Mutation Research 772 (2017) 3–12

Contents lists available at ScienceDirect

Mutation Research/Reviews in Mutation Research

journal homepage: www.elsevier.com/locate/reviewsmr
Community address: www.elsevier.com/locate/mutres

Review

Human papillomavirus molecular biology

Mallory E. Harden[a,b], Karl Munger[b,*]

[a] Program in Virology, Division of Medical Sciences, Harvard Medical School, Boston, MA, 02115, USA
[b] Department of Developmental, Molecular and Chemical Biology, Tufts University School of Medicine, Boston, MA, 02111, USA

ARTICLE INFO

Article history:
Received 12 April 2016
Received in revised form 13 June 2016
Accepted 4 July 2016
Available online 5 July 2016

Keywords:
Human papillomavirus
Cervical cancer
Oncogene
Tumor suppressor
Epithelial differentiation
Vaccine

ABSTRACT

Human papillomaviruses are small DNA viruses with a tropism for squamous epithelia. A unique aspect of human papillomavirus molecular biology involves dependence on the differentiation status of the host epithelial cell to complete the viral lifecycle. A small group of these viruses are the etiologic agents of several types of human cancers, including oral and anogenital tract carcinomas. This review focuses on the basic molecular biology of human papillomaviruses.

Contents

1. Introduction

With a prevalence of 70 million cases and an incidence of 14 million new transmissions each year, human papillomavirus (HPV) infections of the anogenital tract are the most common sexually transmitted diseases in the US [1]. High-risk (HR) HPVs are the causative agents of cervical cancer and, worldwide, it is estimated that 500,000 cases of cervical cancer occur each year, which result in over 250,000 deaths [2]. Cervical cancer is the 4th most common cancer in women and the 7th most common cancer overall [3]. The burden of cervical cancer is disproportionately high in low-income countries due to a scarcity in resources to implement widespread screening, vaccination and treatment programs [4]. While safe and

* Corresponding author at: 150 Harrison Avenue, Jaharis 607, Boston, MA 02111, USA.
E-mail address: karl.munger@tufts.edu (K. Munger).

http://dx.doi.org/10.1016/j.mrrev.2016.07.002
1383-5742/

efficacious vaccines for the prevention of HPV infection are available, they do not protect those already infected with HPV and they do not protect against all HPV types. Therefore, continued studies of the molecular biology of HPV are necessary to develop improved screening techniques and prophylactic vaccines for the prevention of HPV infection, as well as better therapeutic options, including vaccines, for the treatment of HPV infection.

2. HPV classification

HPVs are members of the distinct virus family, the *Papillomaviridae*. The *Papillomaviridae* family is divided into 39 genera [5,6], based on L1 sequence identity of 60% or greater, with each genus designated by a letter of the Greek alphabet. PVs within a genus that share 60–70% L1 sequence identity are termed a species [5,6]. Additionally, within a species, PVs with 71–89% L1 sequence identity are considered a type [5,6]. As of 2016, 205 different HPV types have been identified, which have been categorized into five genera including the following: 65 *Alphapapillomaviruses*, 51 *Betapapillomaviruses*, 84 *Gammapapillomaviruses*, 4 *Mupapillomaviruses* and a single *Nupapillomavirus* [7]. At least 19 more additional types have been identified however, these viruses are currently pending classification. HPVs with 90–98% L1 sequence identity are termed subtypes and those with >98% L1 sequence identity are considered variants [5]. Arguably, HPVs in the alpha genus are of the greatest medical importance given they are associated with oral and mucosal cancers, as well as cancers of the anogenital tract. Table 1 includes a summary of the main HPV genotypes and their associated diseases.

3. Virion and genome structure and organization

HPVs are non-enveloped DNA viruses with a tropism for the squamous epithelium. Each virus particle consists of an icosahedral capsid of about 60 nm in diameter, containing a single molecule of double stranded circular DNA of approximately 8000 base pairs [8]. Only one strand of the double stranded DNA genome is used as a template for transcription and this coding strand contains three genomic regions, including approximately ten open reading frames (ORFs) shown in Fig. 1. Many viral proteins are expressed from polycistronic mRNAs [9,10]. The early region (E) contains up to seven ORFs encoding viral regulatory proteins and the late (L) region encodes the two viral capsid proteins. Each ORF in the early region is designated "E" followed by a numeral, indicative of the length of the ORF. The third region of the genome has been referred to as the long control region (LCR), the upstream regulatory region (URR) or the noncoding region (NCR). This genomic region contains the origin of DNA replication, as well as transcription control sequences [8].

3.1. Viral proteins

The early HPV ORFs include E1, E2, E4, E5, E6, E7 and E8 [10,11] (see Fig. 1). E1 codes for an ATP dependent viral DNA helicase [12] that can bind to the AT-rich origin of replication and E2 proteins function in viral transcription, replication and genome partitioning. The full length E2 protein encodes a transcriptional activator. In contrast, a truncated form of E2 transcribed from an internal ATG and the E8^E2 fusion protein repress transcription [13]. E4 is embedded within the E2 gene and is primarily expressed as an E1^E4 fusion protein during the late stages of the viral life cycle. E4 binds to cytokeratin filaments, disrupting their structure, and is thought to play a role in viral escape from cornified epithelial layers [14]. E5 is a small transmembrane protein, which has been best studied with bovine papillomavirus type 1 (BPV1). BPV1 E5 is an oncogenic small, hydrophobic, single pass transmembrane protein that forms dimers and interacts with and activates receptor tyrosine kinase receptors, including the EGF and PDGF receptors. Similar activities have also been ascribed to HPV E5, which encodes multi pass transmembrane proteins that share only limited sequence similarity with BPV1 E5 [15]. HPV E5 proteins have also been reported to play a role in apoptosis and in evasion of the immune response [15]. HPV E6 and E7 both drive cell cycle entry to allow genome amplification in upper epithelial layers. HR HPV E6 proteins have oncogenic activities. They bind and degrade p53, as well as cellular PDZ proteins, and they activate telomerase [16]. HR HPV E7 proteins bind and degrade the retinoblastoma tumor suppressor, pRB, and contribute to malignant progression by inducing genomic instability [17,18]. The late region encodes the major (L1) and minor (L2) capsid proteins (see Fig. 1). Given the L1

Table 1
Main HPV genotypes and their associated diseases.
This table summarizes information on the main HPV genotypes, their tropism and associated diseases. Information in this table was gathered from several sources including pave.niaid.nih.gov [6,7,135]. Heck's disease, also known as focal epithelial hyperplasia (FEH), is a rare, benign mucosal proliferation that is strongly associated with HPV infection [136]. Other details on HPV classification can be found in the text in Section 2.

Genus	Species	Representative HPV types	Tropism	Associated Diseases
Alpha-PV	**α1**	32	mucosal	Heck's disease
	α2	3, 10, 28	cutaneous	flat warts
	α4	2, 27, 57	cutaneous	common warts
	α7	18, 39, 45, 59, 68	mucosal	intraepithelial neoplasia, invasive carcinoma
	α9	16, 31, 33, 35, 52, 58	mucosal	intraepithelial neoplasia, invasive carcinoma
	α10	6, 11	mucosal	condylomata acuminate
		13		Heck's disease
Beta-PV	**β1c**	5, 8, 12, 14, 19, 20, 21, 24, 25, 36, 47	cutaneous	Epidermodysplasia verruciformis
	β2	9, 15, 17, 22, 23, 37, 38	cutaneous	Epidermodysplasia verruciformis
	β3	49	cutaneous	Epidermodysplasia verruciformis
Gamma-PV	**γ1**	4, 65	cutaneous	Warts
	γ4	60	cutaneous	Warts
Mu-PV	**μ1**	1	cutaneous	plantar warts
	μ2	63	cutaneous	Warts
Nu-PV		41	cutaneous	Warts