Tumour Virus Research 15 (2023) 200258

Contents lists available at ScienceDirect

Tumour Virus Research

journal homepage: www.journals.elsevier.com/tumour-virus-research

Human papillomavirus genomics: Understanding carcinogenicity

Chase W. Nelson [a,b,*], Lisa Mirabello [a,*]

[a] Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, 20850, USA
[b] Institute for Comparative Genomics, American Museum of Natural History, New York, NY, 10024, USA

ARTICLE INFO

Keywords:
Cervical cancer
HPV16
HPV evolution
HPV genomics
Next-generation sequencing (NGS)
Within-host (intrahost) diversity

ABSTRACT

Human papillomavirus (HPV) causes virtually all cervical cancers and many cancers at other anatomical sites in both men and women. However, only 12 of 448 known HPV types are currently classified as carcinogens, and even the most carcinogenic type — HPV16 — only rarely leads to cancer. HPV is therefore necessary but insufficient for cervical cancer, with other contributing factors including host and viral genetics. Over the last decade, HPV whole genome sequencing has established that even fine-scale within-type HPV variation influences precancer/cancer risks, and that these risks vary by histology and host race/ethnicity. In this review, we place these findings in the context of the HPV life cycle and evolution at various levels of viral diversity: between-type, within-type, and within-host. We also discuss key concepts necessary for interpreting HPV genomic data, including features of the viral genome; events leading to carcinogenesis; the role of APOBEC3 in HPV infection and evolution; and methodologies that use deep (high-coverage) sequencing to characterize within-host variation, as opposed to relying on a single representative (consensus) sequence. Given the continued high burden of HPV-associated cancers, understanding HPV carcinogenicity remains important for better understanding, preventing, and treating cancers attributable to infection.

1. Introduction

Human papillomavirus (HPV) causes ~4.5% of all human cancers [1], including tumours of the cervix, anus, vagina, penis, oropharynx, vulva, oral cavity, and larynx [2]. Cervical cancer is the most common of these, with 604,000 new cases and 342,000 deaths per year, virtually all attributable to HPV [3]. HPV is one of the most consequential human carcinogens [4,5]. However, the majority of HPV types (genotypes) do not cause cancer; of 448 types that have been documented [6,7], only 12 are currently classified as carcinogenic: types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, and 59 [8]. Infections by these carcinogenic HPV types are extremely common [9], but ~80% are cleared by the immune system within three years, and only ~3% progress to cervical precancer/cancer within 7 years [10]. HPV is therefore necessary but insufficient for cervical cancer. Further, because infectious virus particles are not produced in tumours [11], cancer cannot provide an evolutionary benefit to the virus [12]. Cancer is therefore a rare and inadvertent consequence — not an objective — of HPV infection.

HPV16 and HPV18 are the most common carcinogenic types [1], together responsible for ~71% of cervical cancers [13,14] and virtually all HPV-associated cancers in males [2]. However, despite advances in genomics [15], pinpointing genetic variants that confer differences in carcinogenicity has remained elusive, due in part to a lack of sufficiently abundant HPV whole genome sequences [16]. Even when genomes are available, data interpretation can be complex. For example, there is a poor correlation between HPV genetic relatedness and carcinogenicity: HPV31 and HPV35 are the types most closely related (genetically similar) to HPV16, but they are much less carcinogenic. At the same time, HPV18 is highly carcinogenic, but it is relatively distantly related to HPV16 and preferentially causes glandular lesions.

In this review, we discuss HPV-related carcinogenesis from the perspective of genomics, focusing on HPV16, cervical cancer, and key concepts necessary for the interpretation of genomics data (see Box 1 for Glossary of **bold** terms). We also document how next-generation sequencing (NGS) has dramatically increased the number of HPV genome sequences over the last decade, leading to new discoveries about genetic differences between HPV **types**, within the same HPV type, and even among HPV genomes that infect a single **host** individual.

* Corresponding authors. Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, 20850, USA.
E-mail addresses: chase.nelson@nih.gov (C.W. Nelson), mirabellol@mail.nih.gov (L. Mirabello).

https://doi.org/10.1016/j.tvr.2023.200258
Received 4 November 2022; Received in revised form 1 February 2023; Accepted 17 February 2023
Available online 20 February 2023

Box 1
Glossary of Key Concepts.

Between-host (interhost): genetic differences between viruses infecting different individuals. Contrast within-host.

Between-type (intertypic): genetic differences between HPV types, with patterns of relatedness typically determined using the L1 ORF. Contrast within-type.

Consensus: a summary nucleotide sequence wherein each genome position has been assigned the most common (major) nucleotide detected in the sequencing reads; generally does not represent within-host polymorphism.

Deep sequencing: next-generation sequencing specifically aimed at producing high sequence coverage (read depth). When applied to a sample containing multiple genomes, can be used to estimate variant allele frequencies within the sample or source population.

Dinucleotide: two contiguous nucleotides (sequence positions) on the same strand of DNA or RNA. Often represented with a 'p' to denote the intervening phosphate group (e.g., TpC).

Divergence (d): the rate of evolutionary substitution (fixation) between lineages, such as HPV types or lineages. Can be estimated separately at sites that are nonsynonymous (d_N) and synonymous (d_S) to detect natural selection, typically long after selection has acted. Contrast nucleotide diversity (π).

Epitope: a molecular pattern that may be recognized as foreign by the host and stimulate an immune response. B cells (antibodies) typically recognize conformational epitopes such as the viral capsid, whereas T cells typically recognize short (e.g., 9–12 amino acid) MHC-bound peptide fragments derived from surveillance of intracellularly translated proteins.

Fitness: generally refers to the reproductive success of a self-replicating entity. Numerous factors influence the fitness of a virus, including genome viability, evasion of immunity, and successful transmission to a new host.

Genetic drift: chance evolution in which allele frequencies fluctuate randomly in a population. Dominates evolution unless overcome by a directional force like natural selection.

Host: an individual organism or population that is infected by a pathogen such as a virus.

Integration: the insertion of full or partial HPV genome sequences into the host (somatic) genome.

iSNV: intrahost single nucleotide variant. Refers specifically to within-host virus polymorphism. By contrast, between-host single nucleotide differences (i.e., different samples or isolates) are often referred to as SNVs or SNPs (single nucleotide polymorphisms). Contrast somatic.

Lineage: an evolutionary line of descent from an ancestor, often visualized as a branch on a tree. In the context of HPV nomenclature, the term refers to distinct groups of related isolates within a single type, denoted by a capitalized letter (e.g., A). HPV lineages typically differ from one another by ~1.0–10.0% at the whole genome level.

Major allele: the most common allele at a given genome position in a sample or population.

Minor allele: the least common allele(s) at a given genome position in a sample or population.

Mutation: a change at one or more nucleotide positions in an individual genome, before the influence of natural selection.

Natural selection: the differential replication success of certain phenotypes. When phenotypes have a genetic basis, selection can shape allele frequencies in a directional manner over time.

Neutral: a nucleotide or amino acid change that produces no change in fitness, whose fate is therefore determined by genetic drift.

Nonsynonymous: a nucleotide change in a protein-coding region that changes the amino acid encoded. More likely than synonymous changes to alter fitness and experience natural selection.

Nucleotide diversity (π): the mean number of differences per site for a randomly chosen pair of sequences in a population. Can be estimated separately at sites that are nonsynonymous (π_N) and synonymous (π_S) to detect natural selection, typically while it is still acting. Contrast divergence (d).

Open reading frame (ORF): a stretch of contiguous codons that begins with a START codon, ends with a STOP codon, and is free of mid-sequence STOP codons. Capable of encoding a complete peptide. Distinct from 'gene', as an ORF may not have its own promoter, may be present in multiple transcripts, and may be co-located in the same transcript with other ORFs (i.e., polycistronic mRNA).

Overlapping ORF: two or more ORFs encoded by the same genome positions, such that distinct protein products are translated from the same nucleotides by using different reading frames. Also called 'overlapping genes' or 'out-of-frame ORFs'.

Positive selection: natural selection that favors an increase in frequency (directional selection) or maintenance at a non-zero frequency (balancing selection) of a particular allele.

Purifying selection: negative natural selection that favors the decrease in frequency and extinction of a particular allele.

Prevalence: the frequency of a virus in a host population.

Quasispecies: a network ('mutant cloud' or 'swarm') of interrelated genotypes produced by very high mutation rates and large population sizes, such that individual genome sequences are unstable.

Somatic: generally refers to body cells in organisms that have a soma vs. germline distinction; in relation to mutations, refers to genetic changes acquired during an individual's lifetime.

Sublineage: an evolutionarily related group nested within a larger lineage. In the context of HPV nomenclature, the term refers to a distinct group of related isolates that form a subset within a single type/lineage, denoted by appending a number to the lineage's capitalized letter (e.g., A1). Sublineages typically differ from one another by 0.5–1.0% at the whole genome level.

Substitution: the evolutionary replacement of one allele by another in a population, resulting in the new allele's fixation (frequency of 100%). Unlike mutation, substitution is the result of evolution after forces like selection have acted. May also refer to point mutations (e.g., single base substitutions).

Synonymous: a nucleotide change in a protein-coding region that does not change the amino acid encoded.

Trinucleotide: three contiguous nucleotides (sequence positions) on the same strand of DNA or RNA; a trimer. Often represented with two 'p' letters to denote the intervening phosphate groups (e.g., TpCpA).

Type: genotype. In HPV genomics, refers to a distinct group of antigenically similar, evolutionarily related viral genomes, denoted with a number (e.g., HPV16). Using current classification criteria, one type differs from all other types by $\geq$ 10% in its L1 nucleotide sequence.

Variant Allele Fraction (VAF): the frequency of a particular allele among all sequencing reads at a given genome position. Refers exclusively to a single sequenced sample. In the case of viral variants, it is an estimate of a variant's allele frequency in the within-host virus population.

Within-host (intrahost): genetic changes occurring within the population of viruses infecting a single host during a single infection, including iSNVs. Contrast between-host and quasispecies.

Within-type (intratypic): genetic changes occurring within one HPV type, with patterns of relatedness categorized as lineages, sublineages, and single nucleotide variants. Includes both between-host and within-host variation.

2. HPV genome and life cycle

2.1. Open reading frames

HPVs have circular, ~7.9 kb double-stranded DNA genomes consisting of an upstream regulatory region (URR), an intergenic noncoding region (NCR) with simple $(AT)_n$ and poly-T repeats, and eight main expressed protein-coding **open reading frames (ORFs)**. The ORFs are named according to their approximate timing of expression during the viral life cycle, where 'E' denotes early and 'L' denotes late: E6, E7, E1, E2, E4, E5, L2, and L1 (listed 5′–3′) (Fig. 1; Fig. 2; Table 1). In addition to the main ORFs, E8 — a sequence often 12 $^2/_3$ codons in length — is spliced to E2 to form E8^E2 at certain stages of infection. All ORFs occupy the sense (forward) strand and are expressed as polycistronic (multi-ORF) mRNAs [17].

E6 and E7 are the primary HPV oncoproteins. In carcinogenic types, E6 and E7 degrade p53 and pRb, respectively [18]. They also interact with numerous other host cell proteins to delay differentiation, promote DNA replication, and evade host immunity [12]. Continued expression of both E6 and E7 is thought to be required for the maintenance of cervical cancer [19,20].

E5 is an accessory oncoprotein that plays a supportive role in, but is not necessary for, oncogenesis [21]. E5s are characterised by high hydrophobicity, transmembrane regions, and downregulation of MHC/HLA (major histocompatibility complex/human leukocyte antigen) class I molecules, thereby disrupting peptide presentation to cytotoxic ($CD8^+$) T cells [22–25]. There are at least four distinct evolutionary groups of E5 ORFs (E5α, E5β, E5γ, E5δ) interspersed among HPV types lacking E5 [26], important to consider for comparative analyses. E5α is the group present in carcinogenic HPVs [22].

E1 (helicase) and E2 (DNA binding protein) are the core viral proteins involved in replication and genome maintenance [27]. Full-length

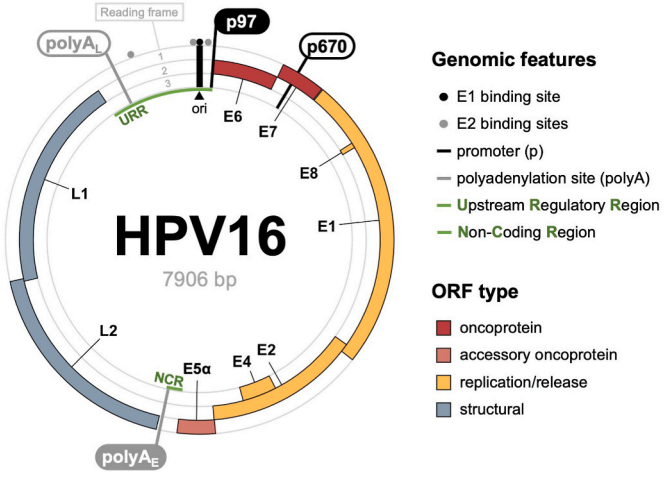

Fig. 1. Human papillomavirus type 16 genome diagram. The circular, ~7.9 kb double-stranded DNA genome of HPV16 is depicted as three sense-strand trinucleotide (codon) reading frames: 1 (outside track), 2 (middle track), and 3 (inside track), where frame 1 begins at position 1 of the genome. Protein-coding open reading frames (ORFs) are depicted as coloured rectangles in the appropriate reading frame. E8 (frame 2) is encoded entirely within E1 (frame 1), and E4 (frame 3) is encoded entirely within E2 (frame 2). E4 is the only ORF occupying frame 3 in HPV16. Early and late promoters (p) are denoted p97 and p670, respectively; early and late polyadenylation sites (polyA) are denoted $polyA_E$ and $polyA_L$, respectively. Black and grey circles denote E1 and E2 binding sites, respectively, where the E1 binding site occurs within the origin of replication (ori) that overlaps position 1. The 3′ terminus of E6 does not overlap the 5′ terminus of E7, in contrast to the overlap observed in non-carcinogenic HPV types. All coordinates correspond to reference genome HPV16REF from PaVE [6,7]. See Table 1 for additional details. Figure made in R [188] (ggplot2; tidyverse; scales; RColorBrewer) and modified in PowerPoint.

E2 tethers virus genomes to host chromosomes for distribution to daughter cells [28,29]. E2 also downregulates E6 and E7 at certain points during the viral life cycle [30]. The shorter E8^E2 splice product is expressed in the basal epithelium to suppress viral replication and maintain low virus copy numbers [31], and this is suggested to play a role in avoiding immune detection [11,32].

E4 is thought to assist in genome amplification and virion release, and is one of the most highly expressed ORFs [33]. Both E8 (38 nucleotides within E1) and E4 (263–284 nucleotides within E2) are out-of-frame **overlapping ORFs**, i.e., their full sequences are encoded in alternative reading frames of other ORFs (Fig. 1; Table 1).

L1 and L2 are the major and minor structural proteins of the viral icosahedral capsid, respectively. Because L1 is generally the most conserved (least variable) ORF, its sequence is used to define HPV types. Specifically, a new HPV type is designated if an isolate's L1 nucleotide sequence differs by $\geq$ 10% from any previously defined type [34]. Nevertheless, L1 does contain five highly variable stretches, ~10–30 codons each, that encode outward-facing loops [35]. These loops contain L1's neutralising antibody **epitopes**, necessary for vaccine-induced immunity [36,37]. Thus, the genetic differences in L1 — used to define different types — correspond to antigenic differences [38], and may reflect **natural selection** for immune escape [9,39].

2.2. Infection

HPV infection of stratified cutaneous and mucosal epithelia (e.g., skin and cervix) is thought to require exposure of long-lived basal (lowermost) cells, including stem or stem-like cells [40]. HPV maintains a stable copy number in this reservoir set of (initially infected) cells, and only later produces infectious virus particles in the upper epithelial layers in coordination with cell differentiation (Fig. 2). Thus, under normal circumstances, no lateral (side-to-side) infection of neighbouring cells occurs in the basal layer; these cells contain virus genomes but not virus particles.

Upon successful infection of the basal layer, viral genomes localise to the nucleus and replicate to a stable number, thought to be an average of ~50–200 copies per cell [41–44]. These genomes persist as virion-free episomes (extrachromosomal circular plasmids) that replicate an average of once per cell cycle [40], but occasionally **integrate** into the host genome [45]. In the basal layer, gene expression remains low, which limits the probability of immune detection [32,40]. However, when daughter cells migrate toward the epithelial surface and differentiate, viral intermediate and late gene expression commences, virus genome copy numbers increase to $>10^3$ [11,44,46,47], and virus particles are formed (Fig. 2). This is all accomplished with no viraemia, no virus-induced cell death, and no inflammation, making the virus practically invisible to the host immune system [46].

Given the above, it is likely that the state and abundance of viral genomes obtained for sequencing depend on the anatomic site and time of sampling. Samples obtained from the epithelial surface during productive infection may include fully viable circular genomes encapsidated within infectious virus particles. On the other hand, samples obtained from cancerous tissue may include partial viral genomes, some or all of which may be integrated into the host cell genome, and which may have incurred deleterious **mutations** or deletions.

2.3. Cancer

HPV viruses replicate their genomes using host polymerases that are normally expressed *before* differentiation, but virion production requires host transcription factors that are expressed *during* differentiation [48]. Both requirements must be met without triggering an immune response that would lead to apoptosis [49]. Strategies used by HPV to achieve these conflicting goals can inadvertently lead to cancer because there is substantial overlap between the cellular functions required for viral success and those which increase the susceptibility of host cells to

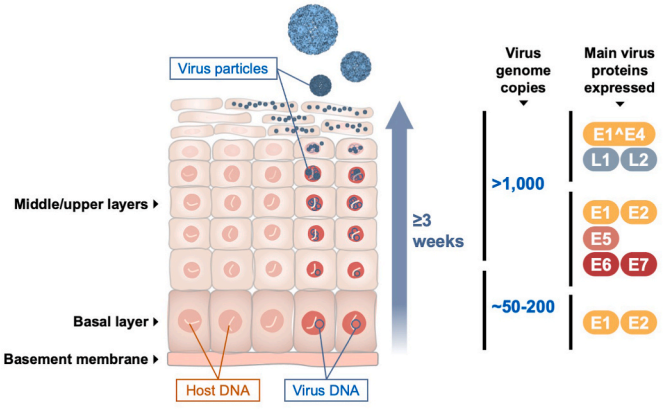

Fig. 2. Life cycle of carcinogenic HPVs in stratified squamous epithelia. Infection is thought to require a microtear exposing the basal (lowermost) layers of the epithelium, which is where host cells susceptible to infection reside. The time required for a basal cell to differentiate and migrate to the epithelial surface is ~3 weeks, placing a lower limit on the length of time required for the viral life cycle [46]. During this time, virus genome copy numbers increase from reservoir levels by at least an order of magnitude. Different viral proteins dominate expression at different levels of the epithelium, in coordination with host cell differentiation. Virus particle formation takes place only in the upper layers, where the capsid proteins L1 and L2 are expressed; no virus particles are formed in the basal layer. Virus genomes are shown as extrachromosomal circular episomes, but integration into the host genome may also occur — effectively ending the virus life cycle by preventing viral genome encapsidation. Figure reflects a synthesis of information presented in the text, primarily refs. [11,41–44,46–48]. Figure made in PowerPoint; virus capsid image modified from Protein Data Bank record 3J6R [189–191].

oncogenesis. Thus, cancer is not the objective of HPV, but rather a 'rare byproduct' [48] that has been termed 'collateral damage' [50] of infection.

Rather than being required for virus propagation, cancer is usually a dead end for HPV: once precancerous lesions form, infectious viral particles are no longer produced [11,12,51]. Thus, cancer — which typically occurs decades after initial infection [52] — does not contribute to viral evolutionary **fitness**. Additionally, HPV-associated 'driver' mutations, most notably integration events, may themselves bring an end to the viral life cycle. Most sequencing methodologies do not distinguish between HPV genomes that are viable or nonviable; or between HPV genomes that exist as virion-encapsidated copies (ready to be transmitted), free episomal copies (may transmit if encapsidated), or integrated copies (unlikely to transmit) [53,54]. Critically, such factors determine how viral variation may be interpreted, e.g., mutations in integrated HPV copies are unlikely to experience onward transmission and contribute to viral evolution.

3. HPV evolution and diversity

3.1. Fitness

Evolutionary fitness refers to reproductive success. Because viruses are self-replicating entities with a genotype/phenotype connection, they can undergo evolution via natural selection to maximise their fitness. The fitness of an HPV type can be estimated by its **prevalence** (frequency in the host population), which is itself a function of persistence (length of productive infection) and incidence (rate of successful transmission to new hosts) [16,50,55]. However, it is important to note that prevalence may also be influenced by chance factors, e.g., a founder effect in which a given viral genotype happens to enter a host population at an earlier date than other genotype(s).

HPV16 is both the most prevalent carcinogenic HPV type and the most prevalent type in cancer. This implies that the replication strategies it employs (or niches it occupies) potentiate cancer. However, some non-carcinogenic types are more or equally prevalent in the general population [56] and likely have even higher fitness than carcinogenic types. Thus, an HPV type may have high fitness without causing cancer — as is true of most viruses.

3.2. Mutation

At least four distinct mechanisms give rise to HPV mutations at different stages of its life cycle. In the basal epithelial layer, copy numbers are maintained using bidirectional replication, which may disproportionately lead to mutation and recombination in the region between E2 and L2 where replication forks meet [26,32,57,58]. Second, when viral copy numbers increase in differentiating cells destined for the surface, HPV switches to unidirectional (rolling circle or recombination-dependent) replication [27,32,57], which may involve distinct mutational processes. Third, when DNA enters the single-stranded state during either transcription or replication, host APOBEC3 enzymes can target TpC **dinucleotides** to induce C→T (G→A) mutations [59] (section 4.4.1). Finally, deamination of methylated CpG dinucleotides, which also occurs in the single-stranded state [60], may also cause C→T (G→A) HPV mutations [61,62].

Because HPV genomes use host DNA polymerases to replicate, they have low mutation rates. Direct estimates of the HPV mutation rate are hindered by the difficulty of growing HPV in cell culture [63–65] and its high replication fidelity. Thus, evolutionary comparisons are used, where mutation rates can be inferred from **substitution** rates. Across the whole papillomavirus genome, evolutionary substitution rates are ~5 times higher than in the genomes of their mammalian hosts [66]. However, mutation and substitution rates are equal only at sites that are **neutral**, i.e., lack functional constraint and therefore evolve predominantly by random **genetic drift** rather than natural selection [67,68].

Two candidates for neutral sites in HPV are the upstream regulatory region (URR) and **synonymous** positions in protein-coding regions. A comparison between the URR of HPV18 and HPV45 yields a single nucleotide substitution rate of $\sim 4.5 \times 10^{-7}$ per site per year [69]. Similarly, an analysis of feline papillomaviruses yields a rate of $\sim 2.69 \times 10^{-8}$ per site per year [70]. Assuming URR neutrality, these serve as estimates of the papillomavirus mutation rate per unit time. However, because the URR encodes regulatory elements that make it subject to **purifying selection**, even these are likely to be underestimates. To our knowledge, no estimates based on synonymous protein-coding sites are available.

The above HPV mutation rate estimates are >1000 times lower than those of RNA viruses ($\sim 10^{-4}$ to 10^{-3} per site per year estimated from synonymous sites [71,72]), but ~500 times higher than the human germline mutation rate ($\sim 4.27 \times 10^{-10}$ per site per year estimated from father/mother/child trios [73]). Numerous factors contribute to these differences, including 1) different generation times; 2) different numbers of genome replications per generation; 3) selection acting on sites assumed to be neutral; 4) acute vs. persistent life cycles, along with any associated latency or replication throughout time; 5) mutagenesis of viral genomes by host enzymes such as APOBEC3; and 6) the enzymes and specific activities involved in DNA replication.

3.3. Nucleotide diversity (π)

Nucleotide diversity (π) [74] is an unbiased metric ideal for measuring the diversity of virus populations [75]. In protein-coding regions, a significant difference between π at **nonsynonymous** (π_N) and **synonymous** (π_S) sites is evidence for ongoing **positive** ($\pi_N/\pi_S > 1$) or **purifying** ($\pi_N/\pi_S < 1$) **selection** [76,77]. Within-population selection is expected to influence substitution rates and therefore **divergence** (d; d_N/d_S) among HPV **lineages** and types over time [78,79]. Thus, π_N/π_S and d_N/d_S are routinely used for detecting functionally important

Table 1
Human papillomavirus open reading frames in three closely related carcinogenic types: HPV16, HPV31, and HPV35.

ORF (5′–3′)	Key protein functions	Key references	Type	Reading frame[a]	Amino acid length	Nucleotide length	CDS start	CDS end	Overlapping ORFs (overlap type)	Overlapping nucleotides (%)
E6	Oncoprotein; targets p53 for degradation; offsets E7's antiviral effects by blocking apoptosis; necessary for maintaining cancer	Vande Pol and Klingelhutz 2013; Vats et al., 2021	HPV16	2	151	456	104[b]	559	–	0
			HPV31	3	149	450	108	557	–	0
			HPV35	2	149	450	110	559	–	0
E7	Oncoprotein; targets pRb for degradation; increases DNA replication; stabilises APOBEC; necessary for maintaining cancer	Roman and Munger 2013; Vats et al., 2021	HPV16	1	98	297	562	858	–	0
			HPV31	2	98	297	560	856	–	0
			HPV35	1	99	300	562	861	–	0
E1	Helicase; essential for replication; interacts with host replication factors; the only HPV enzyme	Bergvall et al., 2013	HPV16	1	649	1950	865	2814	E8 (internal), E2 (terminal)	97 (5%)
			HPV31	1	629	1890	862	2751	E8 (internal), E2 (terminal)	97 (5%)
			HPV35	1	637	1914	868	2781	E8 (internal), E2 (terminal)	103 (5%)
E8 (E8^E2)	Suppresses viral replication in the basal epithelium; spliced to E2 to form E8^E2	McBride 2013; Kuehner and Stubenrauch 2022	HPV16	2	12 $^{2}/_{3}$[c]	38	1265	1302	E1 (full)	38 (100%)
			HPV31	2	12 $^{2}/_{3}$[c]	38	1259	1296	E1 (full)	38 (100%)
			HPV35	2	12 $^{2}/_{3}$[c]	38	1268	1305	E1 (full)	38 (100%)
E2	DNA binding protein; downregulates E6 and E7; partitions viral genomes to daughter cells upon division	McBride 2013; Kuehner and Stubenrauch 2022	HPV16	2	365	1098	2756	3853	E1 (terminal), E4 (internal), E5 (terminal)	326 (30%)
			HPV31	2	372	1119	2693	3811	E1 (terminal), E4 (internal)	343 (31%)
			HPV35	2	366	1101	2717	3817	E1 (terminal), E4 (internal), E5 (terminal)	335 (30%)
E4 (E1^E4)	May assist in virus synthesis and release by disrupting cellular keratin in the upper epithelium; highly expressed biomarker of infection; lacks a conserved start codon; usually spliced to E1 (E1^E4)	Doorbar 2013	HPV16	3[d]	86 $^{2}/_{3}$[d]	263	3358	3620	E2 (full)	263 (100%)
			HPV31	3[d]	93 $^{2}/_{3}$[d]	284	3295	3578	E2 (full)	284 (100%)
			HPV35	3[d]	87 $^{2}/_{3}$[d]	266	3319	3584	E2 (full)	266 (100%)
E5α	Accessory oncoprotein; hydrophobic transmembrane protein; downregulates MHC expression and disrupts presentation of virus T-cell epitopes	DiMiao and Petti 2013; Willemsen et al., 2019	HPV16	1	83	252	3850	4101	E2 (terminal)	4 (2%)
			HPV31	3	84	255	3816	4070	–	0
			HPV35	1	83	252	3814	4065	E2 (terminal)	4 (2%)
L2	Minor capsid; guides virus genomes to nucleus upon infection; up to 72 copies per virus particle	Wang and Roden 2013	HPV16	1	473	1422	4237	5658	L1 (terminal)	20 (1%)
			HPV31	1	466	1401	4171	5571	L1 (terminal)	20 (1%)
			HPV35	2	469	1410	4211	5620	L1 (terminal)	20 (1%)
L1	Major capsid; mediates virus attachment and entry; 360 copies per virus particle; self-assembles into virus-like particles (VLPs) used for vaccines	Buck et al., 2013	HPV16	2	505	1518	5639	7156	L2 (terminal)	20 (1%)
			HPV31	2	504	1515	5552	7066	L2 (terminal)	20 (1%)
			HPV35	3	502	1509	5601	7109	L2 (terminal)	20 (1%)

ORF lengths and positions are given for HPV reference genomes found at PaVE: HPV16REF (7906 bp), HPV31REF (7912 bp), and HPV35REF (7879 bp) [6,7]. Overlapping ORFs refer to out-of-frame protein-coding ORFs.

[a] Reading frames refer to codons occupying the trinucleotides (codons) starting at positions 1, 2, and 3 of the reference genome for each type.

[b] In HPV16, E6 is sometimes annotated as beginning at position 83 of the genome, i.e., 7 additional codons at its 5′ terminus (start); for consistency, we instead employ numbering based on the start site annotated in PaVE.

[c] E8 encodes 12 codons, plus the first two nucleotides of a one additional codon at its 3′ terminus (end). The final nucleotide of the additional codon is spliced from, and maintains the reading frame of, E2.

[d] E4 encodes the end (3′ portion) of E1^E4, beginning with the last 2 nucleotides of a codon; the first nucleotide of the additional codon is spliced from E1.

genome regions (Fig. 3).

Standard methods for estimating π_N/π_S and d_N/d_S are not applicable to overlapping ORFs, such as E2 and E4 in HPV genomes. Because the genome positions encoding E4 also encode E2, the corresponding nucleotides are subject to selective constraints acting on both proteins. Specifically, because the majority of random mutations are non-synonymous, synonymous changes in E2 are likely to be non-synonymous in E4 — and therefore subject to purifying selection. As a consequence, standard π_N/π_S methods [76] tend to underestimate π_S (overestimate π_N/π_S) in such regions, leading to a spurious inference of positive selection [80,81]. Overlapping ORFs therefore require more sophisticated π_N/π_S methods that account for a variant's effects in two proteins [82].

In HPV16, the region of E2 overlapping E4 exhibits $\pi_N/\pi_S > 1$ when analysed using standard methods, suggestive of positive selection (Fig. 3). This has been attributed to the presence of B and T cell epitopes in E2 [83,84], which may experience selection for immune escape. As an alternative explanation, E4 is very highly expressed [33,40] and does

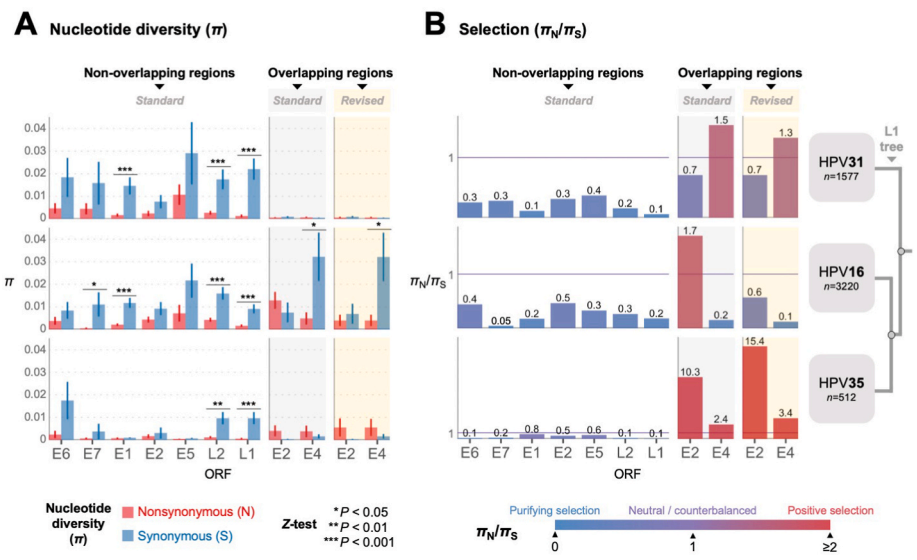

Fig. 3. Human papillomavirus nucleotide diversity and natural selection in three closely related carcinogenic types: HPV16, HPV31, and HPV35. (A) Nonsynonymous (amino acid changing; π_N) and synonymous (not amino acid changing; π_S) nucleotide diversities were calculated as the mean number of pairwise differences per site [74] using SNPGenie [192] based on whole genome consensus sequences (one representative sequence per sample) for HPV16 (n = 3220; [113]), HPV31 (n = 1577; [114]), and HPV35 (n = 512; [115]). Sequences were derived from samples obtained from the NCI-Kaiser Permanente Persistence and Progression (PaP) cohort. The null hypothesis of $\pi_N = \pi_S$ was evaluated using a Z-test (1000 bootstrap replicates, codon unit) [193]. Significance is indicated as $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. **(B)** The ratio of π_N to π_S can provide evidence for positive selection ($\pi_N/\pi_S > 1$) or purifying selection ($\pi_N/\pi_S < 1$). 'Overlapping regions' refers to protein-coding sites that overlap a second protein-coding ORF in another reading frame. Results are only shown for the E2/E4 overlap (~24% of E2 and 100% of E4); sites involved in shorter overlaps yielded highly variable estimates and were excluded (E1/E8, E1/E2, E2/E5, L2/L1). Standard π_N/π_S and d_N/d_S methods were used to analyse non-overlapping regions [76,192]. Revised methods that account for a variant's effects in two proteins were used to analyse overlapping regions, specifically by limiting to sites that are nonsynonymous in the overlapping frame, i.e., the π_{NN}/π_{SN} ratio in OLGenie [81,82]. Positive selection is not significant for any whole ORF, a result that may reflect a counterbalance between sites under positive and purifying selection. The tree topology is that inferred from the L1 ORF (PaVE [6,7]). Figure made in R [188] (ggplot2; tidyverse; scales; RColorBrewer) and modified in PowerPoint. Source data: Supplementary File 1.

not match human codon usage preferences [85], suggesting it may be subject to especially strong purifying selection [86]. Using publicly available HPV16 sequence data and a π_N/π_S method developed for overlapping ORFs [81], we show that E4 maintains a strong signal of purifying selection, whereas the π_N/π_S ratio of E2 drops from 1.7 to 0.6 when limiting to sites that are nonsynonymous in E4 (Fig. 3B). Other evidence indicates that amino acid changes in E2 tend to be tolerated only when they occur in such a way as to produce synonymous changes in E4 [87,88]. Taken together, these results suggest that the functional constraint of E4 outweighs positive selection on E2.

3.4. Recombination

Recombination can produce new combinations of pre-existing mutations, potentially linking adaptive (or maladaptive) variants in the same genome. Although some evidence exists for recombination in HPV [89], it is thought to be very rare. Recent HPV16 genome sequence data were suggested to provide evidence of recombination [90], but the observed patterns were subsequently attributed to co-infection by multiple **sublineages** of the same HPV type — events that can be hard to distinguish given a **consensus** sequence alone.

One obstacle to detecting recombination is that it requires enough dissimilarity between two sequences to infer a breakpoint and rule out sequencing error. Because mutation during the course of a single HPV infection is unlikely to introduce sufficient variation, detectable (and biologically meaningful) recombination would likely require co- or super-infection of the same basal cell by distinct viral types, lineages, or variants. This is expected to be rare. Nevertheless, important recombination events may have occurred at key moments in HPV evolution. Most notably, evolutionary trees inferred from early (E) ORFs cluster carcinogenic HPV species together, whereas those inferred from late (L) ORFs do not (Fig. 4) (section 4.2). This suggests a recombination event between the E5 and L2 ORFs near the root of the *Alphapapillomavirus* genus [91]. However, convergent evolution cannot be ruled out to explain this pattern.

3.5. Natural selection and immunity

Mutations that affect viral persistence and transmission are subject to natural selection. One important selective pressure is host immunity [92,93]. For HPV, it is thought that B cells (antibodies) contribute primarily to the prevention of infection, made possible by the long lag time between virion binding and eventual cell entry [94]. On the other hand, T cells contribute primarily to the control of already-established infections, in part through the presentation of viral epitopes by the MHC [47,95]. Because MHC presentation is determined by an individual's human leukocyte antigen (HLA) genotype [96], hosts likely differ in their ability to control infection for a given viral type, lineage, or variant. Indeed, genome-wide association studies have pointed to specific HLA class I (presentation to $CD8^+$ T cells) and class II (presentation to $CD4^+$ T cells) alleles that confer higher or lower risk of cervical cancer [97–102], including in an HPV type- or variant-specific manner [97–99].

Within a host, HPV genomes are divided into small 'islands' — distinct subpopulations infecting distinct cells. This limits the power of natural selection, because variants in isolated compartments are subject to chance extinction [103], e.g., via host cell death. Nevertheless, **within-host (intrahost)** selection may still occur between genomes infecting the same basal cell, or between genomes infecting different basal cells. Within the same basal cell, selection must be very weak owing to low copy numbers and lack of a genotype/phenotype connection (i.e., no virus particles). Between cells, there is opportunity for HPV genomes to compete via group selection. Specifically, if a virus mutation confers a growth advantage to its host cell, all genomes in the cell will benefit. As a result of such cell-to-cell competition, certain virus genotypes may drive others to extinction as their host cells replace one another. This selection among **somatic** cells may allow viral genome persistence within the host, potentiating clonal expansion and progression to cancer in rare instances [104].

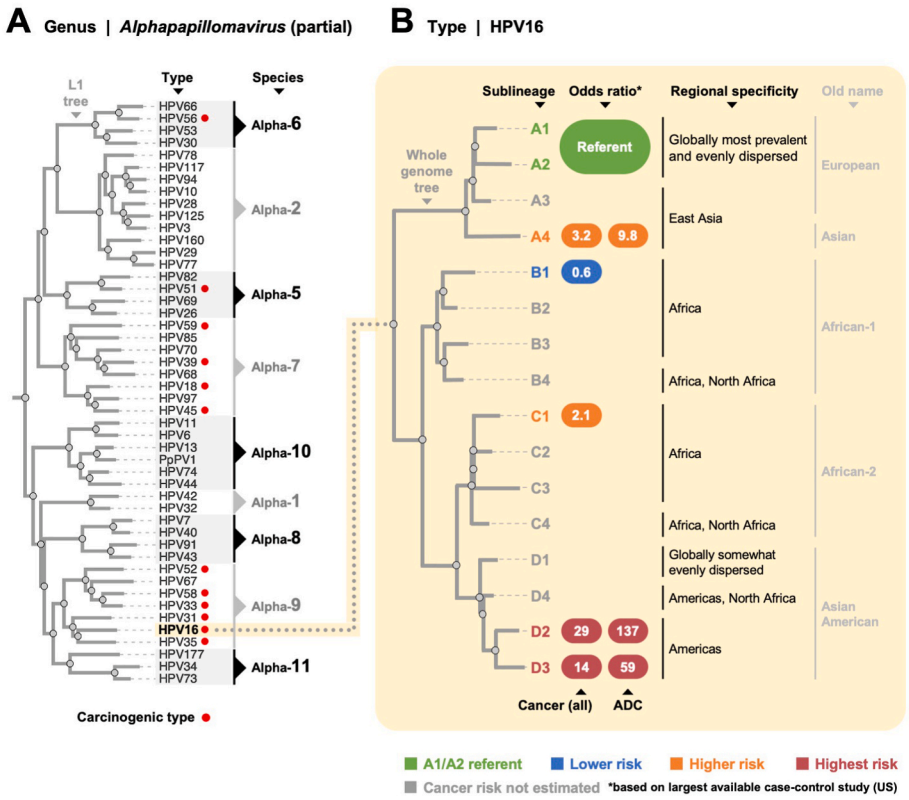

Fig. 4. Evolutionary relationships between and within carcinogenic HPV types. Trees of multiple types are typically inferred using the L1 ORF, reflecting how types are classified, whereas trees of within-type (intratypic) variation are inferred using whole genomes. **(A)** Subtree of the *Alphapapillomavirus* genus including all carcinogenic species (Alpha-5, -6, -7, and -9) and types (red dots), as determined by the L1 ORF (modified from PaVE [6,7]). Note that trees inferred from the early (E) ORFs would instead place the carcinogenic species into one clade [91, 194], possibly due to convergence or recombination early in *Alphapapillomavirus* evolution. **(B)** Lineages (A, B, C, D) and sublineages (A1-4, B1-4, C1-4, D1-4) of HPV16, as reported in Ref. [50]. Key characteristics as determined by HPV genomic and epidemiologic data are noted. Odds ratio estimates for cancer are shown for specific sublineages compared to the most common A1/A2 sublineages, as reported in Mirabello et al. [149] using data from a large U.S. case-control study (colour denotes risk). The odds ratio for B1 is reported for precancer/cancer for statistical power (small sample size). Old sublineage names are shown for reference, but their use is discouraged. ADC = adenocarcinoma. Figure made in PowerPoint. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.6. *Interpreting diversity*

Not all evolutionary change in the HPV genome benefits the virus. As mentioned in section 2.3, infectious virions are not produced in cancer tissues [11]. This implies that many viral functions will freely accumulate mutations in cancers — even if they would render normal virus nonviable. Such *relaxation of purifying selection* may apply to specific genome positions, specific protein residues, or even whole ORFs. Furthermore, because each peptide produced by the virus may potentially encode an epitope that stimulates an immune response, unnecessary protein products may constitute an 'antigenic liability' for the virus and be selected against.

There is a salient historical example of the relaxation of purifying selection: HPV vaccines rely on self-assembling L1 proteins to form virus-like particles (VLPs). However, the first attempts to generate VLPs *in vitro* failed to achieve full-size particles [105] or high yield [106]. It was soon recognized that the L1 sequence being used had been obtained from a cervical cancer, and that this sequence differed from wild type (non-mutated) infectious virions by one amino acid (H202D). When the mutated L1 was reverted to its wild type, full-size VLPs were generated with a 10^3-fold increase in yield [107]. Thus, relaxation of purifying selection had allowed a deleterious nonsynonymous mutation in L1 to freely accumulate in the cancer tissue.

In summary, the HPV life cycle informs interpretation of evolutionary genomics data, and *vice versa*.

4. Key HPV genomics advances

4.1. *HPV whole genome sequencing*

Sanger sequencing yields high-quality HPV genome sequences but is slow, costly, and labour-intensive. As a consequence, only ~100 HPV16 whole genomes were publicly available by the early 2010s [108,109]. Additionally, Sanger's dependency on primers limits its application to known HPV types, and it is not amenable to studying within-host variation, i.e., **minor alleles** of HPV within a single host. Nevertheless, it remains the 'gold standard' and continues to produce important insights [110,111].

Next-generation sequencing (NGS) approaches have potentiated an enormous jump in the number of HPV whole genomes available. An amplicon-based Ion Torrent assay introduced in 2015 [112] is responsible for most of this increase, providing over 5000 HPV16 genomes by 2017 [113], as well as large numbers for other carcinogenic types including HPV31 [114] and HPV35 [115].

Ion Torrent yields fewer single base errors than Illumina, but more indel errors due to the difficulty of sequencing homopolymers (e.g., AAAAA) [116,117]. Improvements to Ion chemistry (Hi-Q) and chips yield single base substitution error rates of only ~0.000129 per base [116], compared to first generation error rates of ~0.00431–0.0110 per base [117]. Nevertheless, even using the older chemistry, the HPV Ion Torrent sequencing assay shows 99.97% (standard deviation [sd] = 0.07%) concordance between sample duplicates, as well as 99.97% (sd = 0.13%) concordance with Sanger sequencing [112]. This compares favourably to concordance between Illumina and Sanger, which has been reported at >99.8% for HPV [110]. Further, whereas ~80% of HPV16 isolates from different women have ≥2 nucleotide differences, ~80% of isolates from the same woman have ≤1 differences (72% are identical) using the Ion assay, confirming its robustness [113]. In fact, the quality of Ion Torrent data is likely sufficient to allow deep (high-coverage) sequencing for detection of within-host HPV polymorphisms (see section 4.4).

Another NGS approach uses full-circle PCR followed by sequencing with Illumina [53,54]. This technology has yielded hundreds of genomes to date, and gives a low error rate of ~0.000076 per base; it is typically used to deep sequence within-host samples [54]. Shotgun metagenomics with Illumina sequencing has also been used to reveal hundreds of new skin HPVs in immunodeficient individuals, approximately doubling the number of known HPV types in recent years [118,119].

4.2. Between-type (intertypic) HPV divergence

Classification of papillomaviruses (family *Papillomaviridae*) follows guidelines from the International Committee on the Taxonomy of Viruses (ICTV) [120]. The current criteria rely on an empirical distribution of pairwise L1 nucleotide sequence identity between HPV isolates that suggests natural groupings into the same genus (>60% identity), species (>70% identity), and type (>90% identity) [120]. However, it was recently noted that bias may have been introduced into this distribution by an overrepresentation of types from the *Alphapapillomavirus* genus, and that these groupings may not hold up for recent genome data (see Ref. [121] for more details).

Five genera contain HPVs: *Alpha-*, *Beta-*, *Gamma-*, *Mu-*, and *Nu-papillomavirus*. These genera roughly reflect tissue tropism, e.g., *Alphapapillomavirus* members tend to infect mucosal/genital epithelia, while *Betapapillomavirus* members tend to infect cutaneous/skin epithelia [40,122], with exceptions [9,61]. All 12 carcinogenic HPV types are present in the *Alphapapillomavirus* genus and are limited to four species: Alpha-5 (HPV51), Alpha-6 (HPV56), Alpha-7 (HPV18, 39, 45, and 59), and Alpha-9 (HPV16, 31, 33, 35, 52, and 58). Because types are substantially diverged (i.e., ≥10% nucleotide difference in L1), immunity to one type offers only limited cross-immunity to closely related types (e.g., HPV16 and HPV31) [47]. Phylogenetic trees built on early ORFs (E6, E7, E1, E2, E5) cluster the four carcinogenic species (Alpha-5, -6, -7, and -9) with a single carcinogenic ancestor [26,55], while trees built on just the late ORFs (L2, L1) instead yield two separate carcinogenic clusters (Alpha-9 vs. Alpha-5/6/7) [91] (Fig. 4). The genetic changes underlying this phylogenetic incongruence (different tree topologies) are concentrated in E6 and L2 (5′-terminal portion), suggesting important **between-type (intertypic)** differences may fall in these regions. Interestingly, each carcinogenic species contains at least one type that does not cause cancer (e.g., HPV67 in Alpha-9; but see Ref. [123]).

It has been noted since the discovery of HPV16 that a type's prevalence differs by geography [124]. For example, globally, HPV16 accounts for ~60% of cervical cancers while one of its closest relatives, HPV35, accounts for only ~2% [13,14]. However, HPV35 is especially prevalent in women with African ancestry, where it accounts for up to 4.9–10.4% of cancers [115,125–129]. Interestingly, HPV35 exhibits low π values in the early (E) genes but extremely high (albeit insignificant) π_N/π_S ratios in E2 and E4 (Fig. 3) — a striking difference from its HPV16 and HPV31 relatives. HPV35 is not included in any current vaccines, and its addition might confer better protection than relying on cross-protection from HPV16 and HPV31.

HPV types also differ in their frequencies of integration into the host genome (see section 4.5). Although integration is observed in ~83% of HPV-positive cervical cancers overall, it is seen in only ~76% of cervical cancers caused by HPV16 but virtually all those caused by HPV18 [130]. Thus, data from evolutionary, epidemiologic, and molecular studies imply that the precise mechanisms of infection and carcinogenesis may differ even among closely related types.

Key features specific to carcinogenic HPV types include (1) the ability to degrade p53 and pRb members; (2) regulation of E6 and E7 expression through differential mRNA splicing rather than separate promoters; (3) the ability to immortalise keratinocytes in cell culture; and (4) a propensity for dysregulated gene expression (reviewed in Refs. [9,18,40,131]). Recently, it has been further noted that, in non-carcinogenic types, the end of E6 overlaps the beginning of E7, similar to other ORFs that overlap at their termini (Fig. 1; Table 1). In contrast, in carcinogenic types, these ORFs no longer overlap due to an insertion that has extended the end of E6 [132]. This region of E6 encodes the protein's PDZ binding motif, which is central to numerous host protein interactions [18]. Thus, the genomic decoupling of E6 and E7 in this region may have potentiated oncogenesis and deserves further attention.

4.3. Within-type (intratypic) HPV diversity

The recent explosion of HPV whole genome sequences has allowed genetic variation within each HPV type to be analysed in unprecedented detail. The simplest way to study this **within-type (intratypic)** diversity is the consensus sequence approach, i.e., one representative HPV sequence per sample. Comparing consensus sequences **between hosts (interhost)** has worked particularly well for studying within-type HPV diversity owing to its low mutation rate. However, it is important to recognize that each host is infected not by a single virus but by a population of viruses, within which consequential variation may exist (see section 4.4).

4.3.1. Lineages, sublineages, and SNPs

Classification of HPV sequences within a type employs an alpha-numeric nomenclature. At the highest level, lineages differ from one another by ~1.0–10% across the whole genome and are denoted with an uppercase letter (e.g., A). Within each lineage, sublineages differ from one another by ~0.5–1.0% and are further denoted with a number (e.g., A1) [34]. For example, HPV16 has a total of four lineages (A, B, C, D) that are divided into 16 sublineages (A1-4, B1-4, C1-4, D1-4) (Fig. 4). Lower levels of classification (e.g., A1.1) have not yet been utilised. The reference sequence for a type is preferentially assigned to lineage A or, if defined, sublineage A1 (e.g., HPV16REF in A1) [34]. Note that lineages and sublineages are best classified using the whole genome rather than L1, because L1 is not sufficiently variable to resolve within-type differences [15,34].

4.3.1.1. HPV16. For HPV16, the existence, geographic clustering, and potential clinical importance of within-type sequence variation has been recognized since at least 1991 [133]. A1 is by far the most common sublineage and is relatively evenly dispersed across the globe. Other sublineages exhibit sometimes extreme clustering by geographic region, often being prevalent where they evolved, most notably A3 and A4 in East Asia; B1-4 and C1-4 in Africa; D2 and D3 in the Americas; and B4, C4, and D4 in North Africa [134,135]. Remarkably, this peculiar distribution is due at least in part to an ancient host split ~500 thousand years ago, when the lineage giving rise to A was carried by the Neanderthals/Denisovans, and BCD by the ancestors of modern humans. After a period of separation, the A lineage was then sexually transmitted to modern humans — at the same time as introgression of host nuclear alleles [136,137].

Before the availability of large numbers of HPV16 whole genomes, it was necessary to group the rarer BCD (previously 'non-European') lineages together for statistical power (Fig. 4B). These earlier pioneering studies revealed an increased risk of cancer for BCD compared to the A lineage [138–148]. Since that time, more fine-scale evaluations of individual sublineages have become possible with the availability of large numbers of cervical samples for sequencing, e.g., the exfoliated cervical cell samples from the Kaiser Permanente Northern California PaP (Persistence and Progression) cohort [149].

Compared to the most common sublineages (A1 and A2, reference), certain sublineages were shown to be significantly associated with increased risks of cervical precancer and cancer: A4 (odds ratio [OR] for cancer = 3.2), C1 (OR = 2.1), D2 (OR = 28.5), and D3 (OR = 13.9) (Fig. 4B). In contrast, D1 and D4 are not associated with precancer/cancer, and B1 is significantly associated with a lower risk of precancer/cancer (OR = 0.6) [149]. Sublineage risks of precancer and cancer also vary by histologic subtype [147,149–153] (but see Refs. [154,155]). This was most strikingly observed in the U.S., with a strong increased risk of adenocarcinoma conferred by A4 (OR = 9.8), D2 (OR = 137.3), and D3 (OR = 59.5), as compared to A1/A2 [149] (Fig. 4B).

Precancer and cancer risks associated with sublineages have also been shown to be influenced by host race/ethnicity [140,156] (but see Refs. [157,158]). Specifically, results suggest that precancer/cancer risk

is highest when there is a match between a woman's self-reported race/ethnicity and the ancestry in which the infecting sublineage evolved: A1/A2 in whites; A4 in Asians; and D2/D3 in Hispanics [149]. Similarly, there are increased cancer risks for A3, A4, and D sublineages in regions where they are common: A3 in East Asia (OR = 2.2); A4 in East Asia (OR = 6.6) and North America (OR = 3.8); and D in North America (OR = 6.2), where D sublineages are also more frequent in adenocarcinoma [134].

The aforementioned risk differences are particularly remarkable given the relatively small number of genetic differences between the sublineages: in HPV16, the A4 and D2/D3 sublineages differ from A1 by only ~60 and ~150 nucleotides, respectively. Early phylogeny-based d_N/d_S analyses identified evidence for positive selection in E6 [159–161], E5 [159], and L2 [161], suggesting particular codons as candidates for important functional differences between sublineages. More recently, two single nucleotide polymorphisms (SNPs) in the URR were shown to greatly reduce risk of precancer/cancer (ORs ≤0.06) [113]. Additionally, individual SNPs have been linked to differences in HPV16-driven oropharyngeal cancer survival: patients with ≥1 high-risk HPV16 SNP had a median survival of only 4 years compared to 19 years for patients without these SNPs [162]. HPV genetic variation has not been evaluated related to cervical cancer prognosis.

4.3.1.2. HPV18. Although HPV18 is the second most common type associated with cancer, much less is known about the relationship between its genetic variation and risk of precancer/cancer. It is less prevalent than HPV16, less commonly detected in precancerous lesions, and found in more adenocarcinoma than squamous cell carcinoma, which has likely limited its thorough study.

HPV18 can be classified into three main lineages (A, B, C) and nine sublineages (A1-A5, B1–B3, C). Two small studies suggest that variants may be differentially associated with adenocarcinoma [150,163], but others do not [164]. A d_N/d_S analysis identified evidence for positive selection in E5 [165]. A worldwide study of HPV18 lineages/sublineages found no major differences in the distribution of lineages between cancer-free controls and cancer cases or histologies; however, when stratified by geographic region, they observed that the A1 sublineage is associated with more cancer in Eastern Asia [163]. More studies are needed to understand the role of genetic variation in HPV18-related disease.

4.3.1.3. HPV31. HPV31 has three lineages (A, B, C) that are divided into eight sublineages (A1-2, B1-2, C1-4). Early studies showed that lineages A/B are associated with precancer compared to C [141,166]. A large analysis of 2093 genomes has since revealed that the A1 (OR = 1.7), A2 (OR = 2.5), and B2 (OR = 1.9) sublineages confer higher risk of precancer/cancer than C3 (most common sublineage) [114]. In addition, a single nonsynonymous change in E7 (H23Y) was shown to increase risk of precancer/cancer (OR = 1.6) [114].

4.3.1.4. HPV35. With only two lineages (A, B) and three sublineages (A1, A2, B), HPV35 has less genetic variation and fewer lineages/sublineages than most other carcinogenic types, including its sibling types HPV16 and HPV31 (Fig. 3). One early study suggested that the A1 sublineage is associated with elevated risk of precancer compared to A2 [141]. A subsequent large analysis of 1053 HPV35 genomes has further revealed important differences in risk associated with viral variation by host race/ethnicity. The A2 sublineage confers higher risk (OR = 5.6) of precancer/cancer specifically in African American women, but not other racial/ethnic groups, compared to A1 (most common sublineage) [115]. Consistent with this, A2 is more prevalent among cancers in Africa compared to other world geographic regions [115]. Further, 12 SNPs are associated with precancer/cancer only in women of African ancestry, and women with two or more of these individual SNPs have a strong increased precancer/cancer risk (OR = 69) [115].

4.3.2. E7 constraint in HPV16

Although the genetic basis of HPV16's unique carcinogenicity is far from understood, comparisons between precancer/cancer cases and cancer-free controls in large studies point to E7 as a key factor. Specifically, examination of 5328 HPV16 consensus genomes shows cancers are characterized by significantly fewer nonsynonymous variants than controls, evidenced by a low odds ratio of 0.16 and a ~5.6-fold reduction in π_N/π_S [113]. An *in vitro* study of these E7 variants showed that the specific variants observed in the controls lead to a reduced level of E7 protein and lower transforming activity [167]. This suggests that E7 may exist in a damaged state in controls, reducing carcinogenicity. However, while E7 conservation appears to be critical for the carcinogenicity of HPV16, this is not necessarily true of other types, e.g., HPV31 [114]. Of note, the increased variation among controls is unlikely to be due to the production of virion early in infection, which would affect all ORFs equally and require new HPV variants to reach high frequencies (i.e., become **major alleles**) specifically in controls but not cancers.

Although the elevation of nonsynonymous changes in HPV16 controls compared to cases is most pronounced in E7, elevation is also observed in E1 and L1 and somewhat in most ORFs [113]. It is possible that specific amino acid changes may promote viral clearance. Such changes could represent random mutations during replication (see section 3.2), but could also represent an antiviral mechanism, namely the mutagenic activity of human APOBEC3 cytidine deaminases (see section 4.4.1). Specifically, consensus-level nonsynonymous differences in E7 are enriched for C→T at TpC dinucleotides (i.e., TpC→TpT) in controls, a change consistent with APOBEC3 activity [168]. Further, HPV genomes exhibit an overall depletion of TpC, particularly at third codon positions where they would have been most likely to cause tolerable synonymous changes [61,62,169,170]. At those TpC sites that remain, the vast majority of possible C→T changes are nonsynonymous [168]. Thus, synonymous APOBEC3 changes have largely been saturated in the HPV genome. Finally, within HPV16, the D2/D3 sublineage has the fewest remaining TpC sites — as a result of having the largest proportion of TpC→TpT changes in its evolutionary history — compared to A1/A2 sublineages [168]. These changes may have contributed to the lower fitness (prevalence) but enhanced carcinogenicity of D2/D3.

4.3.3. Key considerations for evaluating genetic risk associations

As larger studies are published, is it clear that the grouping of disease outcomes (e.g., precancer and cancer; squamous and glandular lesions) and sublineages (e.g., BCD in HPV16) in smaller studies can conceal important qualitative heterogeneity in lineages and disease outcomes, and mask specific associations. These findings raise the exciting prospect that, as whole genome HPV sequences continue to accumulate, we may gain sufficient resolution to pinpoint more specific variants or combinations of variants that modulate cancer risk even below the sublineage level.

In summary, when evaluating viral genetics and precancer/cancer risk, it is important to consider the genetic variation that exists within individual HPV types with respect to geographic distribution, host race/ethnicity, and histologic subtypes (squamous cell carcinoma vs. adenocarcinoma). For HPV16, findings to date imply that sublineages have adapted to the niches (tissues and cell types) and populations in which they historically evolved — likely including strategies for avoiding immune clearance.

4.4. Within-host (intrahost) HPV diversity

Fine-scale analysis is required to study HPV evolution within a single infected individual and go 'beyond the consensus' [171]. Quantification of such within-host viral variation has only recently been made possible by next-generation 'deep sequencing', where a very large number of sequencing reads — often thousands — overlap each position being sequenced. This allows the detection of within-host viral variants such as **intrahost single nucleotide variants (iSNVs)**. The relative frequency

of a particular variant among the viral sequence reads, often referred to as its **variant allele fraction (VAF)**, can then be used to estimate the allele's frequency in the within-host virus population. For example, a C→T iSNV that is present in 10% of reads is inferred to have a relative frequency of 10% in the virus population infecting the host.

Because sequencing error alone can produce low-frequency false-positive variants, appropriate filtering and quality control metrics are essential for within-host analyses. Filtering usually includes a minimum VAF (e.g., 5%), minimum total read coverage (e.g., 200), minimum absolute number of reads containing the variant (e.g., 10), and elimination of variants displaying strand bias. It has been suggested that the total read coverage should be 10 times the reciprocal of the desired minimum VAF, e.g., 10/0.05 = 200 reads to reliably detect variants at a frequency of 5% [172]. Further, amplicons containing mismatches in PCR primer regions should be eliminated because they can experience amplification biases that invalidate frequency estimates [173].

Within-host diversity is not always capable of being transmitted. Transmission to a new host requires a fully functional virion, and is therefore a major selective event. As a result, within-host diversity often includes transient, potentially deleterious mutations that do not transmit to new hosts, evidenced by the fact that π_N/π_S or d_N/d_S ratios are usually higher (closer to 1) within hosts than between hosts [174]. In the case of HPV, this is clearly seen from the accumulation of non-synonymous changes which — even if they contribute to within-host persistence — would fail to produce infectious virion (e.g., in L1).

HPV types have historically been treated as static or fixed sequences. The major insight provided by genomics over the past decade has been that substantial variation within a type can exist, accumulate, and even modulate cancer risk by orders of magnitude. Ultimately, all such viral variation must have initially arisen as a within-host mutation.

4.4.1. *APOBEC3-induced variation*

One of the major causes of within-host HPV polymorphism is the interferon-stimulated host APOBEC3 (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3) family of cytidine deaminases. APOBEC3 is thought to combat infection by introducing deleterious mutations into the viral genome. This could cause viral clearance either through specific changes (e.g., creation of neoantigens that expose the virus to the immune system) or a sufficiently large number of changes that the viral genomes are rendered nonviable (i.e., lethal mutagenesis [175]). For this to be effective, the mutations likely must occur in the viral reservoir in the basal cell layer.

APOBEC3 specifically acts on single-stranded DNA, such as occurs during transcription and replication, to induce C→U changes predominantly at the C of TpCpW (W = A or T) **trinucleotide** motifs. This can lead to C→T changes (via lack of repair or base excision repair/Strauss's A rule) and C→G changes (via base excision repair/REV1), accounting for COSMIC single base substitution (SBS) mutational signatures SBS2 and SBS13, respectively [176]. However, while APOBEC signatures SBS2 and SBS13 are both observed in the host (somatic) genome [177], only SBS2 has been observed in the HPV genome during infection [168].

A role for APOBEC3 in HPV infection was first established when Vartanian et al. showed that HPV16 mutations in cervical precancers correlate with changes induced by APOBEC3 expression *in vitro* [178]. However, it wasn't yet clear how or if these variations contribute to carcinogenesis. More recently, deep sequencing of 151 clinical samples with HPV types 16, 52, and 58 has suggested that the frequency of APOBEC3-compatible iSNVs decreases with progression to cancer [54]. Focusing on HPV16, deep sequencing of 5328 HPV16 samples shows that iSNVs consistent with APOBEC3 activity are enriched in controls compared to cases [168]. These results suggest APOBEC3 may help to reduce viral persistence — and, by extension, progression to cancer — within a host.

Despite APOBEC3's role in controlling viral infection, its mutagenic activity may be a double-edged sword: APOBEC3 signatures are also evident in host (somatic) genomes [177]. Such 'off-target' mutagenesis may contribute to carcinogenesis, i.e., the antiviral mechanism may inadvertently cause cancer and play the role of either 'friend or foe' [59]. Interestingly, a deletion removing the unique portion of APOBEC3B to create an APOBEC3A/APOBEC3B hybrid was found to be very common in East Asian, Native American, and Oceanic populations [179]. Although the effect of this deletion on HPV clearance or cancer risk is unclear (reviewed in Ref. [59]), it is conceivable that it could modulate clearance of different HPV types or variants in different populations.

Beyond contributing to cancer, APOBEC3 may also help to compensate for HPV's evolutionary limitation of a low mutation rate by providing additional mutational resources, e.g., for immune evasion in a present or future host. This is compatible with the overall saturation of nonsynonymous TpC→TpT changes observed in the virus' evolutionary history [168,169], and the fact that APOBEC-induced mutations have been observed to inadvertently benefit other viruses [180,181].

Finally, it is important to recognize that observed APOBEC3 mutations have likely been biased by natural selection; any mutations that eliminate a viral genome within a host or prevent its transmission to a new host will not persist to be sampled and sequenced.

4.4.2. *Neither quasispecies nor invariant*

Within-host variation should not be confused with the concept of **quasispecies** [182]. Briefly, quasispecies theory applies to situations in which mutation rates are so high — typically >1 mutation per genome per replication — that they produce a network ('cloud' or 'swarm') of interrelated genotypes each replication cycle. Such high mutation rates lead to an approximate steady state of unstable sequences, such that selection no longer acts on individual genomes, but rather groups of closely related genomes connected by 'mutational coupling' [183]. This does not describe the situation with HPV, where mutation rates are too low and within-host viral populations too small to give rise to quasispecies dynamics, and where a single viral genome sequence physically exists and forms a consensus in the majority of samples. For perspective, it is even questionable whether quasispecies theory applies to highly mutable RNA viruses [174]. Quasispecies is not synonymous with the presence of within-host viral polymorphism.

4.5. *Integration*

Integration into the host genome exists on a continuum, ranging from none to some to all of the HPV genomes in a cell. It is not a normal part of the HPV life cycle, and often occurs in such a way as to disrupt or delete whole ORFs, representing a dead end for the virus [45]. Nevertheless, integration can lead to cancer by conferring a growth advantage on its host cell. Most notably, integration disrupting E1 and E2 (which together regulate the expression of E6 and E7) is thought to be a major path of HPV-driven oncogenesis [45,48]. Less commonly, integration near host genes (e.g., MYC) may cause aberrant expression that promotes cancer [184,185]. However, other mechanisms such as mutations or methylation may lead to similar results, and a substantial proportion of specifically HPV16-associated cancers do not involve integrants [130].

Short-read technologies like Ion Torrent and Illumina can be used to detect integration breakpoints (sites of fusion between host and virus DNA) but may fail to characterize full integration events (complete stretches of HPV DNA flanked on both sides by host DNA). To address this limitation, Nanopore long-read sequencing has recently been used to describe integration events in HPV16-positive cervical cancers. Integration was observed in 15 of 16 tumour samples, with 0–13 breakpoints and 0–5 events per sample, i.e., the same breakpoint was often observed in multiple events in the same sample [186]. Breakpoints were enriched in E1 and E2, and all samples with integration contained at least one event maintaining E6 and E7 DNA, consistent with the dogma that expression of the oncoproteins is important for cervical cancer maintenance [45]. However, RNA expression of E6 and E7 was relatively low in

one sample, raising the possibility of an alternative oncogenic pathway [186]. Of note, another study of oropharyngeal cancers did not find an enrichment of E2 breakpoints [187], raising the possibility that cancers at different anatomical sites differ in mechanism.

An important caveat applies when interpreting studies of integration. Integration events may occur randomly, but the subset of events maintaining the oncoprotein ORFs may be positively selected because they confer a growth advantage to their host cell. As a result, most studies of viral integration only describe the properties of integration specifically after within-host natural selection of virus and/or host (somatic) genomes has occurred.

5. Conclusions

Despite the availability of highly effective VLP-based vaccines against HPV, carcinogenic HPVs still cause ~604,000 new cervical cancers and ~124,000 new non-cervical cancers globally each year [2, 3]. Thus, it remains important to understand the genetic basis of HPV oncogenicity, particularly that of the uniquely carcinogenic HPV16 type [10].

Like all cancers, HPV-induced cancer development likely involves numerous chance events including transmission, infection by a specific HPV type or variant, mutation, integration, and host (somatic) genetic changes. The stochastic nature of this process suggests there are many unique genetic causes of cervical cancer. Nevertheless, it is hoped that general patterns can be deciphered, and the availability of unprecedented numbers of whole HPV genomes is making this goal increasingly attainable. At the same time, new data are raising a smorgasbord of questions including the relative contributions of virus and host genetics to cancer, the importance of variability between and within HPV types, and the importance of within-host viral polymorphism (see Box 2).

Box 2
Open Questions in HPV Genomics

1. What makes HPV16 uniquely carcinogenic at the cervix and particularly at non-cervical sites?
2. Why are HPV16 sublineages A4, D2, and D3 more associated with adenocarcinoma than A1/A2?
3. Can studies of non-carcinogenic HPV types help to inform us about cancer mechanisms? For example, does the loss of carcinogenicity in some HPV types (e.g., HPV97 in the HPV18/HPV45/HPV97 cluster) help to inform about the genetic basis of cancer in related types?
4. What are the relative contributions of virus vs. host genomic changes in the steps leading to carcinogenesis?
5. What are the relative contributions of virus genetics (e.g., E7 genotype) vs. host genetics (e.g., MHC/HLA alleles) in determining infection outcomes?
6. Can synonymous sites in protein-coding regions be used to derive a better estimate of the HPV mutation rate?
7. What are the relative contributions of natural selection (e.g., immune escape), mutation pressure (e.g., APOBEC3), host/pathogen co-divergence, and genetic drift to HPV evolutionary history?
8. Given APOBEC3 signatures are present in both virus and host (somatic) genomes, does this enzyme primarily promote or impede carcinogenesis?
9. Does a mutation or deletion of APOBEC3 modulate risk of cervical cancer and/or clearance of specific HPV types or variants in different populations?
10. Can infectious virus ever be produced from integrated copies, or is integration always a dead end for the virus life cycle? If integration is not always a dead end, is it ever employed as a strategy for immune avoidance or latency?

Understanding the genetic basis of HPV carcinogenicity will not only assist in the fight against morbidity and mortality associated with cervical cancer, but also increasingly prevalent HPV-driven cancers at other anatomical sites in both men and women — as well as other infection-attributable cancers that may share HPV's mechanisms of oncogenesis.

CRediT author statement

Chase W. Nelson: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization.

Lisa Mirabello: Conceptualization, Validation, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Funding

This research was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics of the National Cancer Institute (NCI), and by the NCI Research Participation Program administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy (DOE) and the National Institute of Health (NIH). ORISE is managed by ORAU under DOE contract number DESC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of NIH, NCBI, DOE, or ORAU/ORISE.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All sequence data are cited and publicly available on GenBank.

Acknowledgments

We thank Meredith Yeager for extensive discussion and feedback; Ming-Hsueh Lin for feedback on figures; Leonardo Varuzza and Felipe Luiz Pereira for feedback on Ion Torrent error rates; the members of the DCEG-NCI HPV Genomics Group (Laurie Burdette, Michael Dean, Aimee Koestler, Elisa Lee, Hong Lou, Sambit Mishra, Maisa Pinheiro, Meredith Yeager), Zachary Ardern, Chen-Hao Kuo, and Xinzhu (April) Wei for discussion; and our reviewers for feedback. We express sincere apologies to those researchers whose work could not be cited due to space limitations and the scope of this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tvr.2023.200258.

References

[1] C. de Martel, M. Plummer, J. Vignat, S. Franceschi, Worldwide burden of cancer attributable to HPV by site, country and HPV type, Int. J. Cancer 141 (2017) 664–670, https://doi.org/10.1002/ijc.30716.
[2] C. de Martel, D. Georges, F. Bray, J. Ferlay, G.M. Clifford, Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis, Lancet Global Health 8 (2020), https://doi.org/10.1016/S2214-109X(19)30488-7 e180–e190.
[3] GLOBOCAN, The Global Cancer Observatory, Cancer Today, 2020 (accessed June 21, 2022), https://gco.iarc.fr/.
[4] V. Bouvard, R. Baan, K. Straif, Y. Grosse, B. Secretan, F. El Ghissassi, L. Benbrahim-Tallaa, N. Guha, C. Freeman, L. Galichet, A review of human

carcinogens—Part B: biological agents, Lancet Oncol. 10 (2009) 321–322, https://doi.org/10.1016/s1470-2045(09)70096-8.
[5] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, Biological agents. Volume 100 B. A review of human carcinogens, IARC Monogr. Eval. Carcinog. Risks Hum. 100 (2012) 1–441.
[6] K. Van Doorslaer, Z. Li, S. Xirasagar, P. Maes, D. Kaminsky, D. Liou, Q. Sun, R. Kaur, Y. Huyen, A.A. McBride, The Papillomavirus Episteme: a major update to the papillomavirus sequence database, Nucleic Acids Res. 45 (2017) D499–D506, https://doi.org/10.1093/nar/gkw879.
[7] PaVE, The Papillomavirus Episteme, the Papillomavirus Episteme, 2022 (accessed June 21, 2022), https://pave.niaid.nih.gov/.
[8] IARC, IARC monographs on the identification of carcinogenic hazards to humans, online database (accessed June 21, 2022), https://monographs.iarc.who.int/, 2022.
[9] J. Doorbar, N. Egawa, H. Griffin, C. Kranjec, I. Murakami, Human papillomavirus molecular biology and disease association, Rev. Med. Virol. 25 (2015) 2–23, https://doi.org/10.1002/rmv.1822.
[10] M. Demarco, N. Hyun, O. Carter-Pokras, T.R. Raine-Bennett, L. Cheung, X. Chen, A. Hammer, N. Campos, W. Kinney, J.C. Gage, B. Befano, R.B. Perkins, X. He, C. Dallal, J. Chen, N. Poitras, M.-H. Mayrand, F. Coutlee, R.D. Burk, T. Lorey, P. E. Castle, N. Wentzensen, M. Schiffman, A study of type-specific HPV natural history and implications for contemporary cervical cancer screening programs, EClinicalMedicine 22 (2020), 100293, https://doi.org/10.1016/j.eclinm.2020.100293.
[11] J. Doorbar, W. Quint, L. Banks, I.G. Bravo, M. Stoler, T.R. Broker, M.A. Stanley, The biology and life-cycle of human papillomaviruses, Vaccine 30 (2012), https://doi.org/10.1016/j.vaccine.2012.06.083. F55–F70.
[12] N.A. Krump, J. You, Molecular mechanisms of viral oncogenesis in humans, Nat. Rev. Microbiol. 16 (2018) 684–698, https://doi.org/10.1038/s41579-018-0064-6.
[13] S. de Sanjose, W.G. Quint, L. Alemany, D.T. Geraets, J.E. Klaustermeier, B. Lloveras, S. Tous, A. Felix, L.E. Bravo, H.-R. Shin, C.S. Vallejos, P.A. de Ruiz, M. A. Lima, N. Guimera, O. Clavero, M. Alejo, A. Llombart-Bosch, C. Cheng-Yang, S. A. Tatti, E. Kasamatsu, E. Iljazovic, M. Odida, R. Prado, M. Seoud, M. Grce, A. Usubutun, A. Jain, G.A.H. Suarez, L.E. Lombardi, A. Banjo, C. Menéndez, E. J. Domingo, J. Velasco, A. Nessa, S.C.B. Chichareon, Y.L. Qiao, E. Lerma, S. M. Garland, T. Sasagawa, A. Ferrera, D. Hammouda, L. Mariani, A. Pelayo, I. Steiner, E. Oliva, C.J. Meijer, W.F. Al-Jassar, E. Cruz, T.C. Wright, A. Puras, C. L. Llave, M. Tzardi, T. Agorastos, V. Garcia-Barriola, C. Clavel, J. Ordi, M. Andújar, X. Castellsagué, G.I. Sánchez, A.M. Nowakowski, J. Bornstein, N. Muñoz, F.X. Bosch, Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study, Lancet Oncol. 11 (2010) 1048–1056, https://doi.org/10.1016/S1470-2045(10)70230-8.
[14] M. Arbyn, M. Tommasino, C. Depuydt, J. Dillner, Are 20 human papillomavirus types causing cervical cancer? J. Pathol. 234 (2014) 431–435, https://doi.org/10.1002/path.4424.
[15] A. Harari, Z. Chen, R.D. Burk, Human papillomavirus genomics: past, present and future, in: M.K. Ramírez-Fort, F. Khan, P.L. Rady, S.K. Tyring (Eds.), Current Problems in Dermatology, S. KARGER AG, Basel, 2014, pp. 1–18, https://doi.org/10.1159/000355952.
[16] R.D. Burk, Z. Chen, K. Van Doorslaer, Human papillomaviruses: genetic basis of carcinogenicity, Public Health Genomics 12 (2009) 281–290, https://doi.org/10.1159/000214919.
[17] L. Yu, V. Majerciak, Z.-M. Zheng, HPV16 and HPV18 genome structure, expression, and post-transcriptional regulation, IJMS 23 (2022) 4943, https://doi.org/10.3390/ijms23094943.
[18] A. Vats, O. Trejo-Cerro, M. Thomas, L. Banks, Human papillomavirus E6 and E7: what remains? Tumour Virus Res. 11 (2021), 200213 https://doi.org/10.1016/j.tvr.2021.200213.
[19] E.C. Goodwin, E. Yang, C.-J. Lee, H.-W. Lee, D. DiMaio, E.-S. Hwang, Rapid induction of senescence in human cervical carcinoma cells, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 10978–10983, https://doi.org/10.1073/pnas.97.20.10978.
[20] E.A. Mesri, M.A. Feitelson, K. Munger, Human viral oncogenesis: a cancer hallmarks analysis, Cell Host Microbe 15 (2014) 266–282, https://doi.org/10.1016/j.chom.2014.02.011.
[21] C.A. Moody, L.A. Laimins, Human papillomavirus oncoproteins: pathways to transformation, Nat. Rev. Cancer 10 (2010) 550–560, https://doi.org/10.1038/nrc2886.
[22] I.G. Bravo, Á. Alonso, Mucosal human papillomaviruses encode four different E5 proteins whose chemistry and phylogeny correlate with malignant or benign growth, J. Virol. 78 (2004) 13613–13626, https://doi.org/10.1128/JVI.78.24.13613-13626.2004.
[23] M.S. Campo, S.V. Graham, M.S. Cortese, G.H. Ashrafi, E.H. Araibi, E.S. Dornan, K. Miners, C. Nunes, S. Man, HPV-16 E5 down-regulates expression of surface HLA class I and reduces recognition by CD8 T cells, Virology 407 (2010) 137–142, https://doi.org/10.1016/j.virol.2010.07.044.
[24] D. DiMaio, L.M. Petti, The E5 proteins, Virology 445 (2013) 99–114, https://doi.org/10.1016/j.virol.2013.05.006.
[25] G.H. Ashrafi, E. Tsirimonaki, B. Marchetti, P.M. O'Brien, G.J. Sibbet, L. Andrew, M.S. Campo, Down-regulation of MHC class I by bovine papillomavirus E5 oncoproteins, Oncogene 21 (2002) 248–259, https://doi.org/10.1038/sj.onc.1205008.
[26] A. Willemsen, M. Félez-Sánchez, I.G. Bravo, Genome plasticity in papillomaviruses and de novo emergence of E5 oncogenes, Genome Biol. Evolut. 11 (2019) 1602–1617, https://doi.org/10.1093/gbe/evz095.
[27] A.A. McBride, Mechanisms and strategies of papillomavirus replication, Biol. Chem. 398 (2017) 919–927, https://doi.org/10.1515/hsz-2017-0113.
[28] A.A. McBride, The Papillomavirus E2 proteins, Virology 445 (2013) 57–79, https://doi.org/10.1016/j.virol.2013.06.006.
[29] T.L. Coursey, A.A. McBride, Hitchhiking of viral genomes on cellular chromosomes, Annu. Rev. Virol. 6 (2019) 275–296, https://doi.org/10.1146/annurev-virology-092818-015716.
[30] J.A. Smith, E.A. White, M.E. Sowa, M.L.C. Powell, M. Ottinger, J.W. Harper, P. M. Howley, Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 3752–3757, https://doi.org/10.1073/pnas.0914818107.
[31] M. Dreer, J. Fertey, S. van de Poel, E. Straub, J. Madlung, B. Macek, T. Iftner, F. Stubenrauch, Interaction of NCOR/SMRT repressor complexes with papillomavirus E8^E2C proteins inhibits viral replication, PLoS Pathog. 12 (2016), e1005556, https://doi.org/10.1371/journal.ppat.1005556.
[32] N. Sakakibara, D. Chen, A.A. McBride, Papillomaviruses use recombination-dependent replication to vegetatively amplify their genomes in differentiated cells, PLoS Pathog. 9 (2013), e1003321, https://doi.org/10.1371/journal.ppat.1003321.
[33] J. Doorbar, The E4 protein; structure, function and patterns of expression, Virology 445 (2013) 80–98, https://doi.org/10.1016/j.virol.2013.07.008.
[34] R.D. Burk, A. Harari, Z. Chen, Human papillomavirus genome variants, Virology 445 (2013) 232–243, https://doi.org/10.1016/j.virol.2013.07.018.
[35] X.S. Chen, R.L. Garcea, I. Goldberg, G. Casini, S.C. Harrison, Structure of small virus-like particles assembled from the L1 protein of human papillomavirus 16, Mol. Cell 5 (2000) 557–567, https://doi.org/10.1016/S1097-2765(00)80449-9.
[36] M. Stanley, D.R. Lowy, I. Frazer, Chapter 12: prophylactic HPV vaccines: underlying mechanisms, Vaccine 24 (2006) S106–S113, https://doi.org/10.1016/j.vaccine.2006.05.110.
[37] P.R. Prabhu, J.J. Carter, D.A. Galloway, B cell responses upon human papillomavirus (HPV) infection and vaccination, Vaccines 10 (2022) 837, https://doi.org/10.3390/vaccines10060837.
[38] V.A. Olcese, Y. Chen, R. Schlegel, H. Yuan, Characterization of HPV16 L1 loop domains in the formation of a type-specific, conformational epitope, BMC Microbiol. 4 (2004) 29, https://doi.org/10.1186/1471-2180-4-29.
[39] S.D. Shah, J. Doorbar, R.A. Goldstein, Analysis of host–parasite incongruence in papillomavirus evolution using importance sampling, Mol. Biol. Evol. 27 (2010) 1301–1314, https://doi.org/10.1093/molbev/msq015.
[40] N. Egawa, J. Doorbar, The low-risk papillomaviruses, Virus Res. 231 (2017) 119–127, https://doi.org/10.1016/j.virusres.2016.12.017.
[41] M.G. Frattini, H.B. Lim, L.A. Laimins, In vitro synthesis of oncogenic human papillomaviruses requires episomal genomes for differentiation-dependent late expression, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 3062–3067, https://doi.org/10.1073/pnas.93.7.3062.
[42] M.A. Bedell, J.B. Hudson, T.R. Golub, M.E. Turyk, M. Hosken, G.D. Wilbanks, L. A. Laimins, Amplification of human papillomavirus genomes in vitro is dependent on epithelial differentiation, J. Virol. 65 (1991) 2254–2260, https://doi.org/10.1128/jvi.65.5.2254-2260.1991.
[43] M.A. Stanley, H.M. Browne, M. Appleby, A.C. Minson, Properties of a non-tumorigenic human cervical keratinocyte cell line, Int. J. Cancer 43 (1989) 672–676, https://doi.org/10.1002/ijc.2910430422.
[44] G.A. Maglennon, P. McIntosh, J. Doorbar, Persistence of viral DNA in the epithelial basal layer suggests a model for papillomavirus latency following immune regression, Virology 414 (2011) 153–163, https://doi.org/10.1016/j.virol.2011.03.019.
[45] A.A. McBride, A. Warburton, The role of integration in oncogenic progression of HPV-associated cancers, PLoS Pathog. 13 (2017), e1006211, https://doi.org/10.1371/journal.ppat.1006211.
[46] M.A. Stanley, Epithelial cell responses to infection with human papillomavirus, Clin. Microbiol. Rev. 25 (2012) 215–222, https://doi.org/10.1128/CMR.05028-11.
[47] R.B.S. Roden, P.L. Stern, Opportunities and challenges for human papillomavirus vaccination in cancer, Nat. Rev. Cancer 18 (2018) 240–254, https://doi.org/10.1038/nrc.2018.13.
[48] M.A. Stanley, M.R. Pett, N. Coleman, HPV: from infection to cancer, Biochem. Soc. Trans. 35 (2007) 1456–1460, https://doi.org/10.1042/BST0351456.
[49] P.S. Moore, Y. Chang, Why do viruses cause cancer? Highlights of the first century of human tumour virology, Nat. Rev. Cancer 10 (2010) 878–889, https://doi.org/10.1038/nrc2961.
[50] L. Mirabello, M.A. Clarke, C.W. Nelson, M. Dean, N. Wentzensen, M. Yeager, M. Cullen, J. Boland, , NCI HPV Workshop, M. Schiffman, R.D. Burk, The intersection of HPV epidemiology, genomics and mechanistic studies of HPV-mediated carcinogenesis, Viruses 10 (2018) 80, https://doi.org/10.3390/v10020080.
[51] K.J. Syrjänen, S. Pyrhönen, Immunoperoxidase demonstration of human papilloma virus (HPV) in dysplastic lesions of the uterine cervix, Arch. Gynecol. 233 (1982) 53–61, https://doi.org/10.1007/BF02110679.
[52] M. Schiffman, J. Doorbar, N. Wentzensen, S. de Sanjosé, C. Fakhry, B.J. Monk, M. A. Stanley, S. Franceschi, Carcinogenic human papillomavirus infection, Nat. Rev. Dis. Prim. 2 (2016), 16086, https://doi.org/10.1038/nrdp.2016.86.
[53] I. Kukimoto, T. Maehama, T. Sekizuka, Y. Ogasawara, K. Kondo, R. Kusumoto-Matsuo, S. Mori, Y. Ishii, T. Takeuchi, T. Yamaji, F. Takeuchi, K. Hanada, M. Kuroda, Genetic variation of human papillomavirus type 16 in individual clinical specimens revealed by deep sequencing, PLoS One 8 (2013), e80583, https://doi.org/10.1371/journal.pone.0080583.

[54] Y. Hirose, M. Onuki, Y. Tenjimbayashi, S. Mori, Y. Ishii, T. Takeuchi, N. Tasaka, T. Satoh, T. Morisada, T. Iwata, S. Miyamoto, K. Matsumoto, A. Sekizawa, I. Kukimoto, Within-host variations of human papillomavirus reveal APOBEC signature mutagenesis in the viral genome, J. Virol. 92 (2018) e00017–e00018, https://doi.org/10.1128/JVI.00017-18.

[55] M. Schiffman, R. Herrero, R. DeSalle, A. Hildesheim, S. Wacholder, A. Cecilia Rodriguez, M.C. Bratti, M.E. Sherman, J. Morales, D. Guillen, M. Alfaro, M. Hutchinson, T.C. Wright, D. Solomon, Z. Chen, J. Schussler, P.E. Castle, R. D. Burk, The carcinogenicity of human papillomavirus types reflects viral evolution, Virology 337 (2005) 76–84, https://doi.org/10.1016/j.virol.2005.04.002.

[56] D. Bzhalava, P. Guan, S. Franceschi, J. Dillner, G. Clifford, A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types, Virology 445 (2013) 224–231, https://doi.org/10.1016/j.virol.2013.07.015.

[57] E.R. Flores, P.F. Lambert, Evidence for a switch in the mode of human papillomavirus type 16 DNA replication during the viral life cycle, J. Virol. 71 (1997) 7167–7179, https://doi.org/10.1128/jvi.71.10.7167-7179.1997.

[58] S.F. Roerink, R. van Schendel, M. Tijsterman, Polymerase theta-mediated end joining of replication-associated DNA breaks in C. elegans, Genome Res. 24 (2014) 954–962, https://doi.org/10.1101/gr.170431.113.

[59] C.J. Warren, M.L. Santiago, D. Pyeon, APOBEC3: friend or foe in human papillomavirus infection and oncogenesis? Annu. Rev. Virol. 9 (2022) 16.1–16.21, https://doi.org/10.1146/annurev-virology-092920-030354.

[60] K.J. Fryxell, E. Zuckerkandl, Cytosine deamination plays a primary role in the evolution of mammalian isochores, Mol. Biol. Evol. 17 (2000) 1371–1383, https://doi.org/10.1093/oxfordjournals.molbev.a026420.

[61] Z. Chen, F. Utro, D. Platt, R. DeSalle, L. Parida, P.K.S. Chan, R.D. Burk, K-mer analyses reveal different evolutionary histories of alpha, Beta, and Gamma papillomaviruses, IJMS 22 (2021) 9657, https://doi.org/10.3390/ijms22179657.

[62] K.M. King, E.V. Rajadhyaksha, I.G. Tobey, K. Van Doorslaer, Synonymous nucleotide changes drive papillomavirus evolution, Tumour Virus Res. 14 (2022), 200248, https://doi.org/10.1016/j.tvr.2022.200248.

[63] K.E.K. Rowson, B.W.J. Mahy, Human papova (wart) virus, Bacteriol. Rev. 31 (1967) 110–131, https://doi.org/10.1128/br.31.2.110-131.1967.

[64] C. Meyers, M.G. Frattini, J.B. Hudson, L.A. Laimins, Biosynthesis of human papillomavirus from a continuous cell line upon epithelial differentiation, Science 257 (1992) 971–973, https://doi.org/10.1126/science.1323879.

[65] S.C. Fausch, D.M. Da Silva, G.L. Eiben, C. Le Poole, W.M. Kast, HPV protein/peptide vaccines: from animal models to clinical trials, Front. Biosci. 8 (2003) s81–s91, https://doi.org/10.2741/1009.

[66] K. Van Doorslaer, Evolution of the Papillomaviridae, Virology 445 (2013) 11–20, https://doi.org/10.1016/j.virol.2013.05.012.

[67] M. Kimura, Evolutionary rate at the molecular level, Nature 217 (1968) 624–626, https://doi.org/10.1038/217624a0.

[68] M. Kimura, The Neutral Theory of Molecular Evolution, Cambridge University Press, 1983.

[69] C.-K. Ong, S.-Y. Chan, M.S. Campo, K. Fujinaga, P. Mavromara-Nazos, V. Labropoulou, H. Pfister, S.-K. Tay, J. ter Meulen, L.L. Villa, H.-U. Bernard, Evolution of human papillomavirus type 18: an ancient phylogenetic root in Africa and intratype diversity reflect coevolution with human ethnic groups, J. Virol. 67 (1993) 6424–6431, https://doi.org/10.1128/jvi.67.11.6424-6431.1993.

[70] A. Rector, P. Lemey, R. Tachezy, S. Mostmans, S.-J. Ghim, K. Van Doorslaer, M. Roelke, M. Bush, R.J. Montali, J. Joslin, R.D. Burk, A.B. Jenson, J.P. Sundberg, B. Shapiro, M. Van Ranst, Ancient papillomavirus-host co-speciation in Felidae, Genome Biol. 8 (2007) R57, https://doi.org/10.1186/gb-2007-8-4-r57.

[71] G.M. Jenkins, A. Rambaut, O.G. Pybus, E.C. Holmes, Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis, J. Mol. Evol. 54 (2002) 156–165, https://doi.org/10.1007/s00239-001-0064-3.

[72] K. Hanada, Y. Suzuki, T. Gojobori, A large variation in the rates of synonymous substitution for RNA viruses and its relationship to a diversity of viral infection and transmission modes, Mol. Biol. Evol. 21 (2004) 1074–1080, https://doi.org/10.1093/molbev/msh109.

[73] H. Jónsson, P. Sulem, B. Kehr, S. Kristmundsdottir, F. Zink, E. Hjartarson, M. T. Hardarson, K.E. Hjorleifsson, H.P. Eggertsson, S.A. Gudjonsson, L.D. Ward, G. A. Arnadottir, E.A. Helgason, H. Helgason, A. Gylfason, A. Jonasdottir, A. Jonasdottir, T. Rafnar, M. Frigge, S.N. Stacey, O. Th Magnusson, U. Thorsteinsdottir, G. Masson, A. Kong, B.V. Halldorsson, A. Helgason, D. F. Gudbjartsson, K. Stefansson, Parental influence on human germline de novo mutations in 1,548 trios from Iceland, Nature 549 (2017) 519–522, https://doi.org/10.1038/nature24018.

[74] M. Nei, W.-H. Li, Mathematical model for studying genetic variation in terms of restriction endonucleases, Proceed. National Acad. Sci. USA 76 (1979) 5269–5273, https://doi.org/10.1073/pnas.76.10.5269.

[75] L. Zhao, C.J.R. Illingworth, Measurements of intrahost viral diversity require an unbiased diversity metric, Virus Evolution 5 (2019) vey041, https://doi.org/10.1093/ve/vey041.

[76] M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions, Mol. Biol. Evol. 3 (1986) 418–426, https://doi.org/10.1093/oxfordjournals.molbev.a040410.

[77] C.W. Nelson, A.L. Hughes, Within-host nucleotide diversity of virus populations: insights from next-generation sequencing, Infect. Genet. Evol. 30 (2015) 1–7, https://doi.org/10.1016/j.meegid.2014.11.026.

[78] A.L. Hughes, Adaptive Evolution of Genes and Genomes, Oxford University Press, New York, NY, 1999.

[79] S. Kryazhimskiy, J.B. Plotkin, The Population Genetics of dN/dS, PLoS Genet. 4 (2008), e1000304, https://doi.org/10.1371/journal.pgen.1000304.

[80] E.C. Holmes, D.J. Lipman, D. Zamarin, J.W. Yewdell, Comment on 'large-scale sequence analysis of avian influenza isolates', Science 313 (2006) https://doi.org/10.1126/science.1131729, 1573b–1573b.

[81] C.W. Nelson, Z. Ardern, X. Wei, OLGenie: estimating natural selection to predict functional overlapping genes, Mol. Biol. Evol. 37 (2020) 2440–2449, https://doi.org/10.1093/molbev/msaa087.

[82] X. Wei, J. Zhang, A simple method for estimating the strength of natural selection on overlapping genes, Genome Biol. Evolut. 7 (2015) 381–390, https://doi.org/10.1093/gbe/evu294.

[83] J. Dillner, Mapping of linear epitopes of human papillomavirus type 16: the E1, E2, E4, E5, E6 and E7 open reading frames, Int. J. Cancer 46 (1990) 703–711, https://doi.org/10.1002/ijc.2910460426.

[84] M. Lehtinen, M.H. Hibma, G. Stellato, T. Kuoppala, J. Paavonen, Human T helper cell epitopes overlap B cell and putative cytotoxic T cell epitopes in the E2 protein of human papillomavirus type 16, Biochem. Biophys. Res. Commun. 209 (1995) 541–546, https://doi.org/10.1006/bbrc.1995.1535.

[85] M. Félez-Sánchez, J.-H. Trösemeier, S. Bedhomme, M.I. González-Bravo, C. Kamp, I.G. Bravo, Cancer, warts, or asymptomatic infections: clinical presentation matches codon usage preferences in human papillomaviruses, Genome Biol. Evol. 7 (2015) 2117–2135, https://doi.org/10.1093/gbe/evv129.

[86] J. Zhang, J.-R. Yang, Determinants of the rate of protein sequence evolution, Nat. Rev. Genet. 16 (2015) 409–420, https://doi.org/10.1038/nrg3950.

[87] A.L. Hughes, M.A.K. Hughes, Patterns of nucleotide difference in overlapping and non-overlapping reading frames of papillomavirus genomes, Virus Res. 113 (2005) 81–88, https://doi.org/10.1016/j.virusres.2005.03.030.

[88] A. Narechania, M. Terai, R.D. Burk, Overlapping reading frames in closely related human papillomaviruses result in modular rates of selection within E2, J. Gen. Virol. 86 (2005) 1307–1313, https://doi.org/10.1099/vir.0.80747-0.

[89] M. Jiang, L.F. Xi, Z.R. Edelstein, D.A. Galloway, G.J. Olsem, W.C.-C. Lin, N. B. Kiviat, Identification of recombinant human papillomavirus type 16 variants, Virology 394 (2009) 8–11, https://doi.org/10.1016/j.virol.2009.08.040.

[90] M. Nikolaidis, D. Tsakogiannis, G. Bletsa, D. Mossialos, C. Kottaridi, I. Iliopoulos, P. Markoulatos, G.D. Amoutzias, HPV16-Genotyper: a computational tool for risk-assessment, lineage genotyping and recombination detection in HPV16 sequences, based on a large-scale evolutionary analysis, Diversity 13 (2021) 497, https://doi.org/10.3390/d13100497.

[91] A. Narechania, Z. Chen, R. DeSalle, R.D. Burk, Phylogenetic incongruence among oncogenic genital alpha human papillomaviruses, J. Virol. 79 (2005) 15503–15510, https://doi.org/10.1128/JVI.79.24.15503-15510.2005.

[92] M.D. Daugherty, H.S. Malik, Rules of engagement: molecular insights from host-virus arms races, Annu. Rev. Genet. 46 (2012) 677–700, https://doi.org/10.1146/annurev-genet-110711-155522.

[93] J.L. Tenthorey, M. Emerman, H.S. Malik, Evolutionary landscapes of host-virus arms races, Annu. Rev. Immunol. 40 (2022) 271–294, https://doi.org/10.1146/annurev-immunol-072621-084422.

[94] J.T. Schiller, D.R. Lowy, Understanding and learning from the success of prophylactic human papillomavirus vaccines, Nat. Rev. Microbiol. 10 (2012) 681–692, https://doi.org/10.1038/nrmicro2872.

[95] R.M. Zinkernagel, H. Hengartner, Antiviral immunity, Immunol. Today 18 (1997) 258–260, https://doi.org/10.1016/S0167-5699(97)80017-5.

[96] H. Trottier, E.L. Franco, The epidemiology of genital human papillomavirus infection, Vaccine 24 (2006), https://doi.org/10.1016/j.vaccine.2005.09.054. S4–S15.

[97] P.J. Leo, M.M. Madeleine, S. Wang, S.M. Schwartz, F. Newell, U. Pettersson-Kymmer, K. Hemminki, G. Hallmans, S. Tiews, W. Steinberg, J.S. Rader, F. Castro, M. Safaeian, E.L. Franco, F. Coutlée, C. Ohlsson, A. Cortes, M. Marshall, P. Mukhopadhyay, K. Cremin, L.G. Johnson, S. Garland, S.N. Tabrizi, N. Wentzensen, F. Sitas, J. Little, M. Cruickshank, I.H. Frazer, A. Hildesheim, M. A. Brown, Defining the genetic susceptibility to cervical neoplasia—a genome-wide association study, PLoS Genet. 13 (2017), e1006866, https://doi.org/10.1371/journal.pgen.1006866.

[98] I. Zehbe, R. Tachezy, J. Mytilineos, G. Voglino, I. Mikyskova, H. Delius, A. Marongiu, L. Gissmann, E. Wilander, M. Tommasino, Human papillomavirus 16 E6 polymorphisms in cervical lesions from different European populations and their correlation with human leukocyte antigen class II haplotypes, Int. J. Cancer 94 (2001) 711–716, https://doi.org/10.1002/ijc.1520.

[99] I. Zehbe, J. Mytilineos, I. Wikström, R. Henriksen, L. Edler, M. Tommasino, Association between human papillomavirus 16 E6 variants and human leukocyte antigen class I polymorphism in cervical cancer of Swedish women, Hum. Immunol. 64 (2003) 538–542, https://doi.org/10.1016/S0198-8859(03)00033-8.

[100] E. Ivansson, I. Juko-Pecirep, H. Erlich, U. Gyllensten, Pathway-based analysis of genetic susceptibility to cervical cancer in situ: HLA-DPB1 affects risk in Swedish women, Gene Immun. 12 (2011) 605–614, https://doi.org/10.1038/gene.2011.40.

[101] Y. Shi, L. Li, Z. Hu, S. Li, S. Wang, J. Liu, C. Wu, L. He, J. Zhou, Z. Li, T. Hu, Y. Chen, Y. Jia, S. Wang, L. Wu, X. Cheng, Z. Yang, R. Yang, X. Li, K. Huang, Q. Zhang, H. Zhou, F. Tang, Z. Chen, J. Shen, J. Jiang, H. Ding, H. Xing, S. Zhang, P. Qu, X. Song, Z. Lin, D. Deng, L. Xi, W. Lv, X. Han, G. Tao, L. Yan, Z. Han, Z. Li, X. Miao, S. Pan, Y. Shen, H. Wang, D. Liu, E. Gong, Z. Li, L. Zhou, X. Luan, C. Wang, Q. Song, S. Wu, H. Xu, J. Shen, F. Qiang, G. Ma, L. Liu, X. Chen, J. Liu, J. Wu, Y. Shen, Y. Wen, M. Chu, J. Yu, X. Hu, Y. Fan, H. He, Y. Jiang, Z. Lei, C. Liu, J. Chen, Y. Zhang, C. Yi, S. Chen, W. Li, D. Wang, Z. Wang, W. Di, K. Shen, D. Lin, H. Shen, Y. Feng, X. Xie, D. Ma, A genome-wide association study identifies two

new cervical cancer susceptibility loci at 4q12 and 17q12, Nat. Genet. 45 (2013) 918–922, https://doi.org/10.1038/ng.2687.
[102] D. Chen, I. Juko-Pecirep, J. Hammer, E. Ivansson, S. Enroth, I. Gustavsson, L. Feuk, P.K.E. Magnusson, J.D. McKay, E. Wilander, U. Gyllensten, Genome-wide association study of susceptibility loci for cervical cancer, JNCI, J. National Cancer Instit. 105 (2013) 624–633, https://doi.org/10.1093/jnci/djt051.
[103] M. Lynch, The Origins of Genome Architecture, Sinauer Associates, Inc. Publishers, Sunderland, MA, 2007.
[104] M.A. Nowak, Evolutionary Dynamics, Belknap/Harvard, Canada, 2006.
[105] J. Zhou, X.Y. Sun, D.J. Stenzel, I.H. Frazer', Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is suffcient for assembly of HPV virion-like particles, Virology 185 (1991) 251–257, https://doi.org/10.1016/0042-6822(91)90772-4.
[106] R. Kirnbauer, F. Booy, N. Cheng, D.R. Lowy, J.T. Schiller, Papillomavirus L1 major capsid protein self-assembles into virus-like particles that are highly immunogenic, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 12180–12184, https://doi.org/10.1073/pnas.89.24.12180.
[107] R. Kirnbauer, J. Taub, H. Greenstone, R. Roden, M. Dürst, L. Gissmann, D. R. Lowy, J.T. Schiller, Efficient self-assembly of human papillomavirus type 16 L1 and L1-L2 into virus-like particles, J. Virol. 67 (1993), https://doi.org/10.1128/JVI.67.12.6929-6936.1993, 6929–6926.
[108] B. Smith, Z. Chen, L. Reimers, K. van Doorslaer, M. Schiffman, R. DeSalle, R. Herrero, K. Yu, S. Wacholder, T. Wang, R.D. Burk, Sequence imputation of HPV16 genomes for genetic association studies, PLoS One 6 (2011), e21375, https://doi.org/10.1371/journal.pone.0021375.
[109] M. Sun, L. Gao, Y. Liu, Y. Zhao, X. Wang, Y. Pan, T. Ning, H. Cai, H. Yang, W. Zhai, Y. Ke, Whole genome sequencing and evolutionary analysis of human papillomavirus type 16 in Central China, PLoS One 7 (2012), e36577, https://doi.org/10.1371/journal.pone.0036577.
[110] P. van der Weele, C.J.L.M. Meijer, A.J. King, Whole-genome sequencing and variant analysis of human papillomavirus 16 infections, J. Virol. 91 (2017), https://doi.org/10.1128/JVI.00844-17 e00844-17.
[111] P. van der Weele, C.J.L.M. Meijer, A.J. King, High whole-genome sequence diversity of human papillomavirus type 18 isolates, Viruses 10 (2018) 68, https://doi.org/10.3390/v10020068.
[112] M. Cullen, J.F. Boland, M. Schiffman, X. Zhang, N. Wentzensen, Q. Yang, Z. Chen, K. Yu, J. Mitchell, D. Roberson, S. Bass, L. Burdette, M. Machado, S. Ravichandran, B. Luke, M.J. Machiela, M. Andersen, M. Osentoski, M. Laptewicz, S. Wacholder, A. Feldman, T. Raine-Bennett, T. Lorey, P.E. Castle, M. Yeager, R.D. Burk, L. Mirabello, Deep sequencing of HPV16 genomes: a new high-throughput tool for exploring the carcinogenicity and natural history of HPV16 infection, Papillomavirus Res. 1 (2015) 3–11, https://doi.org/10.1016/j.pvr.2015.05.004.
[113] L. Mirabello, M. Yeager, K. Yu, G.M. Clifford, Y. Xiao, B. Zhu, M. Cullen, J. F. Boland, N. Wentzensen, C.W. Nelson, T. Raine-Bennett, Z. Chen, S. Bass, L. Song, Q. Yang, M. Steinberg, L. Burdett, M. Dean, D. Roberson, J. Mitchell, T. Lorey, S. Franceschi, P.E. Castle, J. Walker, R. Zuna, A.R. Kreimer, D. C. Beachler, A. Hildesheim, P. Gonzalez, C. Porras, R.D. Burk, M. Schiffman, HPV16 E7 genetic conservation is critical to carcinogenesis, Cell 170 (2017) 1164–1174, https://doi.org/10.1016/j.cell.2017.08.001.
[114] M. Pinheiro, A. Harari, M. Schiffman, G.M. Clifford, Z. Chen, M. Yeager, M. Cullen, J.F. Boland, T. Raine-Bennett, M. Steinberg, S. Bass, Y. Xiao, V. Tenet, K. Yu, B. Zhu, L. Burdett, S. Turan, T. Lorey, P.E. Castle, N. Wentzensen, R. D. Burk, L. Mirabello, Phylogenomic analysis of human papillomavirus type 31 and cervical carcinogenesis: a study of 2093 viral genomes, Viruses 13 (2021) 1948, https://doi.org/10.3390/v13101948.
[115] M. Pinheiro, J.C. Gage, G.M. Clifford, M. Demarco, L.C. Cheung, Z. Chen, M. Yeager, M. Cullen, J.F. Boland, X. Chen, T. Raine-Bennett, M. Steinberg, S. Bass, B. Befano, Y. Xiao, V. Tenet, J. Walker, R. Zuna, N.E. Poitras, M.A. Gold, T. Dunn, K. Yu, B. Zhu, L. Burdett, S. Turan, T. Lorey, P.E. Castle, N. Wentzensen, R.D. Burk, M. Schiffman, L. Mirabello, Association of HPV35 with cervical carcinogenesis among women of African ancestry: evidence of viral-host interaction with implications for disease intervention, Int. J. Cancer 147 (2020) 2677–2686, https://doi.org/10.1002/ijc.33033.
[116] F.L. Pereira, S.C. Soares, F.A. Dorella, C.A.G. Leal, H.C.P. Figueiredo, Evaluating the efficacy of the new Ion PGM Hi-Q Sequencing Kit applied to bacterial genomes, Genomics 107 (2016) 189–198, https://doi.org/10.1016/j.ygeno.2016.03.004.
[117] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M.J. Milgrew, M. Edwards, J. Hoon, J.F. Simons, D. Marran, J.W. Myers, J.F. Davidson, A. Branting, J.R. Nobile, B.P. Puc, D. Light, T.A. Clark, M. Huber, J.T. Branciforte, I.B. Stoner, S.E. Cawley, M. Lyons, Y. Fu, N. Homer, M. Sedova, X. Miao, B. Reed, J. Sabina, E. Feierstein, M. Schorn, M. Alanjary, E. Dimalanta, D. Dressman, R. Kasinskas, T. Sokolsky, J.A. Fidanza, E. Namsaraev, K.J. McKernan, A. Williams, G.T. Roth, J. Bustillo, An integrated semiconductor device enabling non-optical genome sequencing, Nature 475 (2011) 348–352, https://doi.org/10.1038/nature10242.
[118] D.V. Pastrana, A. Peretti, N.L. Welch, C. Borgogna, C. Olivero, R. Badolato, L. D. Notarangelo, M. Gariglio, P.C. FitzGerald, C.E. McIntosh, J. Reeves, G. J. Starrett, V. Bliskovsky, D. Velez, I. Brownell, R. Yarchoan, K.M. Wyvill, T. S. Uldrick, F. Maldarelli, A. Lisco, I. Sereti, C.M. Gonzalez, E.J. Androphy, A. A. McBride, K. Van Doorslaer, F. Garcia, I. Dvoretzky, J.S. Liu, J. Han, P. M. Murphy, D.H. McDermott, C.B. Buck, Metagenomic discovery of 83 new human papillomavirus types in patients with immunodeficiency, mSphere 3 (2018), https://doi.org/10.1128/mSphereDirect.00645-18 e00645-18.
[119] O. Tirosh, S. Conlan, C. Deming, S.-Q. Lee-Lin, X. Huang, , NISC Comparative Sequencing Program, H.C. Su, A.F. Freeman, J.A. Segre, H.H. Kong, Expanded skin virome in DOCK8-deficient patients, Nat. Med. 24 (2018) 1815–1821, https://doi.org/10.1038/s41591-018-0211-7.
[120] E.-M. de Villiers, C. Fauquet, T.R. Broker, H.-U. Bernard, H. zur Hausen, Classification of papillomaviruses, Virology 324 (2004) 17–27, https://doi.org/10.1016/j.virol.2004.03.033.
[121] K. Van Doorslaer, Revisiting papillomavirus taxonomy: a proposal for updating the current classification in line with evolutionary evidence, Viruses 14 (2022) 2308, https://doi.org/10.3390/v14102308.
[122] S.Y. Chan, H. Delius, A.L. Halpern, H.U. Bernard, Analysis of genomic sequences of 95 papillomavirus types: uniting typing, phylogeny, and taxonomy, J. Virol. 69 (1995) 3074–3083, https://doi.org/10.1128/jvi.69.5.3074-3083.1995.
[123] G. Kogure, M. Onuki, Y. Hirose, M. Yamaguchi-Naka, S. Mori, T. Iwata, K. Kondo, A. Sekizawa, K. Matsumoto, I. Kukimoto, Whole-genome analysis of human papillomavirus 67 isolated from Japanese women with cervical lesions, Virol. J. 19 (2022) 157, https://doi.org/10.1186/s12985-022-01894-z.
[124] M. Dürst, L. Gissmann, H. Ikenberg, H. zur Hausen, A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions, Proc. Natl. Acad. Sci. U.S.A. 80 (1983) 3812–3815, https://doi.org/10.1073/pnas.80.12.3812.
[125] X. Castellsagué, J. Klaustermeier, C. Carrilho, G. Albero, J. Sacarlal, W. Quint, B. Kleter, B. Lloveras, M.R. Ismail, S. de Sanjosé, F.X. Bosch, P. Alonso, C. Menéndez, Vaccine-related HPV genotypes in women with and without cervical cancer in Mozambique: burden and potential for prevention, Int. J. Cancer 122 (2007) 1901, https://doi.org/10.1002/ijc.23292. –1904.
[126] C. Okolo, S. Franceschi, I. Adewole, J.O. Thomas, M. Follen, P.J. Snijders, C. J. Meijer, G.M. Clifford, Human papillomavirus infection in women with and without cervical cancer in Ibadan, Nigeria, Infect. Agents Cancer 5 (2010) 24, https://doi.org/10.1186/1750-9378-5-24.
[127] P. Guan, R. Howell-Jones, N. Li, L. Bruni, S. de Sanjosé, S. Franceschi, G. M. Clifford, Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer, Int. J. Cancer 131 (2012) 2349–2359, https://doi.org/10.1002/ijc.27485.
[128] L. Denny, I. Adewole, R. Anorlu, G. Dreyer, M. Moodley, T. Smith, L. Snyman, E. Wiredu, A. Molijn, W. Quint, G. Ramakrishnan, J. Schmidt, Human papillomavirus prevalence and type distribution in invasive cervical cancer in sub-Saharan Africa: cervical cancer in sub-Saharan Africa, Int. J. Cancer 134 (2014) 1389–1398, https://doi.org/10.1002/ijc.28425.
[129] G.M. Clifford, H. de Vuyst, V. Tenet, M. Plummer, S. Tully, S. Franceschi, Effect of HIV infection on human papillomavirus types causing invasive cervical cancer in Africa, JAIDS J. Acquired Immune Deficiency Syndrom. 73 (2016) 332–339, https://doi.org/10.1097/QAI.0000000000001113.
[130] The Cancer Genome Atlas Research Network, Integrated genomic and molecular characterization of cervical cancer, Nature 543 (2017) 378–384, https://doi.org/10.1038/nature21386.
[131] A.J. Klingelhutz, A. Roman, Cellular transformation by human papillomaviruses: lessons learned by comparing high- and low-risk viruses, Virology 424 (2012) 77–98, https://doi.org/10.1016/j.virol.2011.12.018.
[132] N. Auslander, Y.I. Wolf, S.A. Shabalina, E.V. Koonin, A unique insert in the genomes of high-risk human papillomaviruses with a predicted dual role in conferring oncogenic risk, F1000Res 8 (2019) 1000, https://doi.org/10.12688/f1000research.19590.2.
[133] L. Ho, S.-Y. Chan, V. Chow, T. Chong, S.-K. Tay, L.L. Villa, H.-U. Bernard, Sequence variants of human papillomavirus type 16 in clinical samples permit verification and extension of epidemiological studies and construction of a phylogenetic tree, J. Clin. Microbiol. 29 (1991) 1765–1772, https://doi.org/10.1128/jcm.29.9.1765-1772.1991.
[134] G.M. Clifford, V. Tenet, D. Georges, L. Alemany, M.A. Pavón, Z. Chen, M. Yeager, M. Cullen, J.F. Boland, S. Bass, M. Steinberg, T. Raine-Bennett, T. Lorey, N. Wentzensen, J. Walker, R. Zuna, M. Schiffman, L. Mirabello, Human papillomavirus 16 sub-lineage dispersal and cervical cancer risk worldwide: whole viral genome sequences from 7116 HPV16-positive women, Papillomavirus Res. 7 (2019) 67–74, https://doi.org/10.1016/j.pvr.2019.02.001.
[135] S. Nicolás-Párraga, L. Alemany, S. de Sanjosé, F.X. Bosch, I.G. Bravo, RIS HPV TT and HPV VVAP Study Groups, Differential HPV16 variant distribution in squamous cell carcinoma, adenocarcinoma and adenosquamous cell carcinoma: HPV16 variants in different cervical cancer histologies, Int. J. Cancer 140 (2017) 2092–2100, https://doi.org/10.1002/ijc.30636.
[136] V.N. Pimenoff, C.M. de Oliveira, I.G. Bravo, Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16, Mol. Biol. Evol. 34 (2017) 4–19, https://doi.org/10.1093/molbev/msw214.
[137] Z. Chen, R. DeSalle, M. Schiffman, R. Herrero, C.E. Wood, J.C. Ruiz, G.M. Clifford, P.K.S. Chan, R.D. Burk, Niche adaptation and viral transmission of human papillomaviruses from archaic hominins to modern humans, PLoS Pathog. 14 (2018), e1007352, https://doi.org/10.1371/journal.ppat.1007352.
[138] A. Hildesheim, M. Schiffman, C. Bromley, S. Wacholder, R. Herrero, A. C. Rodriguez, M.C. Bratti, M.E. Sherman, U. Scarpidis, Q.-Q. Lin, M. Terai, R. L. Bromley, K. Buetow, R.J. Apple, R.D. Burk, Human papillomavirus type 16 variants and risk of cervical cancer, JNCI J. Nat. Cancer Instit. 93 (2001) 315–318, https://doi.org/10.1093/jnci/93.4.315.
[139] C. Pientong, P. Wongwarissara, T. Ekalaksananan, P. Swangphon, P. Kleebkaow, B. Kongyingyoes, S. Siriaunkgul, K. Tungsinmunkong, C. Suthipintawong, Association of human papillomavirus type 16 long control region mutation and

cervical cancer, Virol. J. 10 (2013) 30, https://doi.org/10.1186/1743-422X-10-30.
[140] L.F. Xi, L.A. Koutsky, A. Hildesheim, D.A. Galloway, C.M. Wheeler, R.L. Winer, J. Ho, N.B. Kiviat, Risk for High-Grade Cervical Intraepithelial Neoplasia Associated with Variants of Human Papillomavirus Types 16 and 18, Cancer Epidemiology, Biomarkers & Prevention, vol. 16, 2007, pp. 4–10, https://doi.org/10.1158/1055-9965.EPI-06-0670.
[141] M. Schiffman, A.C. Rodriguez, Z. Chen, S. Wacholder, R. Herrero, A. Hildesheim, R. Desalle, B. Befano, K. Yu, M. Safaeian, M.E. Sherman, J. Morales, D. Guillen, M. Alfaro, M. Hutchinson, D. Solomon, P.E. Castle, R.D. Burk, A population-based prospective study of carcinogenic human papillomavirus variant lineages, viral persistence, and cervical neoplasia, Cancer Res. 70 (2010) 3159–3169, https://doi.org/10.1158/0008-5472.CAN-09-4179.
[142] I. Cornet, T. Gheit, M.R. Iannacone, J. Vignat, B.S. Sylla, A. Del Mistro, S. Franceschi, M. Tommasino, G.M. Clifford, IARC HPV Variant Study Group, HPV16 genetic variation and the development of cervical cancer worldwide, Br. J. Cancer 108 (2013) 240–244, https://doi.org/10.1038/bjc.2012.508.
[143] T. Gheit, I. Cornet, G.M. Clifford, T. Iftner, C. Munk, M. Tommasino, S.K. Kjaer, Risks for persistence and progression by human papillomavirus type 16 variant lineages among a population-based sample of Danish women, cancer epidemiology, Biomarkers Prevent. 20 (2011) 1315–1321, https://doi.org/10.1158/1055-9965.EPI-10-1187.
[144] I. Zehbe, G. Voglino, H. Delius, E. Wilander, M. Tommasino, Risk of cervical cancer and geographical variations of human papillomavirus 16 E6 polymorphisms, Lancet 352 (1998) 1441–1442, https://doi.org/10.1016/S0140-6736(05)61263-9.
[145] R.E. Zuna, W.E. Moore, R.P. Shanesmith, S.T. Dunn, S.S. Wang, M. Schiffman, G. L. Blakey, T. Teel, Association of HPV16 E6 variants with diagnostic severity in cervical cytology samples of 354 women in a US population, Int. J. Cancer 125 (2009) 2609–2613, https://doi.org/10.1002/ijc.24706.
[146] L. Sichero, S. Ferreira, H. Trottier, E. Duarte-Franco, A. Ferenczy, E.L. Franco, L. L. Villa, High grade cervical lesions are caused preferentially by non-European variants of HPVs 16 and 18, Int. J. Cancer 120 (2007) 1763–1768, https://doi.org/10.1002/ijc.22481.
[147] J. Berumen, R.M. Ordoñez, E. Lazcano, J. Salmeron, S.C. Galvan, R.A. Estrada, E. Yunes, A. Garcia-Carranca, G. Gonzalez-Lira, A. Madrigal-de la Campa, Asian-American variants of human papillomavirus 16 and risk for cervical cancer: a case-control study, JNCI J. Nat. Cancer Instit. 93 (2001) 1325–1330, https://doi.org/10.1093/jnci/93.17.1325.
[148] L.B. Freitas, Z. Chen, E.F. Muqui, N.A.T. Boldrini, A.E. Miranda, L.C. Spano, R. D. Burk, Human papillomavirus 16 non-European variants are preferentially associated with high-grade cervical lesions, PLoS One 9 (2014), e100746, https://doi.org/10.1371/journal.pone.0100746.
[149] L. Mirabello, M. Yeager, M. Cullen, J.F. Boland, Z. Chen, N. Wentzensen, X. Zhang, K. Yu, Q. Yang, J. Mitchell, D. Roberson, S. Bass, Y. Xiao, L. Burdett, T. Raine-Bennett, T. Lorey, P.E. Castle, R.D. Burk, M. Schiffman, HPV16 sublineage associations with histology-specific cancer risk using HPV whole-genome sequences in 3200 women, J. National Cancer Instit. 108 (2016) djw100, https://doi.org/10.1093/jnci/djw100.
[150] R.D. Burk, M. Terai, P.E. Gravitt, L.A. Brinton, R.J. Kurman, W.A. Barnes, M. D. Greenberg, O.C. Hadjimichael, L. Fu, L. McGowan, R. Mortel, P.E. Schwartz, A. Hildesheim, Distribution of human papillomavirus types 16 and 18 variants in squamous cell carcinomas and adenocarcinomas of the cervix, Cancer Res. 63 (2003) 7215–7220.
[151] K.D. Quint, M.N.C. de Koning, L.-J. van Doorn, W.G.V. Quint, E.C. Pirog, HPV genotyping and HPV16 variant analysis in glandular and squamous neoplastic lesions of the uterine cervix, Gynecol. Oncol. 117 (2010) 297–301, https://doi.org/10.1016/j.ygyno.2010.02.003.
[152] S.H. Rabelo-Santos, L.L. Villa, S.F. Derchain, S. Ferreira, L.O.Z. Sarian, L.A. L. Ângelo-Andrade, M.C. do Amaral Westin, L.C. Zeferino, Variants of human papillomavirus types 16 and 18: histological findings in women referred for atypical glandular cells or adenocarcinoma in situ in cervical smear, Int. J. Gynecol. Pathol. 25 (2006) 393–397, https://doi.org/10.1097/01.pgp.0000215302.17029.0c.
[153] S. Nicolás-Párraga, L. Alemany, S. de Sanjosé, F.X. Bosch, I.G. Bravo, RIS HPV TT and HPV VVAP Study Groups, Differential HPV16 variant distribution in squamous cell carcinoma, adenocarcinoma and adenosquamous cell carcinoma: HPV16 variants in different cervical cancer histologies, Int. J. Cancer 140 (2017) 2092, https://doi.org/10.1002/ijc.30636. –2100.
[154] M.A. De Boer, L.A.W. Peters, M.F. Aziz, B. Siregar, S. Cornain, M.A. Vrede, E. S. Jordanova, G.J. Fleuren, Human papillomavirus type 18 variants: histopathology and E6/E7 polymorphisms in three countries, Int. J. Cancer 114 (2005) 422–425, https://doi.org/10.1002/ijc.20727.
[155] M. Lizano, E. De la Cruz-Hernández, A. Carrillo-García, A. García-Carrancá, S. Ponce de Leon-Rosales, A. Dueñas-González, D.M. Hernández-Hernández, A. Mohar, Distribution of HPV16 and 18 intratypic variants in normal cytology, intraepithelial lesions, and cervical cancer in a Mexican population, Gynecol. Oncol. 102 (2006) 230–235, https://doi.org/10.1016/j.ygyno.2005.12.002.
[156] L.F. Xi, N.B. Kiviat, A. Hildesheim, D.A. Galloway, C.M. Wheeler, J. Ho, L. A. Koutsky, Human Papillomavirus Type 16 and 18 Variants: Race-Related Distribution and Persistence, JNCI, vol. 98, Journal of the National Cancer Institute, 2006, pp. 1045–1052, https://doi.org/10.1093/jnci/djj297.
[157] E.A. Lopera, A. Baena, V. Florez, J. Montiel, C. Duque, T. Ramirez, M. Borrero, C. M. Cordoba, F. Rojas, R. Pareja, A.M. Bedoya, G. Bedoya, G.I. Sanchez, Unexpected inverse correlation between Native American ancestry and Asian American variants of HPV16 in admixed Colombian cervical cancer cases, Infect. Genet. Evol. 28 (2014) 339–348, https://doi.org/10.1016/j.meegid.2014.10.014.
[158] K. Junes-Gill, L. Sichero, P.C. Maciag, W. Mello, V. Noronha, L.L. Villa, Human papillomavirus type 16 variants in cervical cancer from an admixtured population in Brazil, J. Med. Virol. 80 (2008) 1639–1645, https://doi.org/10.1002/jmv.21238.
[159] Z. Chen, M. Terai, L. Fu, R. Herrero, R. DeSalle, R.D. Burk, Diversifying selection in human papillomavirus type 16 lineages based on complete genome analyses, J. Virol. 79 (2005) 7014–7023, https://doi.org/10.1128/JVI.79.11.7014-7023.2005.
[160] V.R. DeFilippis, F.J. Ayala, L.P. Villarreal, Evidence of diversifying selection in human papillomavirus type 16 E6 but not E7 oncogenes, J. Mol. Evol. 55 (2002) 491–499, https://doi.org/10.1007/s00239-002-2344-y.
[161] A. Carvajal-Rodríguez, Detecting recombination and diversifying selection in human alpha-papillomavirus, Infect. Genet. Evol. 8 (2008) 689–692, https://doi.org/10.1016/j.meegid.2008.07.002.
[162] K.A. Lang Kuhs, D.L. Faden, L. Chen, D.K. Smith, M. Pinheiro, C.B. Wood, S. Davis, M. Yeager, J.F. Boland, M. Cullen, M. Steinberg, S. Bass, X. Wang, P. Liu, M. Mehrad, T. Tucker, J.S. Lewis, R.L. Ferris, L. Mirabello, Genetic variation within the human papillomavirus type 16 genome is associated with oropharyngeal cancer prognosis, Ann. Oncol. 33 (2022) 638–648, https://doi.org/10.1016/j.annonc.2022.03.005.
[163] A.A. Chen, T. Gheit, S. Franceschi, M. Tommasino, G.M. Clifford, Human papillomavirus 18 genetic variation and cervical cancer risk worldwide, J. Virol. 89 (2015) 10680–10687, https://doi.org/10.1128/JVI.01747-15.
[164] H. Arias-Pulido, C.L. Peyton, N. Torrez-Martínez, D.N. Anderson, C.M. Wheeler, Human papillomavirus type 18 variant lineages in United States populations characterized by sequence analysis of LCR-E6, E2, and L1regions, Virology 338 (2005) 22–34, https://doi.org/10.1016/j.virol.2005.04.022.
[165] Z. Chen, R. DeSalle, M. Schiffman, R. Herrero, R.D. Burk, Evolutionary dynamics of variant genomes of human papillomavirus types 18, 45, and 97, J. Virol. 83 (2009) 1443–1455, https://doi.org/10.1128/JVI.02068-08.
[166] L.F. Xi, M. Schiffman, L.A. Koutsky, A. Hulbert, S.-K. Lee, V. DeFilippis, Z. Shen, N.B. Kiviat, Association of human papillomavirus type 31 variants with risk of cervical intraepithelial neoplasia grades 2-3, Int. J. Cancer 131 (2012) 2300–2307, https://doi.org/10.1002/ijc.27520.
[167] H. Lou, J.F. Boland, H. Li, R. Burk, M. Yeager, S.K. Anderson, N. Wentzensen, M. Schiffman, L. Mirabello, M. Dean, HPV16 E7 nucleotide variants found in cancer-free subjects affect E7 protein expression and transformation, Cancers 14 (2022) 4895, https://doi.org/10.3390/cancers14194895.
[168] B. Zhu, Y. Xiao, M. Yeager, G. Clifford, N. Wentzensen, M. Cullen, J.F. Boland, S. Bass, M.K. Steinberg, T. Raine-Bennett, D. Lee, R.D. Burk, M. Pinheiro, L. Song, M. Dean, C.W. Nelson, L. Burdett, K. Yu, D. Roberson, T. Lorey, S. Franceschi, P. E. Castle, J. Walker, R. Zuna, M. Schiffman, L. Mirabello, Mutations in the HPV16 genome induced by APOBEC3 are associated with viral clearance, Nat. Commun. 11 (2020) 886, https://doi.org/10.1038/s41467-020-14730-1.
[169] C.J. Warren, K. Van Doorslaer, A. Pandey, J.M. Espinosa, D. Pyeon, Role of the host restriction factor APOBEC3 on papillomavirus evolution, Virus Evol. 1 (2015) vev015, https://doi.org/10.1093/ve/vev015.
[170] C.J. Warren, D. Pyeon, APOBEC3 in papillomavirus restriction, evolution and cancer progression, Oncotarget 6 (2015) 39385–39386, https://doi.org/10.18632/oncotarget.6324.
[171] E.C. Holmes, B.T. Grenfell, Discovering the phylodynamics of RNA viruses, PLoS Comput. Biol. 5 (2009), e1000505, https://doi.org/10.1371/journal.pcbi.1000505.
[172] A.S. Lauring, Within-host viral diversity: a window into viral evolution, Annu. Rev. Virol. 7 (2020) 63–81, https://doi.org/10.1146/annurev-virology-010320-061642.
[173] N.D. Grubaugh, K. Gangavarapu, J. Quick, N.L. Matteson, J.G. De Jesus, B. J. Main, A.L. Tan, L.M. Paul, D.E. Brackney, S. Grewal, N. Gurfield, K.K.A. Van Rompay, S. Isern, S.F. Michael, L.L. Coffey, N.J. Loman, K.G. Andersen, An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar, Genome Biol. 20 (2019) 8, https://doi.org/10.1186/s13059-018-1618-7.
[174] E.C. Holmes, The Evolution and Emergence of RNA Viruses, Oxford University Press, New York, 2009.
[175] L.A. Loeb, J.M. Essigmann, F. Kazazi, J. Zhang, K.D. Rose, J.I. Mullins, Lethal mutagenesis of HIV with mutagenic nucleoside analogs, Proc. Natl. Acad. Sci. USA 96 (1999) 1492–1497, https://doi.org/10.1073/pnas.96.4.1492.
[176] G. Koh, A. Degasperi, X. Zou, S. Momen, S. Nik-Zainal, Mutational signatures: emerging concepts, caveats and clinical applications, Nat. Rev. Cancer 21 (2021) 619–637, https://doi.org/10.1038/s41568-021-00377-7.
[177] L.B. Alexandrov, J. Kim, N.J. Haradhvala, M.N. Huang, A.W. Tian Ng, Y. Wu, A. Boot, K.R. Covington, D.A. Gordenin, E.N. Bergstrom, S.M.A. Islam, N. Lopez-Bigas, L.J. Klimczak, J.R. McPherson, S. Morganella, R. Sabarinathan, D. A. Wheeler, V. Mustonen, G. Getz, S.G. Rozen, M.R. Stratton, P.C.A.W. G. Consortium, The repertoire of mutational signatures in human cancer, Nature 578 (2020) 94–101, https://doi.org/10.1038/s41586-020-1943-3.
[178] J.-P. Vartanian, D. Guétard, M. Henry, S. Wain-Hobson, Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions, Science 320 (2008) 230–233, https://doi.org/10.1126/science.1153201.
[179] J.M. Kidd, T.L. Newman, E. Tuzun, R. Kaul, E.E. Eichler, Population stratification of a common APOBEC gene deletion polymorphism, PLoS Genet. 3 (2007) e63, https://doi.org/10.1371/journal.pgen.0030063.
[180] S.K. Pillai, J.K. Wong, J.D. Barbour, Turning up the volume on mutational pressure: is more of a good thing always better? (A case study of HIV-1 Vif and

APOBEC3), Retrovirology 5 (2008) 26, https://doi.org/10.1186/1742-4690-5-26.
[181] H.A. Sadler, M.D. Stenglein, R.S. Harris, L.M. Mansky, APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis, J. Virol. 84 (2010) 7396–7404, https://doi.org/10.1128/JVI.00056-10.
[182] M. Eigen, P. Schuster, The hypercycle: a principle of natural self-organization. Part A: emergence of the hypercycle, Naturwissenschaften 64 (1977) 541–565, https://doi.org/10.1007/BF00450633.
[183] M. Eigen, On the nature of virus quasispecies, Trends Microbiol. 4 (1996) 216–218, https://doi.org/10.1016/0966-842X(96)20011-3.
[184] M.J. Ferber, E.C. Thorland, A.A. Brink, A.K. Rapp, L.A. Phillips, R. McGovern, B. S. Gostout, T.H. Cheung, T.K.H. Chung, W.Y. Fu, D.I. Smith, Preferential integration of human papillomavirus type 18 near the c-myc locus in cervical carcinoma, Oncogene 22 (2003) 7233–7242, https://doi.org/10.1038/sj.onc.1207006.
[185] C. Bodelon, M.E. Untereiner, M.J. Machiela, S. Vinokurova, N. Wentzensen, Genomic characterization of viral integration sites in HPV-related cancers, Int. J. Cancer 139 (2016) 2001–2011, https://doi.org/10.1002/ijc.30243.
[186] L. Zhou, Q. Qiu, Q. Zhou, J. Li, M. Yu, K. Li, L. Xu, X. Ke, H. Xu, B. Lu, H. Wang, W. Lu, P. Liu, Y. Lu, Long-read sequencing unveils high-resolution HPV integration and its oncogenic progression in cervical cancer, Nat. Commun. 13 (2022) 2563, https://doi.org/10.1038/s41467-022-30190-1.
[187] D.E. Symer, K. Akagi, H.M. Geiger, Y. Song, G. Li, A.-K. Emde, W. Xiao, B. Jiang, A. Corvelo, N.C. Toussaint, J. Li, A. Agrawal, E. Ozer, A.K. El-Naggar, Z. Du, J. B. Shewale, B. Stache-Crain, M. Zucker, N. Robine, K.R. Coombes, M.L. Gillison, Diverse tumorigenic consequences of human papillomavirus integration in primary oropharyngeal cancers, Genome Res. 32 (2022) 55–70, https://doi.org/10.1101/gr.275911.121.
[188] R Core Team, A Language and Environment for Statistical Computing, 2018. https://www.R-project.org/.
[189] G. Cardone, A.L. Moyer, N. Cheng, C.D. Thompson, I. Dvoretzky, D.R. Lowy, J. T. Schiller, A.C. Steven, C.B. Buck, B.L. Trus, Electron Cryo-Microscopy of Human Papillomavirus Type 16 Capsid, 2014, https://doi.org/10.2210/pdb3J6R/pdb.
[190] G. Cardone, A.L. Moyer, N. Cheng, C.D. Thompson, I. Dvoretzky, D.R. Lowy, J. T. Schiller, A.C. Steven, C.B. Buck, B.L. Trus, Maturation of the human papillomavirus 16 capsid, mBio 5 (2014) e01104–e01114, https://doi.org/10.1128/mBio.01104-14.
[191] wwPDB consortium, S.K. Burley, H.M. Berman, C. Bhikadiya, C. Bi, L. Chen, L. D. Costanzo, C. Christie, J.M. Duarte, S. Dutta, Z. Feng, S. Ghosh, D.S. Goodsell, R. K. Green, V. Guranovic, D. Guzenko, B.P. Hudson, Y. Liang, R. Lowe, E. Peisach, I. Periskova, C. Randle, A. Rose, M. Sekharan, C. Shao, Y.-P. Tao, Y. Valasatava, M. Voigt, J. Westbrook, J. Young, C. Zardecki, M. Zhuravleva, G. Kurisu, H. Nakamura, Y. Kengaku, H. Cho, J. Sato, J.Y. Kim, Y. Ikegawa, A. Nakagawa, R. Yamashita, T. Kudou, G.-J. Bekker, H. Suzuki, T. Iwata, M. Yokochi, N. Kobayashi, T. Fujiwara, S. Velankar, G.J. Kleywegt, S. Anyango, D. R. Armstrong, J.M. Berrisford, M.J. Conroy, J.M. Dana, M. Deshpande, P. Gane, R. Gáborová, D. Gupta, A. Gutmanas, J. Koča, L. Mak, S. Mir, A. Mukhopadhyay, N. Nadzirin, S. Nair, A. Patwardhan, T. Paysan-Lafosse, L. Pravda, O. Salih, D. Sehnal, M. Varadi, R. Vařeková, J.L. Markley, J.C. Hoch, P.R. Romero, K. Baskaran, D. Maziuk, E.L. Ulrich, J.R. Wedell, H. Yao, M. Livny, Y.E. Ioannidis, Protein Data Bank: the single global archive for 3D macromolecular structure data, Nucleic Acids Res. 47 (2019), https://doi.org/10.1093/nar/gky949. D520–D528.
[192] C.W. Nelson, L.H. Moncla, A.L. Hughes, SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data, Bioinformatics 31 (2015) 3709–3711, https://doi.org/10.1093/bioinformatics/btv449.
[193] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford University Press, New York, NY, 2000.
[194] S. García-Vallvé, Á. Alonso, I.G. Bravo, Papillomaviruses: different genes have different histories, Trends Microbiol. 13 (2005) 514–521, https://doi.org/10.1016/j.tim.2005.09.003.