Biomedicine & Pharmacotherapy 106 (2018) 1537–1556

Contents lists available at ScienceDirect

Biomedicine & Pharmacotherapy

journal homepage: www.elsevier.com/locate/biopha

Review

The human papillomavirus (HPV)-related cancer biology: An overview

Rodrigo Pinheiro Araldi[a], Thalita Araujo Sant'Ana[b], Diego Grando Módolo[b], Thatiana Correa de Melo[c], Diva Denelle Spadacci-Morena[d], Rita de Cassia Stocco[b], Janete Maria Cerutti[a], Edislane Barreiros de Souza[e,*]

[a] Genetics Basis of Thyroid Cancer Laboratory, São Paulo Federal University (UNIFESP), São Paulo, SP, Brazil
[b] Genetics Laboratory, Butantan Institute, São Paulo, SP, Brazil
[c] Haematology and Molecular Biology Laboratory, Medical School, University of São Paulo (USP), São Paulo, SP, Brazil
[d] Pathophysiology Laboratory, Butantan Institute, São Paulo, SP, Brazil
[e] Genetics Laboratory, Biotechnology Department, Faculty of Science and Languages, São Paulo State University (UNESP), Assis, SP, Brazil

ARTICLE INFO

Keywords:
Oncology
HPV
Mutation
Cancer progression
Metastasis

ABSTRACT

Despite the novel diagnostic methods and therapies implemented in oncology, the number of patients that succumb by the cancer remains high globally. Currently studies point out that 20–25% of all human malignancies are related to micro-organism infections. Among these cancer-related pathogens, the human papillomavirus (HPV) has a prominent position, since the virus is responsible for about 30% of all infectious agent-related cancers. Thus, an amount of cancers could be avoided by means prophylactic and/or therapeutic measures. However, these measures required a holistic comprehension about HPV-related cancer biology. Based on this, this review aims to summarize the last evidences of HPV on cancer biology (from initiation to metastasis), focus on molecular and biochemical deregulations associated with viral infection, and discuss the viral etiology in different malignancies.

1. Infectious agent and cancer

Despite the novel diagnostic and therapeutic methods implemented in oncology, the cancer incidence have been grown, making the disease an important public health problem globally [1]. In 2012, 14.1 million of people were diagnosed with cancer and, 8.2 million dead due to the disease [2].

Theories aiming to describe the carcinogenesis have been proposed for centuries [3]. In this context, the identification of the intricate relation between virus and cancer was a landmark in the discoveries that led to the formulation of the modern cancer biology concepts [4].

Currently it is widely accepted that infectious agents are responsible for 20–25% of all cancer cases globally registered [4]. Among these agents, the virus stands out to be responsible for about 12% of all human cancers [5]. Among these viruses, the human papillomavirus (HPV) occupies a prominent position, being responsible for about 30% of all infectious agent-related cancers [6].

It is estimated that HPV causes 610,000 incident cancer cases [7] and 250,000 deaths every year [8]. According to the World Health Organization (WHO), 85% of these deaths occur in low- and middle-income countries (http://www.who.int/mediacentre/factsheets/fs380/en/). However, the HPV is not only a public health problem of developing countries, since the virus infects about 6.2 million of people in United States of America annually [9].

Based on the HPV impacts on oncology, this review aimed to summarize the most recent advances in HPV-related cancer biology, describing the viral action in each step of carcinogenesis, focus on metabolic deregulations induced by the viral oncoproteins. This review also discusses the HPV participation in different malignancies, including evidences of viral participation in different malignancies.

2. Basic aspects of human papillomavirus biology

2.1. Morphological and genetics aspects

The HPV belongs to *Papillomaviridae* family, one of the oldest viral family known [10], able to infect epithelial cells of the skin, oral and genital mucosa [11]. The HPV viral particles (virions) exhibit a conserved icosahedral morphology [11], with 50–55 nm of diameter [12,13] and a molecular weight of 5×10^6 Da [14]. HPV is evolutionarily conserved, presenting a divergence rate of 1% per 40,000–80,000 years [15,16]. The HPV has a circular double-strand

* Corresponding author at: Genetics Laboratory, Biotechnology Department, Faculty of Science and Languages, São Paulo State University (UNESP), Assis, São Paulo, 19806-900, Brazil.
E-mail address: edislane@assis.unesp.br (E.B. de Souza).

https://doi.org/10.1016/j.biopha.2018.06.149
Received 26 April 2018; Received in revised form 24 June 2018; Accepted 27 June 2018

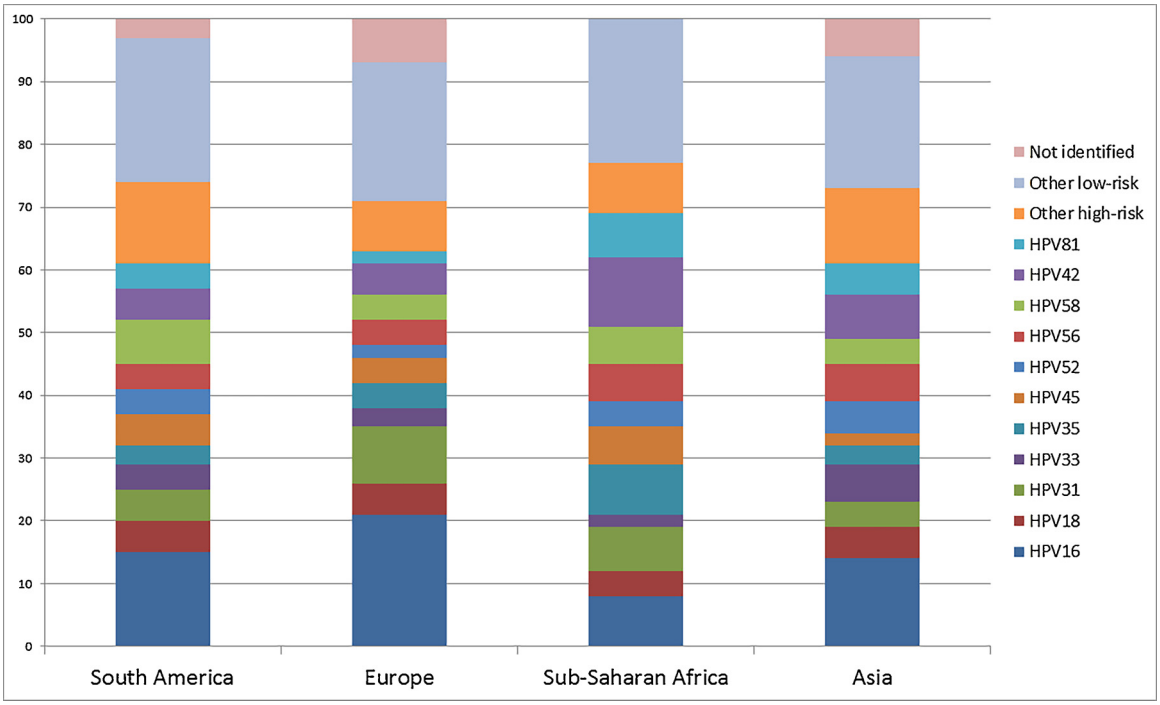

Fig. 1. Prevalence of the most common HPV types in different geographical areas. Adapted from Clifford et al. [25].

DNA, with approximately 8000 bp, which is associated with histone-like proteins [17].

The viral genome is divided into three regions: (1) early (E) region - containing the genes E1, E2, E4, E5, E6 and E7, which are associated with viral replication; (2) late (L) region, which encodes the major (L1) and minor (L2) capsid proteins; and (3) a non-coding region (NCR), also known as long control region (LCR), which is located between the L1 and E6 open reading frames (ORFs) [11]. Although not encoding, the LCR contains most of the regulatory elements involved in viral DNA replication and transcription, including the replication origin (*ori*) [11].

Nearly 280 papillomavirus types were already described in vertebrates [18]. More than 200 types infect human [18,19]. All HPVs can induce proliferative benign lesions [20]. However, 12 viral types are closely related with malignant neoplasms (HPV-16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58 and 59). For this reason, the International Agency for Research on Cancer (IARC) classified these HPVs as high-risk (HR) [11]. Altogether, the HR-HPVs-16, 18, 31, 33, 45 and the low-risk (LR)-HPVs-6 and 11 comprise the most prevalent virus types [21–23]. Although the distribution of these HPVs change according to geographical area [24], as showed in Fig. 1, the HR-HPVs-16, 18, 31, 33 and 45 are responsible for 75% of all HPV-associated squamous cell carcinoma and 94% of all adenocarcinomas [22,24].

2.2. Proteins codified by late region (L1 and L2)

The L1 ORF codifies the major capsid protein (L1) [25], with 55 kDa [26]. The L1 ORF is also the most conserved genome sequence and, for this reason, it is used for phylogenetic classification of viral types and subtypes [27,28]. The L1 protein is crucial for the viral infection, since it promotes the virion binding to heparin sulfate receptors present in the basal membrane [25]. The L2 is a 64–78 kDa protein necessary for viral infection and assembly, facilitating the DNA packaging [29,30]. Considering that these proteins are exclusively related to the viral infection and assembly, they are lately expressed, can be detected in differentiated epithelium layers [31].

2.3. Proteins codified by early region (E1, E2 and E4)

The E1 and E2 ORFs codify the E1 and E2 proteins, which are essential for viral replication [32]. The E1 ORF is the second most conserved sequence among the papillomaviruses and codifies the E1 protein with 68 kDa [33,34]. The E1 protein has three functional domains: (1) N-terminal domain, that binds to motif region of Cdk2; (2) central domain, that binds to E2 protein, forming the E1-E2 complex and (3) C-terminal domain, that acts as an ATP-dependent helicase [30,35,36]. After the E1-E2 complex binding to viral *Ori* [34,37], the E1 protein forms a dihexameric complex [38], attracting topoisomerase I, DNA polymerase α and replication protein A (RPA), which are required for the viral replication [37]. The E1 protein also promotes DNA breaks in host chromatin, contributing to viral integration [37].

The E2 is a 48 kDa modular protein [39] that presents two domains: (1) C-terminal and (2) N-terminal transactivation domain [35]. The C-terminal domain binds to the Brd4 protein, which bromodomains interact with lysine residues of acetylated histones, resulting in chromatin remodeling by acting as a super-enhancer [40,41]. The E2-Brd4 complex binds to mitotic chromosomes, allowing the distribution of viral copies to daughter cells after cytokinesis [29,35].

The E2 acts as a transcriptional regulator of E6 and E7 ORFs [30,42–44]. When present in high levels, the E2 binds to 5′-ACCG(N)$_4$CGGT-3′ palindromic sequence present in E2 binding sites (E2BS) in LCR, including in the P97 promoter [42,45]. This action prevents the binding of RNA polymerase II to P97, repressing the E6 and E7 expression [30]. However, when in low levels, the E2 recruits transcription factors required to replication [40,41]. Thus, the E2 also regulates the number of viral copies [46]. The E2 is also an epigenetic regulator, being able to interact with p300/CBP-p/CAF complex, recognized as a global transcription activator [47], promoting the *TP53* hypo-acetylation and, therefore, reducing the p53 transcriptional activity [35].

The E4 ORF codifies a protein family produced by splicing followed by post-translational modifications [48]. The E4 protein is expressed in suprabasal and granular epithelium layers [49], being the most expressed viral protein [50,51]. The E4 interacts with keratin-associated amyloid fibers [51], leading to cell fragility, contributing to virion