integration event in HeLa is quite complex, as various portions of the HPV18 genome are differentially amplified between 8 and 32 times. As the result of this integration, the entire amplified and altered region spans over 200 kb. The proto-oncogene *c-MYC* is situated just 500 kb upstream of this locus and phased RNA-seq data indicate that *c-MYC* is highly expressed [Adey et al., 2013]. WGS in other HPV-positive human cancer cell lines, including both cervical and head and neck cancers, have revealed that HPV integrants flank and bridge extensive host genomic amplifications and rearrangements including deletions, inversions, and chromosome translocations [Akagi et al., 2014]. These studies indicate that viral integration events could cause dramatic rearrangements in both the human and viral genomes and this could cause gene deletion or amplification at or surrounding the integration sites.

The sites of the HPV integrations in the human genome might play very important roles in the carcinogenesis of HPV-driven cancers as it has also been revealed that gene expression levels at HPV integration sites were statistically significantly higher in tumors with HPV integrations compared with expression of the same genes in tumors without viral integrations at the same sites [Ojesina et al., 2014]. By conducting WGS and high-throughput viral integration detection of cervical intraepithelial neoplasias, cervical carcinomas, and cervical cancer cell lines, Hu et al. [2015] also revealed several hotspots for HPV integrations which include the previously reported

frequent integration sites *POU5F1B*, *FHIT*, *KLF12*, *KLF5*, *LRP1B*, and *LEPREL1* and several new hotspots like *HMGA2*, *DLG2*, and *SEMA3D*. Among these, *FHIT*, *LRP1B*, and *DLG2* are known to be very large CFS genes located at 3p14.2, 2q22.1, and 11q14.1, respectively, and all the other genes have also been shown to play important roles in cancer development. Hu et al. [2015] also reported that protein expression was downregulated when HPV integrated into the introns of *FHIT* and *LRP1B*, while protein expression was elevated when HPV had integrated into the flanking regions of other genes, as was observed with *c-MYC* and *HMGA2*.

Even with the dramatic increases in sequence output made possible with the current generation of Illumina sequencers, the overall cost for WGS remains too high to analyze a large number of specimens to determine the physical status of HPV in OPSCC. However, a much easier and cheaper alternative to WGS is to construct mate-pairs (MP) libraries from fragments that are 5 kb in size and to do MP-Seq analysis of a large number of OPSCCs. The sequencing company Illumina produces a 5 kb MP library construction kit (Nextera) that can be used to generate MP libraries which are directly suitable for paired-end next-generation sequencing (NGS). By sequencing the ends of fragments which were originally 5 kb in size, one can obtain detailed genomic information with significantly less sequencing than WGS (WGS usually requires 100 Gb of sequencing, while MP-Seq can be done with just 5 Gb of sequence data). Our colleagues in the Biomarker Discovery Program at the Mayo Clinic have developed powerful algorithms to analyze these data [Feldman et al., 2011]. These algorithms readily reveal the physical status of HPV in any cancer genome and can also be used to determine changes in copy number throughout the human genome independent of the physical status of HPV. We utilized MP-Seq to characterize the physical status of HPV as well as other genomic alterations in HPV-positive OPSCCs and observed that only 30% of HPV-positive OPSCCs had HPV integrated into the human genome, which is in stark contrast to cervical cancer, where HPV integrations occur in over 90% of the specimens [Gao et al., 2014]. Our results are very similar to what Olthof et al. [2015] detected using APOT-PCR. Thus, the majority of HPV-positive OPSCCs develop just in the presence of episomal copies of HPV. We also observed that the HPV copy number was not associated with the physical status, as we found several samples with HPV integration into the human genome but still having a high episomal viral copy number. However, as was noticed in cervical cancers, both local human and viral ge-

nome rearrangements were also present in OPSCCs when integration had occurred [Gao et al., 2017].

Recent findings from NGS studies in both cervical cancer and OPSCC suggest that HPV integration can result in dramatic rearrangements of both human and viral sequences at the site of integration. Furthermore, certain integrations can be quite disruptive to genes at or near those integration sites, and this could potentially contribute to the eventual cancer that develops.

NGS studies on a large number of cancers not only presented powerful information about sites of integration into the human genome but also revealed a great deal about any potential specificity for where the HPV genome is disrupted and/or altered. Hu et al. [2015] found that integration could occur in any part of the viral genome, and there really was no specificity for alterations or integration to occur within the E2 gene. Similarly, out of 12 validated HPV integrations in OPSCC from our study, not a single one disrupted the HPV E2 gene. Instead, we found disruption of the viral genome could occur throughout the HPV genome, including the late genes L1, L2, as well as even E6 and E7 [Gao et al., 2017].

The Cancer Genome Atlas (TCGA) projects have used different platforms to comprehensively characterize many different cancers including head and neck cancers. Unfortunately, only a small proportion of the 279 head and neck cancers are from the oropharyngeal region and only 35 of those are HPV positive. The sequence data suggested that 25 of them had HPV integrated into the human genome. Although their reported HPV integration frequency is higher than in our study, their analysis is based solely on TCGA data bioinformatics analysis and these integration events were not validated by PCR and Sanger sequencing. Thus, they have not determined how many putative viral integrations are real events [Parfenov et al., 2014]. However, they did show that the viral integration could have a significant impact on the human genome at the site of integration, including alterations in DNA copy number, mRNA transcript abundance and splicing, and both inter- and intrachromosomal rearrangements, which is similar to what we and other groups have observed [Parfenov et al., 2014].

In addition to WGS and MP-Seq and their utility in examining structural variations and copy number changes at the whole-genome level, whole-exome sequencing has also provided gene mutation information for both cervical cancers and OPSCC. Whole-exome sequencing analysis of 115 cervical carcinoma-normal paired samples revealed recurrent somatic mutations in *EP300* (16%), *FBXW7* (15%), *PIK3CA* (14%), *HLA-B* (9%), *TP53* (9%),

MAPK1 (8%), *EBRR2* (5%), and *NFE2L2* (4%). Lechner et al. [2013] performed paired-end sequencing from laser-captured microdissected OPSCC, targeting 3,230 exons in 182 genes that were often observed to be mutated in OPSCC. They found that *PI3KCA* mutations or amplifications and *PTEN* inactivation by gene copy loss or mutations were seen in over 55% of HPV-positive tumors. This is consistent, as it was previously reported that HPV infections could cause activation of PI3K and deactivation of PTEN pathways in carcinogenesis, suggesting that these pathways are frequently involved in the development of HPV-positive OPSCCs. The mutations in *PI3KCA* are also frequently observed in HPV-positive OPSCCs, and this is similar to what has been reported in cervical cancer.

Re-Visiting HPV's Role(s) in Cancers

The new findings obtained from the recent NGS studies in cervical cancer certainly revealed new insights and have challenged some of the traditionally accepted models for HPV-driven cervical cancer: (1) HPV integration in the viral genome does not necessarily occur or disrupt the E2 gene but rather can occur at any place of the viral genome. (2) The HPV integration event itself could cause viral genome rearrangements, including copy number alterations close to the junction sites, as observed in the HeLa genome, where the partial viral genome could be amplified as many as 30 times. (3) The high level of the viral copy number is not derived necessarily from the episomal copy of the virus; it could be also from the partial amplified viral genome even after the virus has integrated into the human genome. (4) In addition to important roles that the E6 and E7 oncoproteins play in tumorigenesis, the site of the HPV integration into the human genome could indeed also be very important. It is especially interesting that Hu et al. [2015] also pointed out that the protein expression of genes was generally downregulated when HPV integrated into their introns, but protein expression was elevated when HPV integrated into the flanking regions of genes, as was observed for *c-MYC* and *HMGA2*. Thus, it is suggested that the sites of HPV integration into the human genome could be very important for the eventual cancer that develops. The significance of the viral integration sites in the human genome could be that HPV integration causes the deletion or the decreased expression of the gene where the virus integrated and the rearrangements or amplification of the surrounding genes near the integration sites. This would explain why decreased FHIT expression is often observed in cervical cancer, while there is dramatically increased ex-

pression of MYC in HeLa cells, as *c-MYC* is located some 500 kb upstream to the HPV integration site.

There are clearly some distinctions between HPV's role in the development of OPSCCs as compared to cervical cancers; however, there are also some similarities. When there is viral integration in OPSCC, the CFS regions appear to be hotspots for integration events, as is observed in cervical cancer. The CFS regions are also found to be hotspots for deletions and other alterations in OPSCCs independent of HPV integration events. What remains unclear in both cervical cancer and OPSCC is the importance of the site of HPV integration and the importance of the different types of alterations in expression of genes that are either at or near the sites of these integrations towards the cancers that develop.

CFSs and Large CFS Genes and HPV-Driven Cancers

There are a number of important associations between cancers that have an HPV etiology and the unstable CFS regions. Most cervical cancers have HPV integrated into the human genome, and over half of these occur within one of the CFS regions. While most OPSCCs do not have HPV integrated into the human genome, the CFS regions are indeed hotspots for HPV integration in those 30% tumors that do have HPV integration. Furthermore, over half of the genome-wide alterations including deletions and rearrangements detected in OPSCCs occur within one of the CFS regions. There are also a number of very large CFS genes which appear to be frequent targets of alteration in both cervical cancers and OPSCC. Some of these genes, which are located at the site of HPV integrations, were also found to have frequent deletions independent of HPV integrations. Some CFS regions including FRA1A at 1p36, FRA2C at 2q24.2, FRA7G at 7q31.1, and FRA16C at 16q22 were found to be viral integration sites in both cervical cancer and OPSCCs. However, some frequently disrupted CFS regions and their large genes are differentially targeted in these 2 types of cancers. For example, *FHIT* was frequently observed as an HPV integration sites in cervical cancer, but we have not observed the disruption of this gene in OPSCC. Deletions and rearrangements of the third largest known human gene, *CSMD1*, are frequently observed in OPSCC but not in cervical cancer. This suggests that cells with different origins might have different CFS regions. On the other hand, due to the low HPV integration frequency, the absolute number of viral integration sites observed in OPSCC is limited; thus it is difficult to make very thorough conclusions.

There are multiple questions that remain. For example, whether there is a selection for HPV integration events that cause some type of phenotypic change due to disruption of important large CFS genes or other non-CFS genes that play an important role in cancer development. Would determining the physical status of HPV in any given cancer and also the sites(s) of integration have any clinical significance? It is clear that NGS now provides some very powerful tools for the characterization of alterations in a cancer genome, and the clinical applications of this technology will definitely transform how we characterize cancers and treat our patients.

Conclusions

NGS provides very powerful tools to characterize the genomic alterations that have occurred during the development of HPV-driven cancers and also to reveal the

physical status of HPV in different cancers. Recent NGS studies showed that the site of integration within the HPV genome is random. On the contrary, it appears that the sites of integration within the human genome are the ones that are nonrandom. The integration events themselves can be quite disruptive and can dramatically change gene expression in and around the integration site.

OPSCC appears distinct from cervical cancer, as most HPV-positive OPSCCs develop just in the presence of episomal copies of HPV. HPV may therefore play variable roles in the development of different HPV-positive OPSCCs. One important similarity, however, between cervical cancers and HPV-positive OPSCCs is that the CFS regions and the large genes contained within appear to be important for the development of both types of cancers.

Disclosure Statement

The authors have no conflicts of interest to declare.

References

- Adey A, Burton JN, Kitzman JO, Hiatt JB, Lewis AP, et al: The haplotype-resolved genome and epigenome of the aneuploid HeLa cancer cell line. *Nature* 500:207–211 (2013).
- Akagi K, Li J, Broutian TR, Padilla-Nash H, Xiao W, et al: Genome-wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. *Genome Res* 24: 185–199 (2014).
- Anorlu RI: What is the significance of the HPV epidemic? *Can J Urol* 15:3860–3965 (2008).
- Badaracco G, Venuti A, Sedati A, Marcante ML: HPV16 and HPV18 in genital tumors: significantly different levels of viral integration and correlation to tumor invasiveness. *J Med Virol* 67:574–582 (2002).
- Becker NA, Thorland EC, Denison SR, Phillips LA, Smith DI: Evidence that instability within the FRA3B region extends four megabases. *Oncogene* 21:8713–8722 (2002).
- Brescia RJ, Jenson AB, Lancaster WD, Kurman RJ: The role of human papillomaviruses in the pathogenesis and histologic classification of precancerous lesions of the cervix. *Hum Pathol* 17:552–559 (1986).
- Chaturvedi AK, Engels EA, Pfeiffer RM, Hernandez BY, Xiao W, et al: Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol* 29:4294–4301 (2011).
- Chen J: Signaling pathways in HPV-associated cancers and therapeutic implications. *Rev Med Virol* 25 Suppl 1:24–53 (2015).
- Cullen AP, Reid R, Campion M, Lörincz AT: Analysis of the physical state of different human papillomavirus DNAs in intraepithelial and invasive cervical neoplasm. *J Virol* 65: 606–612 (1991).
- Dall KL, Scarpini CG, Roberts I, Winder DM, Stanley MA, et al: Characterization of naturally occurring HPV16 integration sites isolated from cervical keratinocytes under non-competitive conditions. *Cancer Res* 68:8249–8259 (2008).
- Deng WG, Nishizaki M, Fang B, Roth JA, Ji L: Induction of apoptosis by tumor suppressor FHIT via death receptor signaling pathway in human lung cancer cells. *Biochem Biophys Res Commun* 355:993–999 (2007).
- Deng Z, Hasegawa M, Kiyuna A, Matayoshi S, Uehara T, et al: Viral load, physical status, and E6/E7 mRNA expression of human papillomavirus in head and neck squamous cell carcinoma. *Head Neck* 35:800–808 (2013).
- Deng Z, Hasegawa M, Aoki K, Matayoshi S, Kiyuna A: A comprehensive evaluation of human papillomavirus positive status and p16^{INK4a} overexpression as a prognostic biomarker in head and neck squamous cell carcinoma. *Int J Oncol* 45:67–76 (2014).
- Denison SR, Callahan G, Becker NA, Phillips LA, Smith DI: Characterization of FRA6E and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. *Genes Chromosomes Cancer* 38:40–52 (2003).
- DiMaio D, Mattoon D: Mechanisms of cell transformation by papillomavirus E5 proteins. *Oncogene* 20:7866–7873 (2001).
- Dyson N, Howley PM, Münger K, Harlow E: The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 243:934–947 (1989).
- Feldman AL, Dogan A, Smith DI, Law ME, Ansell SM, et al: Discovery of recurrent t(6;7)(p25.3;q32.3) translocations in ALK-negative anaplastic large cell lymphomas by massively-parallel genomic sequencing. *Blood* 117:915–919 (2011).
- Ferber MJ, Thorland EC, Brink AA, Rapp AK, Phillips LA, et al: Preferential integration of human papillomavirus type 18 near the *c-myc* locus in cervical carcinoma. *Oncogene* 22: 7233–7242 (2003).
- Flaherty A, Kim T, Giuliano A, Magliocco A, Hakky TS, et al: Implications for human papillomavirus in penile cancer. *Urol Oncol* 53:e1–e8 (2014).
- Gao G, Chernock RD, Gay HA, Thorstad WL, Zhang TR, et al: A novel RT-PCR method for quantification of human papillomavirus transcripts in archived tissues and its application in oropharyngeal cancer prognosis. *Int J Cancer* 132:882–890 (2013).
- Gao G, Johnson SH, Kasperbauer JL, Eckloff BW, Tombers NM, et al: Mate pair sequencing of oropharyngeal squamous cell carcinomas reveals that HPV integration occurs much less frequently than in cervical cancer. *J Clin Virol* 59:195–200 (2014).
- Gao G, Johnson SH, Vasmataz G, Pauley CE, Tombers NM, et al: Common fragile sites (CFS) and extremely large CFS genes are targets for human papillomavirus integrations and chromosome rearrangements in oropharyngeal squamous cell carcinoma. *Genes Chromosomes Cancer* 56:59–74 (2017).

- Giarrè M, Caldeira S, Malanchi I, Ciccolini F, Leão MJ, Tommasino M: Induction of pRb degradation by the human papillomavirus type 16 E7 protein is essential to efficiently overcome p16^{INK4a}-imposed G₁ cell cycle arrest. *J Virol* 75:4705–4712 (2001).
- Glover TW, Berger C, Coyle J, Echo B: DNA polymerase alpha inhibition by aphidicolin induces gaps and breaks at common fragile sites in human chromosomes. *Hum Genet* 67:136–142 (1984).
- Hidalgo A, Monroy A, Arana RM, Taja L, Vázquez G, Salcedo M: Chromosomal imbalances in four new uterine cervix carcinoma derived cell lines. *BMC Cancer* 3:8 (2003).
- Hoffmann M, Klose N, Gottschlich S, Görög T, Fazel A, et al: Detection of human papillomavirus DNA in benign and malignant sinonasal neoplasms. *Cancer Lett* 239:64–70 (2006).
- Hoots BE, Palefsky JM, Pimenta JM, Smith JS: Human papillomavirus type distribution in anal cancer and anal intraepithelial lesions. *Int J Cancer* 124:2375–2383 (2009).
- Hu Z, Zhu D, Wang W, Li W, Jia W, et al: Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology-mediated integration mechanism. *Nat Genet* 47:158–163 (2015).
- Imai Y, Soda M, Takahashi R: Parkin suppresses unfolded protein stress-induced cell death through its E3 ubiquitin-protein ligase activity. *J Biol Chem* 275:35661–35664 (2000).
- Jenson AB, Geyer S, Sundberg JP, Ghim S: Human papillomavirus and skin cancer. *J Investig Dermatol Symp Proc* 6:203–206 (2001).
- Joannes A, Grelet S, Duca L, Gilles C, Kileztky C, et al: Fhit regulates EMT targets through an EGFR/Src/ERK/Slug signaling axis in human bronchial cells. *Mol Cancer Res* 12:775–783 (2014).
- Koskinen WJ, Chen RW, Leivo I, Mäkitie A, Bäck L, et al: Prevalence and physical status of human papillomavirus in squamous cell carcinomas of the head and neck. *Int J Cancer* 107:401–406 (2003).
- Kreimer AR, Clifford GM, Boyle P, Franceschi S: Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. *Cancer Epidemiol Biomarkers Prev* 14:467–475 (2005).
- Krummel KA, Denison SR, Calhoun E, Phillips LA, Smith DI: The common fragile site FRA16D and its associated gene WWOX are highly conserved in the mouse at Fra8E1. *Genes Chromosomes Cancer* 34:154–167 (2002).
- Kumar N: Cervical cancer; a nightmare for womanhood: review of recent advances. *Womens Health Gynecol* 2:17 (2016).
- Kuroki T, Yendamuri S, Trapasso F, Matsuyama A, Aqeilan RI, et al: The tumor suppressor gene WWOX at FRA16D is involved in pancreatic carcinogenesis. *Clin Cancer Res* 10:2459–2465 (2004).
- Lechner M, Frampton GM, Fenton T, Feber A, Palmer G, et al: Targeted next-generation sequencing of head and neck squamous cell carcinoma identifies novel genetic alterations in HPV+ and HPV- tumors. *Genome Med* 5:49 (2013).
- Licitra L, Bernier J, Grandi C, Merlano M, Bruzzi P, et al: Cancer of the oropharynx. *Crit Rev Oncol Hematol* 41:107–22 (2002).
- Liu X, Schiffman M, Hulbert A, He Z, Shen Z, et al: Association of human papillomavirus 31 DNA load with risk of cervical intraepithelial neoplasia grades 2 and 3. *J Clin Microbiol* 53:3451–3457 (2015).
- Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, et al: Clonal genetic alterations in the lungs of current and former smokers. *J Natl Cancer Inst* 89:857–862 (1997).
- Marur S, D'Souza G, Westra WH, Forastiere AA: HPV-associated head and neck cancer: a virus-related cancer epidemic. *Lancet Oncol* 11:781–789 (2010).
- Matovina M, Sabol I, Grubisic G, Gasperov NM, Grce M: Identification of human papillomavirus type 16 integration sites in high-grade precancerous cervical lesions. *Gynecol Oncol* 113:120–127 (2009).
- Mellin H, Friesland S, Auer G, Dalianis T, Munk-Wikland E: Human papillomavirus and DNA ploidy in tonsillar cancer—correlation to prognosis. *Anticancer Res* 23:2821–2828 (2003).
- Mooren JJ, Kremer B, Claessen SM, Voogd AC, Bot FJ, et al: Chromosome stability in tonsillar squamous cell carcinoma is associated with HPV16 integration and indicates a favorable prognosis. *Int J Cancer* 132:1781–1789 (2013).
- Näsmän A, Attner P, Hammarstedt L, Du J, Eriksson M: Incidence of human papillomavirus (HPV) positive tonsillar carcinoma in Stockholm, Sweden: an epidemic of viral-induced carcinoma? *Int J Cancer* 125:362–366 (2009).
- Ni S, Hu J, Duan Y, Shi S, Li R, et al: Down expression of LRP1B promotes cell migration via RhoA/Cdc42 pathway and actin cytoskeleton remodeling in renal cell cancer. *Cancer Sci* 104:817–825 (2013).
- Ojesina AI, Lichtenstein L, Freeman SS, Pedamallu CS, Imaz-Rosshandler I, et al: Landscape of genomic alterations in cervical carcinomas. *Nature* 506:371–375 (2014).
- Olthof NC, Speel EJ, Kolligs J, Haesevoets A, Henfling M, et al: Comprehensive analysis of HPV16 integration in OSCC reveals no significant impact of physical status on viral oncogene and virally disrupted human gene expression. *PLoS One* 9:e88718 (2014).
- Olthof NC, Huebbers CU, Kolligs J, Henfling M, Ramaekers FC, et al: Viral load, gene expression and mapping of viral integration sites in HPV16-associated HNSCC cell lines. *Int J Cancer* 136:E207–E218 (2015).
- Panwar A, Batra R, Lydiatt WM, Ganti AK: Human papilloma virus positive oropharyngeal squamous cell carcinoma: a growing epidemic. *Cancer Treat Rev* 40:215–219 (2014).
- Parfenov M, Pedamallu CS, Gehlenborg N, Freeman SS, Danilova L, et al: Characterization of HPV and host genome interactions in primary head and neck cancers. *Proc Natl Acad Sci USA* 111:15544–15549 (2014).
- Pirami L, Giachè V, Becciolini A: Analysis of HPV16, 18, 31, and 35 DNA in pre-invasive and invasive lesions of the uterine cervix. *J Clin Pathol* 50:600–604 (1997).
- Rietbergen MM, Leemans CR, Bloemena E, Heideman DA, Braakhuis BJ, et al: Increasing prevalence rates of HPV attributable oropharyngeal squamous cell carcinomas in the Netherlands as assessed by a validated test algorithm. *Int J Cancer* 132:1565–1571 (2013).
- Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM: The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 63:1129–1136 (1990).
- Scheffner M, Huibregtse JM, Vierstra RD, Howley PM: The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* 75:495–505 (1993).
- Sozzi G, Pastorino U, Moiraghi L, Tagliabue E, Pezzella F, et al: Loss of FHIT function in lung cancer and preinvasive bronchial lesions. *Cancer Res* 58:5032–5037 (1998).
- Srodon M, Stoler MH, Baber GB, Kurman RJ: The distribution of low and high-risk HPV types in vulvar and vaginal intraepithelial neoplasia (VIN and VaIN). *Am J Surg Pathol* 30:1513–1518 (2006).
- Thorland EC, Myers SL, Gostout BS, Smith DI: Common fragile sites are preferential targets for HPV16 integrations in cervical tumors. *Oncogene* 22:1225–1237 (2003).
- Tommasino M: The human papillomavirus family and its role in carcinogenesis. *Semin Cancer Biol* 26:13–21 (2014).
- Ukpo OC, Moore EJ, Smith DI: Human papillomavirus and oropharyngeal cancer. *N Engl J Med* 357:1156–1157 (2007).
- Vega-Peña A, Illades-Aguir B, Flores-Alfaro E, López-Bayghen E, Leyva-Vázquez MA, et al: Risk of progression of early cervical lesions is associated with integration and persistence of HPV-16 and expression of E6, Ki-67, and telomerase. *J Cytol* 30:226–232 (2013).
- Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, et al: Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 189:12–19 (1999).
- Watson JE, Doggett NA, Albertson DG, Andaya A, Chinnaiyan A, et al: Integration of high-resolution array comparative genomic hybridization analysis of chromosome 16q with expression array data refines common regions of loss at 16q23-qter and identifies underlying candidate tumor suppressor genes in prostate cancer. *Oncogene* 23:3487–3494 (2004).
- zur Hausen H: Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2:342–350 (2002).