Detection of Viral RNA Splicing in Diagnostic Virology

Vladimir Majerciak and Zhi-Ming Zheng

Introduction

Diagnostic virology is the process in which the viral etiologic cause of infection is identified from a patient's clinical sample. In the past, diagnostic virology relied on three classical techniques to make this diagnosis of viral infection: (a) virus isolation by direct virus cultivation, (b) viral antigen detection, and (c) indirect detection of virus-specific antibodies. While the remaining important tools are utilized by diagnostic virology laboratories today, these techniques are time-consuming and require specific reagents/methods such as cultivation media, cell or tissue cultures, antibodies, or purified antigens. In the past several decades, the number of new molecular-based methods increased rapidly and became widely used in diagnostic virology laboratories. The core of these techniques constitutes of techniques based on nucleic acid detection by specific amplification, hybridization, and/or sequencing (reviewed in [1]). The majority of these nucleic acid-based diagnostic methods are simple, speedy, sensitive, and specific and thus meet the gold "four-S standard" for their application in any diagnostic laboratory. The methods are simple and speedy because only a specific primer pair and a PCR machine are required by the laboratory, and identification of a viral pathogen takes only a few hours. They are sensitive and specific and require only a small amount of patient's clinical specimen to detect a specific nucleotide sequence region. In general, these techniques can be used to detect almost all types of viral pathogens and can even identify multiple viral pathogens or their variants at the same time. In this chapter, we will focus on detection of viral RNA splicing as a new tool for diagnostic virology.

V. Majerciak · Z.-M. Zheng (✉)
Tumor Virus RNA Biology Section, RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD, USA
e-mail: zhengt@exchange.nih.gov

Y.-W. Tang, C. W. Stratton (eds.), *Advanced Techniques in Diagnostic Microbiology*, https://doi.org/10.1007/978-3-319-95111-9_15

Principle of RNA Splicing

Definition of RNA Splicing

RNA splicing was discovered 40 years ago by Susan Berget [2] and Louise Chow [3]; these investigators mapped adenovirus transcription and identified intervening sequences (introns) in type 2 adenovirus primary transcripts. Subsequently, RNA splicing was recognized as an essential nuclear event for mammalian gene expression and for virus replication of almost all DNA viruses and some RNA viruses. Most mammalian genes consist of multiple segments called exons which are separated by noncoding or intervening sequences named introns. Genes which are composed of exons and introns are "split" genes. After transcription, a nascent or primary transcript (pre-mRNA) contains both exons and introns. The introns are removed from the pre-mRNA by a molecular process called "RNA splicing" resulting in production of spliced mature mRNA. RNA splicing takes place both in coding as well as in noncoding primary transcripts. RNA splicing has been considered a posttranscriptional event; however, recent studies have demonstrated that RNA splicing often occurs co-transcriptionally [4, 5]. Only those transcripts which are fully processed are eventually exportable from the nucleus to the cytoplasm for protein synthesis.

Molecular Mechanism of RNA Splicing

All introns are defined by three *cis*-elements: a 5' splice site (donor site), a branch point, and a 3' splice site (acceptor site) with a polypyrimidine track immediately upstream (Fig. 1a). These *cis*-elements allow cellular splicing machinery to recognize and remove the intron from pre-mRNA. Most mammalian introns start with GU dinucleotide on its 5' end and an AG dinucleotide on its 3' end ("GU-AG" introns). The GU-AG pairs are conserved sequences and define the exon-intron boundaries. Introns with an AU on its 5' end and an AC on its 3' end are rare, and this set of the introns are known as "AU-AC" introns [6, 7]. The presence of splice

Fig. 1 (continued) These initial recognitions of the intron elements by the components of RNA splicing machinery are essential for spliceosome formation on a pre-mRNA, leading to intron removal. (**b**) RNA splicing is catalyzed by two transesterification reactions. During RNA splicing, the intron (gray) between two exons (black) is removed by two transesterification reactions. First, the intron is recognized by cellular splicing machinery via splicing factors binding to intron-specific sequences as described in (**a**). Splicing factors carry out the first transesterification reaction between the branch point and the 5' donor site, resulting of an RNA cleavage at the 5' donor site and releasing exon 1 and formation of a lariat intermediate. Subsequently, the free 3'end OH group of exon 1 attacks the 5' end phosphate of exon 2 and joins with the 5' end of exon 2 via the second transesterification reaction to form a mature mRNA. Intron is removed in the form of lariat structure and quickly degraded

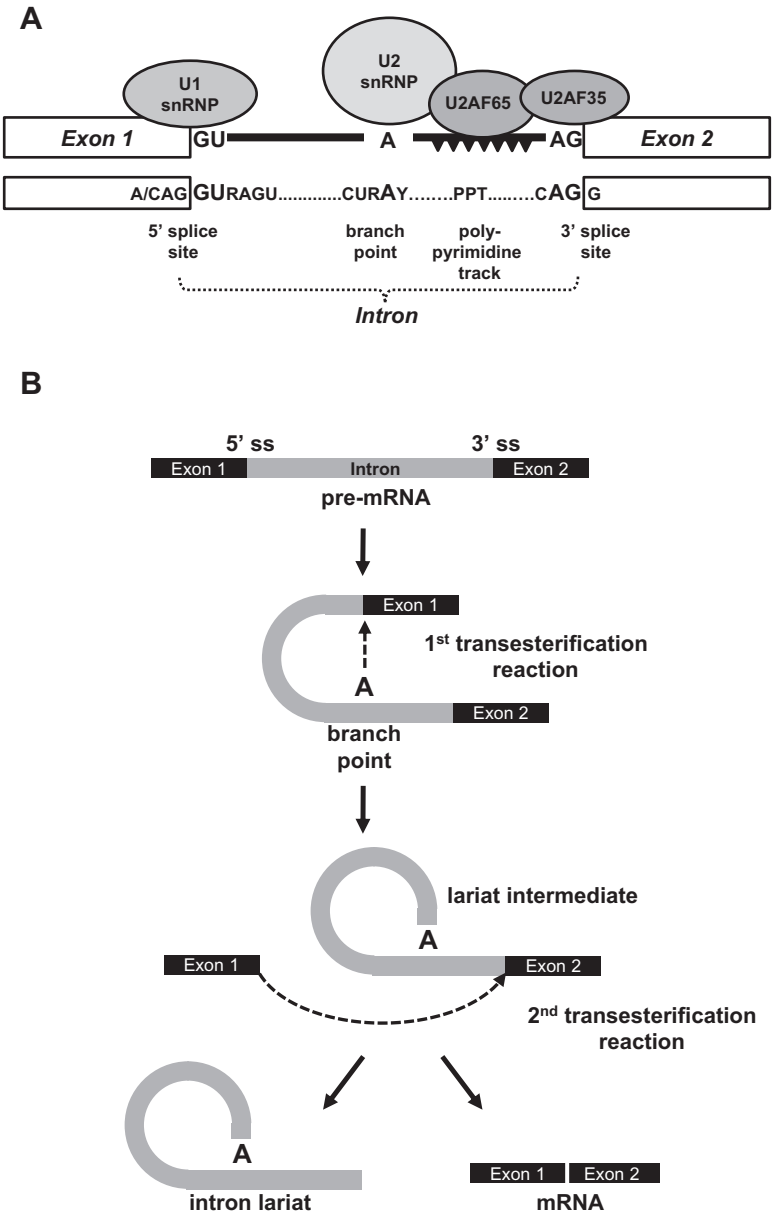


Fig. 1 Pre-mRNA structure and splicing reactions. (**a**) Structure of a pre-mRNA containing an intron (solid line) and two exons (empty boxes). An intron is defined by several specific sequence motifs which allow intron recognition by cellular splicing machinery. A 5' splice site or donor site GU at the intron 5' end is recognized by U1 small nuclear ribonucleoprotein (snRNP), a major component of the cellular splicing machinery. A 3' splice site or acceptor site at the intron 3' end consists of an AG dinucleotide, an upstream polypyrimidine track (PPT) and a further upstream branch point, which are recognized correspondently by U2AF35, U2AF65, and U2 snRNP.

sites is not sufficient for intron definition. All introns must contain an additional element called a "branch point," which is located 20–50 nts upstream from the 3'splice site, and have a consensus sequence CU(A/C)A(C/U) where A is a most conserved base. The sequence between the branch point and the acceptor site is a run of 15–40 pyrimidines (mostly U) and is referred to as a "polypyrimidine track."

RNA splicing is catalyzed by cellular splicing machinery, which consists of the following components: (1) small nuclear ribonucleoproteins (snRNPs, U1, U2, U4, U5, and U6) and (2) splicing factors. During the initial step, snRNP, U1, and U2 recognize the intron sequences at the 5' splice site and the branch point via complementary base-pairing involving the U2 accessory proteins U2AF65 and U2AF35, which associate with the polypyrimidine track and 3'splice site, respectively (Fig. 1a). Intron recognition is a signal for the formation of a large protein complex called the "spliceosome" where intron removal takes place [8, 9] by two transesterification reactions (Fig. 1b). First, the primary transcript is cleaved at the intron 5' end (5' splice site) to leave the upstream exon free; this step is followed by branching of the cleaved intron 5' end to the branch point (usually A) to create a looped structure named "lariat intermediate". In the second step, the hydroxyl group of the free exon attacks the intron 3' splice site leading to the 3' splice site cleavage and lariat formation. Simultaneously, a covalent bond is created between two exons to create a mature mRNA. In general, the lariats are quickly released from the spliceosome and degraded in the nucleus.

The efficiency of RNA splicing is regulated at multiple levels, and both RNA *cis*-elements and cellular splicing factors play major roles in the regulation of RNA splicing. As described above, the level of conservation of the sequences at the splice sites and the branch point will affect the strength of the binding of core splicing factors and thus determine the splicing efficiency. RNA splicing is also modulated by a large family of cellular splicing factors containing serine-arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs). Most of the splicing factors that are differentially expressed in a specified tissues and/or in a development stage are the RNA-binding proteins, which bind to specific RNA *cis*-element (splicing enhancers or silencers) located within introns and exons [10]. It is now well documented that splicing factors binding to the *cis*-elements either increase or decrease RNA splicing efficiency depending on the type of splicing factors, the positions of the binding sites, and the overall spliceosome composition [11, 12]. Current studies demonstrate that in addition to splicing factors, other processes such as RNA polymerase rate and chromatin structure also affect RNA splicing [13, 14].

Alternative RNA Splicing

Although all introns in a pre-mRNA could be constitutively spliced out, and all exons are supposedly included in a mature mRNA, there are many examples where an RNA splice site may be not selected constitutively, but instead skipped, during

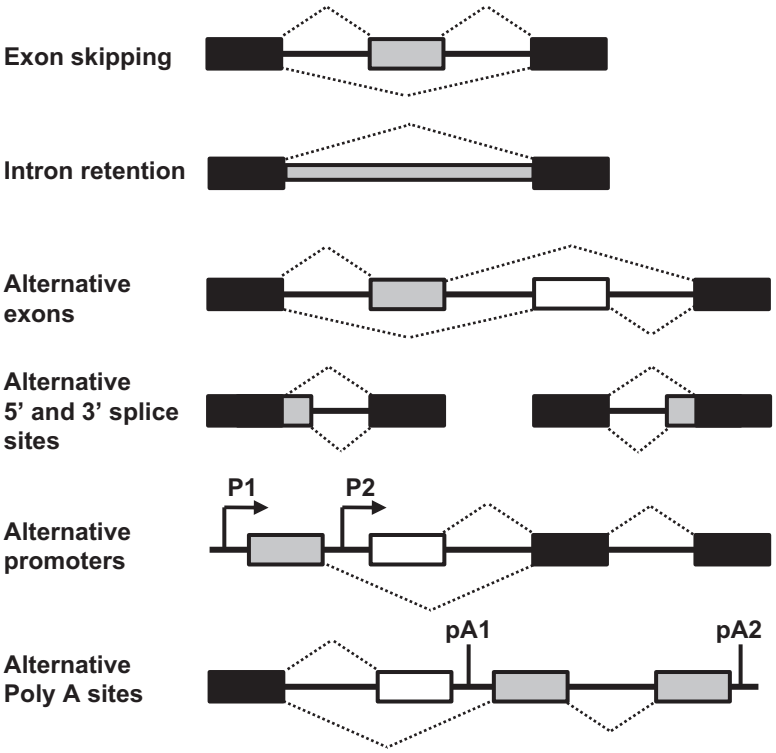


Fig. 2 Alternative RNA splicing. Alternative RNA splicing allows production of multiple splicing isoforms from a single pre-mRNA species. Constitutive exons (black boxes) are included in all splicing products, while alternative exons (gray and white boxes) are either included or excluded in various isoforms of mature mRNAs. Major forms of alternative RNA splicing include exon skipping, intron retention, usage of alterative exons, and usage of alternative splice sites. In addition, usage of alternative promoters (P1 or P2) or polyadenylation sites (pA1 or pA2) may affect exon composition in a final mature transcript.

RNA splicing. Consequently, this alternative RNA splicing leads to the production of RNA isoforms with different exon compositions and production of different protein isoforms. In general, there are four major classes of alternative RNA splicing (Fig. 2), including exon skipping, intron retention, usage of alternative exons, and usage of alternative 5' and alternative 3' splice site [15]. In addition, alternative promoter or polyadenylation usage can further complicate the alternative RNA splicing. The reason why some of the exons or introns in a pre-mRNA are alternatively spliced is either because of the presence of weak or suboptimal splice signals in the pre-mRNA or due to the lack of a particular splicing factor. It has been noticed that usage of weak splice sites is highly dependent on auxiliary splicing factors binding to the regulatory *cis*-elements. Since the expression of these factors is variable from cell to cell and from tissue to tissue, alternative RNA splicing is often

associated with a specific cell type, tissue, or stage of cell differentiation [16]; this results in the production of various isoforms of transcripts from one gene [17]. Current analysis also revealed that the number of genes with alternative RNA splicing increases with complexity of the organism and ranges from 0.05% in yeast up to 66.8% in human. Thus, alternative RNA splicing is the driving force behind the complexity of the proteome in higher organisms in addition to the total number of the mapped genes.

Molecular Methods for Detection of RNA Splicing

The mRNA generated by RNA splicing is different from its pre-mRNA. First, mRNA is smaller in size than its pre-mRNA due to RNA splicing which removes the introns found in the pre-mRNA. In contrast, the pre-mRNA is not only larger than the spliced mRNA but also has the same size as its DNA template. Second, mRNA contains exon-exon junctions with the sequence not present in DNA or its primary transcript, allowing designing primers or probes to specifically identify a particular mRNA isoform due to alternative RNA splicing. Although an alternatively spliced mRNA may translate a truncated protein, which could be detectable with a specific antibody, the molecular techniques based on detection of nucleic acids are more commonly used to detect RNA splicing.

Northern Blot

Northern blot is one of the oldest techniques used to detect RNA splicing. First, RNA molecules isolated from samples are separated based on their size by electrophoresis in agarose or polyacrylamide gel. After transfer to a nitrocellulose or nylon membrane, the individual RNA transcripts are detected by an antisense probe specific for the detecting RNA. The probes used for the Northern blot are usually labeled with ^{32}P isotope, enzyme (e.g., alkaline phosphatase), digoxigenin (DIG), or biotin and can be derived from a constitutive exon or an exon-exon junction (Fig. 3a). Constitutive exon-based probes would detect all spliced RNA isoforms and the remaining, unspliced pre-mRNA; these exon-based probes are recommended when the size difference between spliced RNA isoforms and unspliced pre-mRNA is sufficient enough to be separated. If the size difference is too small for two RNA isoforms to be distinguished, an exon junction probe can be used to specifically detect a spliced product. In addition, a specific probe from an alternative exon or intron can be also designed for detection of individual splicing isoforms derived from exon/intron inclusion.

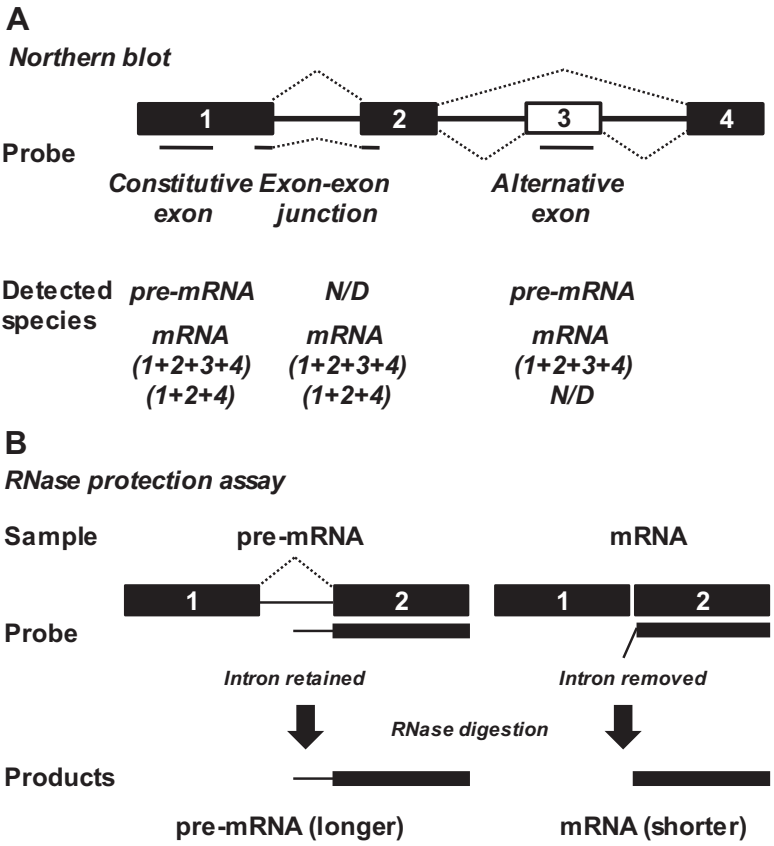


Fig. 3 Detection of RNA splicing products by Northern blot and RNase protection assays (RPA). (**a**) Diagram of a pre-mRNA with constitutive exons (black boxes), introns (solid lines), and an alternative exon (empty box). Solid lines below indicate positions of antisense probes commonly designed for Northern blot to detect products by each probe. N/D – not detectable. (**b**) Detection of spliced RNA products by RPA. Diagram shows an antisense riboprobe spanning over an intron region (a solid thin line) between two exons (black boxes or solid thick lines) and possible detection products. As an RNase used in the assay digests a single-stranded RNA region only, an RNA region base-paired with an antisense probe uniformly labeled with isotope ^{32}P will be protected from RNase digestion. In this diagram, the probe remains intact when binding to the pre-mRNA, while the probe binding to the spliced mRNA (lack of the intron) will cause the digestion of the probe intron region (single stranded) resulting in production of smaller protected products corresponding to each exon

RNase Protection Assay

The RNase protection assay (RPA) requires ^{32}P-labeled single-stranded antisense RNA probes complementary to the transcripts of interest. The prepared probe(s) is consequently hybridized with sample RNA to form an RNA-RNA hybrid.

Unhybridized single-stranded RNA is then removed by RNases A and T1, which digest single-stranded RNA only. The protected RNA fragments are separated in the gel by electrophoresis, and their sizes are determined by molecular markers. To distinguish a spliced RNA product, the probe should contain at least one partial intron region that will be digested from the probe due to the lack of the intron sequence in the spliced mRNA. As a result, the probe protected by the corresponding exon regions of the detecting mRNA is shorter and will run faster in the gel (Fig. 3b). In general, RPA is more sensitive than Northern blot in detection of RNA splicing.

Both Northern blot and RPA are commonly used in research laboratories. Their main advantage is high specificity. However, both methods are very laborious and low-throughput requiring isolation of large amount (usually a few micrograms) of the total RNA from samples and preparation of specific probes often labeled with radioisotopes, which limits their use in clinical diagnostics.

RT-PCR

The RT-PCR (reverse transcription-polymerase chain reaction) is one of the most commonly used methods for detection and quantification of RNA molecules. During RT-PCR, RNA transcripts are converted into complementary DNA (cDNA) by reverse transcription using random hexamers or oligo-dT- (a short sequence of deoxy-thymidine nucleotides) or transcript-specific primers. The resulted cDNA is then used as a template in subsequent PCR with a pair of transcript-specific primers.

In principle, the detection of spliced RNA transcripts by RT-PCR depends on amplicon selection and primer design. The most common approach is the amplification over the intron regions by a set of primers in flanking exons. The resulted RT-PCR products vary in sizes depending on how the detecting transcript is spliced. A larger product than the predicted size may represent an unspliced pre-mRNA, or contaminating genomic DNA. The latter can be determined by a minus RT amplification (PCR). A spliced mRNA always gives a smaller RT-PCR product than its pre-mRNA due to removal of intron sequences by RNA splicing (Fig. 4a). Another approach is to specifically amplify a spliced product by using an exon junction primer because the sequence at exon-exon junction is not present in pre-mRNA, nor in genomic DNA. Similarly, a primer based on an alternative exon would amplify only the transcript with the inclusion of that exon (Fig. 4b). After amplification, the size and amount of RT-PCR products are analyzed by gel electrophoresis. Because of nonlinear nature of PCR amplification, classical PCR only provides semiquantitative data on the abundance of various spliced RNA isoforms.

Introduction of real-time quantitative RT-PCR (RT-qPCR) with a broad (10 [7]) dynamic range has significantly improved the sensitivity of RT-PCR. Because of its high sensitivity, real-time RT-qPCR is able to detect and amplify RNA directly from

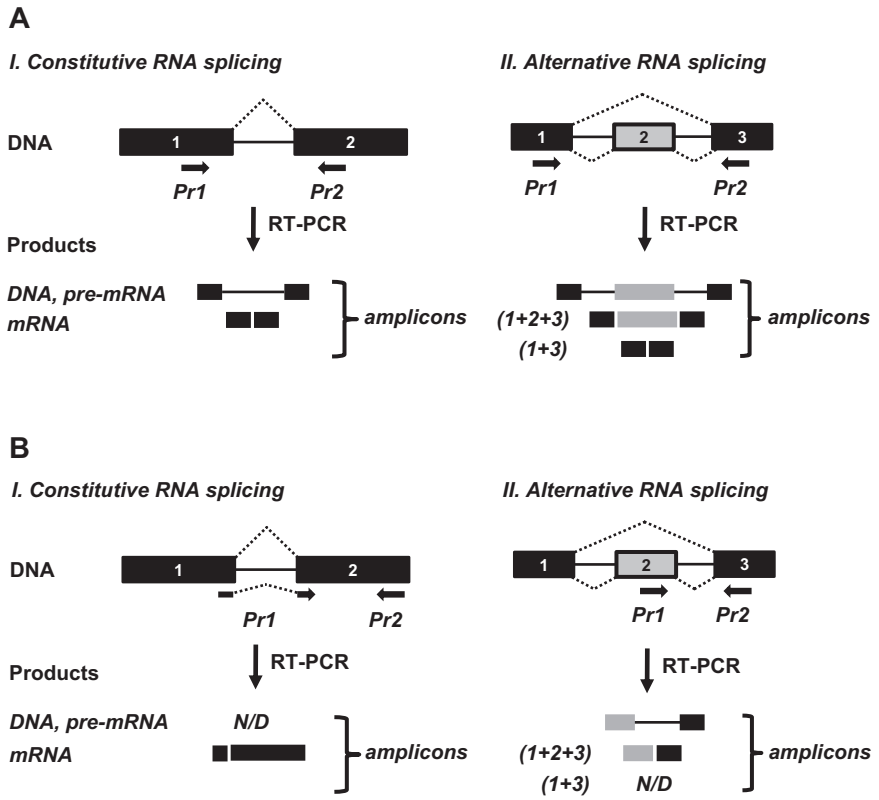


Fig. 4 Detection of spliced RNA products by RT-PCR. Intron removal from a pre-mRNA could be detected by RT-PCR based on the size of a product (**a**) or in a selective amplification (**a** and **b**). (**a**) A pair of primers (*Pr1* and *Pr2*) used for amplification are derived from two constitutive exons (black boxes) over an intron (solid line), and the amplified mRNA product (spliced) will be smaller in size than the products amplified from template DNA or unspliced pre-mRNA. In an alternative RNA splicing assay, a pair of primers is derived from exon 1 and exon 3 spanning over alternative exon 2 (gray box). Alternative RNA splicing will result in multiple-spliced RNA products of various sizes amplified by RT-PCR. (**b**) Specific amplification of spliced RNA products during constitutive or alternative RNA splicing. The specific splicing products could be obtained by using a set of primers in which one represents a splicing junction (*Pr1*). Because of the lack of this sequence in unspliced pre-mRNA or DNA, only the spliced product will be selectively amplified. A selective amplification could be also used to amplify by RT-PCR a specific RNA isoform derived by alternative RNA splicing by using one primer in an alternative exon (exon 2) in combination with a primer in exon 3. In this case, only the spliced product with exon 2 inclusion will be selectively amplified by RT-PCR

a single cell without RNA extraction. In addition, it automates the quantification and does not require electrophoretic separation of RT-PCR products.

Currently there are four major detection chemistries widely used in real-time RT-PCR: SYBR green (Molecular Probes), *Taq*Man™ probes [18], Molecular Beacons [19], and Scorpions™ probes [20]. The principles of these different

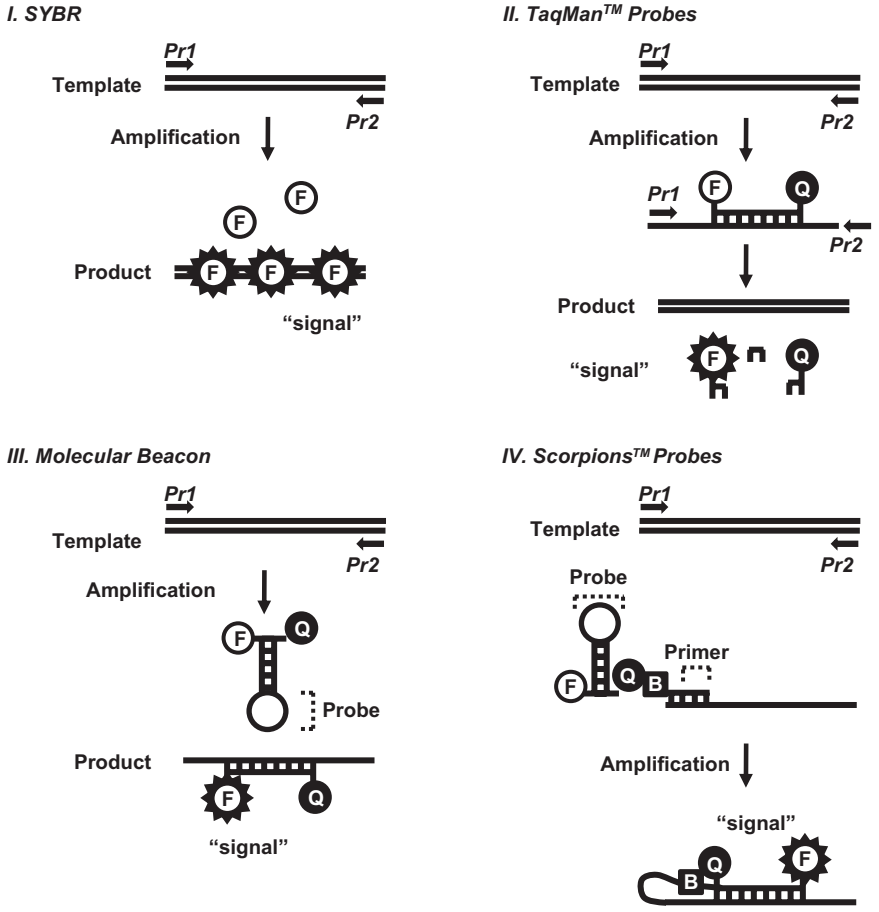


Fig. 5 Major chemistries of real-time PCR. PCR product detection by four commercially available chemistries in real-time PCR: SYBR green (*I*), *Taq*Man™ (*II*), Molecular Beacon (*III*), and Scorpions™ probes (*IV*). The mechanism of each chemistry is described in detail in the chapter section III-in the corresponding chapter. The sections are not numbered anymore. Double lines represent a double-stranded DNA template generated by RT-PCR, while a single line is correspondent to single-stranded DNA. Horizontal arrows mark primer (*Pr1* and *Pr2*) positions. *F* stands for quenched "fluorophore moiety" and could be present in two stages: quenched nonfluorescent stage (empty circles) or in activated stage (empty circles with spikes) with generation of detectable signal. *Q* in black circles represents "quencher." Scorpion probes contain amplification stop sequences named "blocker" (*B* in black box)

detection chemistries are described in Fig. 5. As of today, SYBR green and *Taq*Man™ probes represent the most common detection chemistries. SYBR green is a fluorescent dye which has low fluorescence when in solution; however, it becomes highly fluorescent upon binding to double-stranded DNA. On the other hand, *Taq*Man™ probes, Molecular Beacons, and Scorpions probes employ fluores-

cence resonance energy transfer or Förster resonance energy transfer (FRET, also known as resonance energy transfer [RET] or electronic energy transfer [EET]) to generate a fluorescent signal. The FRET combines the donor fluorescent dye (fluorophore) with a nonfluorescent quenching moiety (quencher). When the fluorescent dye is in close proximity of the quencher, the quencher molecule absorbs the energy and thus blocks fluorescence emission from the fluorophore when excited by light. TaqMan probes are 18–22 bp oligonucleotide probes that are labeled with a reporter fluorophore at the 5' end and a quencher at the 3' end; these are thus in close proximity. Each probe is complementary to a region in the middle of the detecting target between the two primers used in the PCR reaction. When *Taq* polymerase extends the primer to synthesize the nascent strand, the 5' to 3' exonuclease activity of the *Taq* polymerase degrades the *Taq*Man™ probe annealed to the targeted region and releases the fluorophore from *Taq*Man™ probe and thereby breaks the close proximity to the quencher. As a result, the fluorophore when excited by cycler's light emits fluorescence, which marks the presence of PCR product. The method determines the amount of product by generation of fluorescent signal, which is measured in "real time" during the entire cycle time of the RT-PCR reaction, which allows the calculation of the amount of PCR product after each amplification cycle. Similar to *Taq*Man™, the Molecular Beacon and the Scorpions™ both use probes to detect specific PCR product. However, instead of probe degradation, the signal is generated by physical separation of the fluorophore and quenching moieties after hybridization of the specific probe to the PCR product during amplification (Fig. 5).

Each detection chemistry has its own advantages and disadvantages, which need to be considered during experimental design. SYBR green represents a simple, easy-to-use method and is the most economical real-time RT-PCR method. The disadvantage of SYBR green is its binding non-specifically to any DNA including primer dimers and non-specific PCR products; therefore, it is not useful for multiplex amplification of several products in the same reaction. In contrast, *Taq*Man™, Molecular Beacons, and Scorpions™ probes specifically detect only a PCR product complementary to probe sequence enabling to distinguish specific from non-specific products. The disadvantage of these detection chemistries is that each PCR product requires synthesis of its own specific probe, which increases the cost per reaction. On the other hand, labeling individual probes with fluorophores of different emission spectrums allows multiplexing with simultaneous detection of several products and thus reduces the cost and labor.

The usage of real-time RT-PCR for splicing detection requires special considerations. Since real-time RT-PCR techniques omit electrophoretic separation, the spliced product cannot be distinguished based on size. Therefore, it is important that only the desired product is amplified. In this case, the usage of SYBR green chemistry is the most challenging due to the lack of specificity. Probe-based methods provide higher specificity due to probe hybridization to selected sequences that are not present in non-specific products. To detect only the desired spliced product, the probe and primer should be complementary to a specific exon-exon junction or to an alternatively spliced region. Many commercial manufacturers of synthetic oligos

provide free online tools for the design of the most optimal and specific primer pairs based on the provided template sequence. Several other factors must be considered when using RT-PCR-based techniques in diagnostics; these include (1) RNA sample quality and preparation, (2) *Taq* polymerase inactivating contaminants in clinical samples, and (3) amplification bias. Other considerations are false positivity and PCR cross-contamination.

Splicing Microarrays

DNA microarrays (also known as DNA chips) are composed of large number of probes (often several thousand probes) spotted on very small area in 2D format on a solid surface (glass or plastic). The probes represent DNA oligos of various length and chemistry. Each probe has specific DNA sequence allowing detection of corresponding DNA with a complementary sequence. Currently, there are two major technologies involved with DNA arrays and the manufacturing of a microarray: (a) direct synthesis of probes on the array and (b) printing arrays from a library of presynthesized probes. Each DNA microarray allows rapid profiling of large number of DNA molecules at the same time. Today DNA microarrays are widely used to study gene expression profiling and RNA posttranscriptional modifications including RNA splicing [21].

The analysis of RNA transcripts by DNA microarrays requires a conversion of RNA samples to DNA by reverse transcription, following amplification and labeling with fluorescent dye. After labeling, the samples are hybridized with the probes on the array. Unbound samples are washed away, and the fluorescent signal is captured and analyzed by the microarray reader (Fig. 6a). The intensity of the fluorescent signal corresponds to the number of bound molecules and thus allows the determination of the level of RNA in the original sample by use of a mathematical algorithm.

There are two different approaches to probe design for the study of RNA splicing using DNA microarrays: (1) tiling and (2) exon arrays [22] (Fig. 6b). In tiling arrays, the set of overlapping probes cover the full length of the nascent primary transcript including exons and introns. The analysis of fluorescence for each probe allows the identification of exons and introns based on the difference in signal intensity (Fig. 6b*i*). The advantage of tiling arrays is their ability to identify known splice

Fig. 6 (continued) (*ii*) Detection of RNA splicing by co-hybridization of two probes labeled with acceptor (*A* in circle) or donor (*D* in circle) fluorophores binding to exonic regions (black boxes) flanking the intervening intron (solid line). When bound to unspliced transcript, the binding sites of two probes are separated by intron preventing energy transfer by FRET, and thus, no signal is generated. After intron removal by RNA splicing, two probes are bought to proximity for FRET to occur. The energy transfer from the donor to the acceptor leads to excitation of the acceptor fluorophore (A in circle with spikes) and generation of detectable signal. The diagram is modified from Blanco and Artero [31]

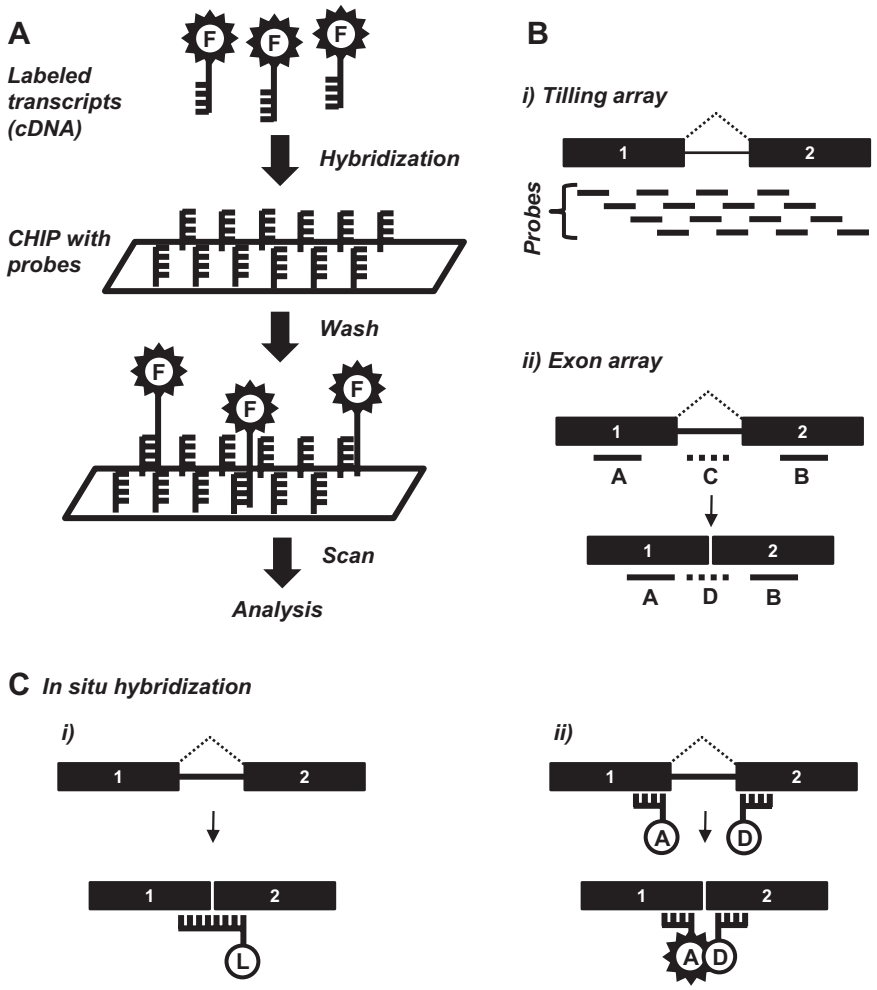


Fig. 6 Splicing microarrays and in situ hybridization. (**a**) A workflow for microarray assay. First, sample RNAs are converted into cDNAs simultaneously labeled with specific fluorophore *(F)*. The labeled cDNAs are hybridized with oligonucleotide probes attached to a solid surface. The unbound cDNAs are washed away, and the remaining fluorescence signals resulted from specific hybridization are collected by an array-scanning device and analyzed. (**b**) Microarrays in RNA splicing detection. (*i*) Tiling arrays represent a large set of overlapping probes (short solid lines) to cover the selected genes. Because an mRNA has no intron after RNA splicing, introns (lines) can be detected with drop on signal intensity when compared to neighboring exons (solid boxes). (*ii*) Exon arrays consist of combination of two types of probes: probes binding to exon regions (short solid lines) and probes binding to intron or exon-exon junction (short dashed lines). Splicing events are calculated by analysis of signal intensity between exon, intron, and junction probes. (**c**) Detection of spliced transcript by in situ hybridization. (*i*) Probe spanning over the exons (solid boxes) junction specifically binds to spliced product but not to unspliced pre-mRNA containing intron (solid line) or genomic DNA. The probe detection depends on type of labeling (L in empty circle) including isotope, biotin, digoxigenin, fluorophore, or others.

events as well as new splice events. Therefore, the tiling arrays are often used as discovery tools. The disadvantages are the requirement of a large amount of probes, which results in time-consuming data analysis. The exon arrays are more commonly used but require the knowledge of splicing events. Several types of probes hybridizing to flanking exons, intron, and exon-exon junctions are designed to detect each splicing event (Fig. 6b*ii*). The fluorescence intensity is detected for each probe, and a mathematical model is applied to determine the occurrence of splicing event. The advantage of exon arrays is the smaller number of probes required, which means simpler data analysis. However, the exon arrays detect only known or predicted splicing variants. Due to their large capacities, the exon arrays can be designed to detect splicing in multiple viral pathogens simultaneously.

In Situ Hybridization

Tissue sections historically represent an important tool for the diagnosis of pathological changes during viral infection as well as the detection of viral pathogens at the cellular level. There are two major types of tissue sections: frozen and formalin-fixed, paraffin-embedded (FFPE). Both are routinely used for the detection of viral antigens by various types of staining, but their use in the detection of nucleic acids including spliced transcripts is still relatively rare. The improved sensitivity of current nucleic acid isolation and amplification techniques allows the recovery of nucleic acid from tissue sections for further analysis by PCR and RT-PCR with selective isolation of only the cells of interest by the use of laser capture microdissection to add an additional level of specificity [23, 24]. However, the detection of nucleic acids by in situ hybridization (ISH) directly on tissue sections can provide additional information about gene expression linked with spatial distribution of specific RNA transcripts in a morphological context often at the cellular or even subcellular level. In the past, the nucleic acid molecules including RNA transcripts by ISH were detected by DNA probes labeled with radioisotope (^{35}S, ^{33}P, ^{3}H) [25], which were later replaced by nonradioactive DNA probes labeled with biotin or digoxigenin and detected by chromogenic methods using enzyme-labeled antibodies (CISH) [26]. Labeling probes specific for different transcripts with different fluorophores (FISH) allows detection of multiple targets at the same tissue section. However, the sensitivity was always a limiting factor of ISH techniques. This was caused mainly by the use of DNA probes, which suffer from low affinity to complementary RNA targets and are sensitivity to degradation of RNA-DNA hybrids by RNase H. Development of tyramide signal amplification (TSA) has dramatically improved the sensitivity of DNA probes [27]. Further improvement was seen with the introduction of locked nucleic acid (LNA) and peptide nucleic acid (PNA) probes with high affinity to RNA molecules and resistance to RNase H degradation [28–30]. Detection of RNA splicing by ISH requires a probe to specifically bind

only to spliced mRNA without binding to unspliced pre-mRNA or to the genomic DNA in the sample. Historically, this was achieved by designing a probe over exon-exon junction containing sequences present only in spliced transcripts (Fig. 6c*i*). Another approach in the detection of spliced transcripts by IHS is using co-hybridization of two probes labeled with donor and acceptor fluorophore and the generation of signals by FRET. In principle, each splicing event is monitored by a set of two probes complementary to exonic sequences flanking an intron region. One probe carries a fluorophore acceptor, while the second probe is labeled by a fluorophore donor. When probes bind to genomic DNA or unspliced nascent transcript, their binding sites are separated by intron regions resulting in the distance between the donor and acceptor being too big for the two fluorophores to engage in FRET. However, intron removal by splicing brings the probe binding sites to a proximity close enough for FRET to occur resulting in generation of measurable fluorescence [31] (Fig. 6c*ii*).This results in high specificity and low background. Using a set of probes with different fluorophores allows detection of multiple spliced transcripts or various spliced isoforms of the transcript. In summary, ISH hybridization methods provide a useful tool for the investigation of the distribution not only of protein-encoding transcripts but also of the rapidly growing number of virus-encoded noncoding RNAs, which their role in viral pathogenesis often remains elusive [32]. In situ hybridization methods could be especially suitable in retrospective analysis of archived samples in collections.

RNA-seq

Next-generation sequencing (NGS) represents a new generation of analytical tools for genome and transcriptome analysis [33]. This method is based on generation of a large amount of short sequences in parallel sequencing reactions. Advantages of NGS are the requirement of less amount of the initial sample, deep coverage, and nucleotide resolution. NGS also does not require any previous knowledge of the detecting sequence. Currently the main platforms are the Illumina HiSeq 2500 or HiSeq 3000/HiSeq4000 for high-throughput sequencing and generating the sequencing reads of various lengths.

Sequencing of RNA samples converted to cDNA is called *RNA-seq*. *RNA-seq* provides a comprehensive picture of whole genome transcriptome and has been successfully used for analysis on gene expression and posttranscriptional processing including RNA splicing. However, NGS may be costly and time-consuming as well as requiring sophisticated data analysis; this currently makes NGS less suitable for clinical diagnostics. However, *RNA-seq* does not require any prior knowledge of detecting sequence composition and therefore allows detecting unknown or unpredicted RNA sequences. This may be especially beneficial in discovery of new pathogens including viruses [34]. In addition, *RNA-seq* instantly analyzes a transcriptome including spliced transcripts in any type of cell or tissue.

RNA Splicing in Clinical Virology

RNA splicing does not occur in prokaryotes and is a hallmark of the eukaryotic gene expression. In eukaryotes the number of genes which undergo splicing varies highly from organism to organism, with only about 5% of all genes being spliced in yeasts to 95% in human [35, 36]. Viruses as intracellular parasites replicate inside of host cells and hitchhike many cellular processes for their replication including RNA splicing. By using constitutive and/or alternative RNA splicing, most of DNA viruses and some of RNA viruses increase the complexity of their proteome without the requirement of additional genetic materials.

Detection of spliced viral mRNAs in clinical samples would provide several benefits. While detection of viral genomes in clinical samples indicates virus infection, the result does not provide information about the stage and dynamic of the virus infection. In many cases the progress of viral replication can be assumed from changes in viral load, but this approach requires multiple sampling during infection and varies between individuals. One major advantage for detection of spliced viral transcripts is that viral RNA splicing reflects viral gene expression and thus indicates active viral infection, providing important information about the status of infection without requiring multiple sampling. The production of viral transcripts and their RNA splicing products are often the first sign of virus replication detectable before the increase of viral load or occurrence of viral-specific antigens or antibodies. Therefore, the detection of active viral infection by RNA splicing may be particularly important for early diagnosis of viral infection and may be critical for successful treatment. Because of direct association of spliced viral transcripts with the level of active viral replication and by monitoring viral RNA, one might be able to provide essential information early enough for initiation of antiviral therapy. A rapid shutoff of viral transcription and RNA splicing could be also the first sign of the blockage of viral replication visible even before the change in viral load by genome copy numbers. In the case of ubiquitous and common viruses, such as members of herpesvirus or parvovirus family, which establish latent infection in the host, detection of RNA splicing of a viral early gene would assist to distinguish viral latent infection from active lytic infection. Such a diagnosis is critical for recipients of the transplant organs where reactivation of latent viruses often leads to transplant rejection.

In addition, interpretation of the detection of RNA splicing results is straightforward without concern for carry-over DNA contamination, because spliced RNA is smaller than its corresponding DNA template. As described above, there are many techniques currently available for RNA splicing assay. These techniques are not only easy to set up with a low cost compared to virus isolations and immunological methods, but can be quickly applied to detect new emerging viruses for which the cultivation of the virus is difficult or impossible and/or no immunological method is available. This is particularly true for the combination with RNA-*seq*, which can rapidly provide sequence information about a viral transcriptome and RNA splicing of the viral messages.

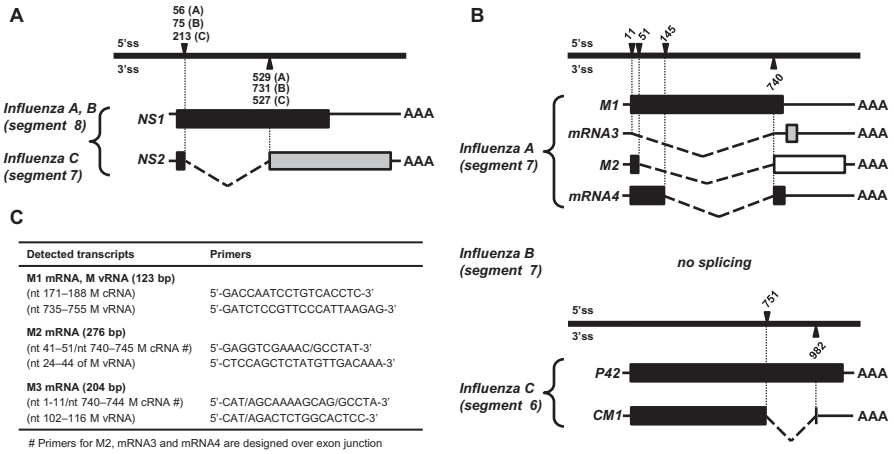
RNA Splicing in RNA Viruses

Influenza Viruses

Influenza virus infection affects millions of people every year. Influenza viruses, including influenza virus A, B, and C, are the members of the *Orthomyxoviridae* family. Influenza viruses are enveloped RNA viruses with a segmented, single-stranded RNA (ssRNA) genome of negative polarity. The number of segments may vary between virus species, with influenza viruses A and B genome having eight segments and influenza C seven segments. In contrast to the majority of RNA viruses, the influenza viruses replicate in the nucleus of host cells because of their dependence on cellular expression machinery [37]. During replication, the viral RNAs are produced by viral RNA-dependent RNA polymerase. However, viral RNA genomes use short sequences with a cap structure generated by host RNA polymerase II for priming to initiate viral transcription. During infection viral polymerase produces two types of RNAs: one for protein synthesis and the other serving as a template for viral genome replication (see review [38]).

RNA splicing in influenza viruses was first detected in an RNA transcript from the smallest segment 8 in influenza A and B as well as in their corresponding segment 7 in influenza C. This transcript encodes two nonstructural viral proteins: larger NS1 space and smaller NS2 [39]. In influenza A, NS1 protein is encoded by an unspliced primary RNA transcript (~890 nts), whereas NS2 protein is expressed from a spliced RNA (~350 nts) generated by removal of a 473-nt intron from its primary RNA transcript. This results in influenza A NS1 and NS2 proteins sharing the same AUG start codon as well as the first nine amino acid residues. Translation of NS2 protein continues in +1 frame after RNA splicing, that results in the C-terminal NS2 partially overlapping the NS1 by 70 amino acid residues [40] (Fig. 7a). A similar splicing event for production of NS1 and NS2 proteins has been detected from influenza B infections [41] as well as from influenza C [42] infections.

The influenza A segment 7, which encodes M1 and M2 proteins, produces 3 RNA species by alternative RNA splicing events. The unspliced RNA, which is collinear with the genome, encodes M1 nucleoprotein composed of 252 amino acid residues. The two alternatively spliced RNAs, M2 and mRNA3 [43], share the same 3' splice site at nt 740 position, but use different 5'splice sites for alternative RNA splicing (Fig. 7b). M2 RNA uses a 5' splice site at nt 51 position, whereas mRNA3 employs another 5' splice site at nt 11 position from the beginning of viral-specific sequences. The M2 protein has ion channel activity and shares 8 amino acid residues with the M1 N-terminus; it also overlaps with 14 amino acid of M1 C-terminus. The mRNA3 contains a short open reading frame in its exon 2 with the potential to encode a short peptide of 9 amino acid residues. However, the expression of this peptide has never been experimentally confirmed. The role of this transcript during virus replication remains unknown.



C

Detected transcripts	Primers
M1 mRNA, M vRNA (123 bp)	
(nt 171–188 M cRNA)	5'-GACCAATCCTGTCACCTC-3'
(nt 735–755 M vRNA)	5'-GATCTCCGTTCCCATTAAGAG-3'
M2 mRNA (276 bp)	
(nt 41–51/nt 740–745 M cRNA #)	5'-GAGGTCGAAAC/GCCTAT-3'
(nt 24–44 of M vRNA)	5'-CTCCAGCTCTATGTTGACAAA-3'
M3 mRNA (204 bp)	
(nt 1-11/nt 740–744 M cRNA #)	5'-CAT/AGCAAAAGCAG/GCCTA-3'
(nt 102–116 M vRNA)	5'-CAT/AGACTCTGGCACTCC-3'

Primers for M2, mRNA3 and mRNA4 are designed over exon junction

Fig. 7 Diagrams of RNA splicing of influenza virus NS and M transcripts. Thin lines represent noncoding sequences. Dashed lines indicate splicing directions. Transcripts are polyadenylated (AAA) at the 3' end. Black boxes mark ORF in primary transcripts. Small black boxes and white boxes are ORF in other frames created by RNA splicing. RNA splicing of NS1 (**a**) and M1 (**b**) transcripts in influenza A, B, and C viruses is diagramed according to Lamb and Horvath [38]. (**c**) Oligo primers used for RT-PCR to detect spliced RNA products from segment 7 of influenza A/WSN/33 (H1N1) as described [207]

While M2 and mRNA3 transcripts are detectable in cells infected with all influenza A viruses, some strains, such as A/WSN/33, produce an additional spliced transcript named mRNA4 [44]. The spliced mRNA4 transcript is generated by usage of additional 5' splice site at position nt 146 and shares the same 3'splice site with M2 and mRNA3 at position nt 740 (Fig. 7b). The mRNA4 has the potential to encode a peptide with 54 amino acid residues, and its first 37 amino acid residues are identical with M1 protein. Sequence analysis of more than 6000 influenza strains revealed that about 20 influenza A strains have this conserved mRNA4 splice site [45]. Sequence information of all influenza viruses can be found at https://www.fludb.org/.

The primary RNA transcript of segment 7 in influenza virus B does not undergo alternative splicing to produce M2 protein as happens in the case of influenza A. RNA splicing in the M transcript takes place with segment 6 of influenza C [46]. The two transcripts generated from segment 6 in the infected cells are the full-length primary and a single-spliced transcript created by removal of an intron located at the 3'end of the primary transcript. The primary transcript contains a 374 aa-long ORF (P42), but the spliced message contains a shorter ORF encoding 242-aa residues (CM1) due to generation of a stop codon after RNA splicing from a 5' splice site at nt 751 to a 3' splice site at nt 982 (Fig. 7b). The P42 protein is consequently processed by internal cleavage, resulting in production of a predominant CM2 protein containing the C-terminal 115-aa residues of P42 protein [47, 48].

In summary, there are two instances of viral RNA splicing in influenza infection: one conserved between all three species (NS1) and the second being highly variable

in each species (M). Thus, the combination of NS1 and M RNA splicing assays would reveal not only active influenza virus infection but also be able to specify the infection with a specific influenza virus species.

Human Retroviruses

HIV-1 and *HIV-2*

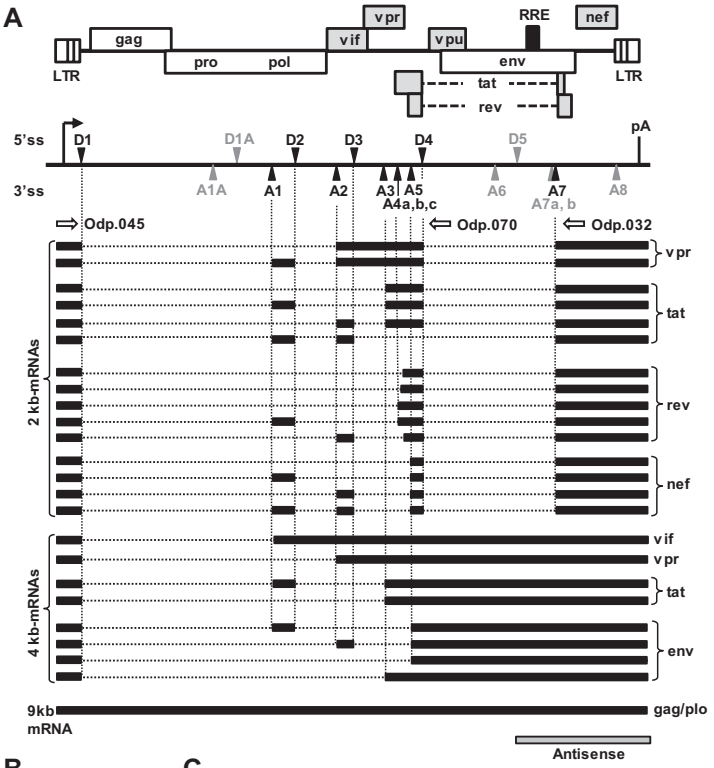
Human immunodeficiency virus (HIV) is a member of *Lentivirus* genus in retrovirus family and causes the acquired immunodeficiency syndrome (AIDS). HIV infects cells of the immune system and consequently causes the failure of immunity that is associated with occurrence of opportunistic infections that may result in death. HIV infection is considered to be a pandemic with about 0.6% of the world population being infected. Two types of HIV viruses have been characterized. Although closely related, HIV-1 differs from HIV-2 in infectivity and geographical distribution, with HIV-2 being much less pathogenic and predominantly occurring in several West African countries.

HIV is an enveloped virus and carries two copies of a single-stranded RNA genome of 9 kb having positive polarity (ssRNA+). After initial infection, the viral genomic RNA is converted by virus-encoded reverse transcriptase into DNA, which then can integrate into the host genome where this integrated viral genome subsequently resides as a provirus. Later, the integrated provirus serves as a template for continued transcription of viral transcripts.

In contrast to simple retroviruses, the HIV genome has a high coding capacity. In addition to encoding viral structural and replication proteins (*gag, pol, env*), HIV encodes a large number of accessory proteins by production of over 40 RNA isoforms, which are derived from a single RNA transcript due to extensive alternative RNA splicing [49] (Fig. 8a).

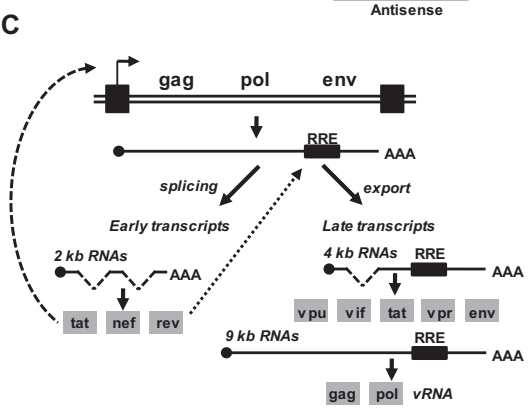
Three groups of HIV transcripts can be observed by size in Northern blot analysis. The first group represents an unspliced 9-kb transcript, which serves a template for expression of *gag* and *gag/pol* as well as serving as genomic RNA for newly formed virions. The second group represents single-spliced RNA transcripts of ~4 kb, which encode *env*, *vif*, *vpr*, and *vpu* proteins. The third group of transcripts of ~2 kb consists of multiple-spliced RNA transcripts that encode accessory proteins *tat*, *rev*, *nef*, and *vpr*. During virus infection, HIV generates a wide variety of RNA transcripts by using at least five alternative 5' splice sites as well as eight to nine alternative 3’ splice sites [50, 51] (Fig. 8b). In addition, several antisense transcripts from several 3’ long terminal repeats (3’ LTR) have been detected in HIV-1-infected cells [52].

Recent studies have demonstrated that HIV alternatively RNA splicing is largely regulated by viral RNA *cis*-elements as well as cellular splicing factors and is orchestrated for completion of the HIV life cycle during virus infection. Multiple-spliced transcripts of the 2-kb family are expressed in the early stage of virus infec-



B

pNL4-3	
D1	743
D2	4962
D3	5463
D4	6044
D5	6724
A1	4913
A2	5390
A3	5777
A4a	5936
A4b	5954
A4c	5960
A5	5976
A6	6611
A7a	8335
A7b	8341
A7	8369
A8	9167



D

Primers	Position (nt)	Sequence	Detected transcripts
Odp.045	477-499	5'-CTGAGCCTGGGAGCTCTCTGGC-3'	2.0 kb (early)
Odp.032	8477-8498	5'-CCGCAGATCGTCCCAGATAAG-3'	
Odp.045	477-499	5'-CTGAGCCTGGGAGCTCTCTGGC-3'	4.0 kb (late)
Odp.070	6094-6115	5'-ACTATTGCTATTATTATTGCTACTAC-3'	

Fig. 8 RNA splicing of HIV-1 transcripts. (**a**) Schematic diagram of HIV-1 genome with structural and replication protein ORFs marked with empty boxes and accessory protein ORFs with gray boxes. *LTR*, *long terminal repeats.* Shown below the diagram are positions of four 5' splice sites (D1-4) and seven 3' splice sites (A1-7) identified in a 9-kb full-length primary transcript which encodes *gag* and *pol* proteins. Alternatively spliced HIV-1 transcripts in the size of

tion and express *tat*, *rev*, and *nef*. This group of spliced RNAs is produced by using the 3' splice site A3–A5 located in the central part of the viral genome with an A3 site for expression of *tat*, A4a-c for *rev*, and A5 splice site for *nef* proteins. During late stage of HIV infection, nuclear import and accumulation of *tat* together with *rev* protein allow the *rev* protein to bind to a *rev-responsive element* (RRE) in partially spliced 4-kb and unspliced 9-kb RNA transcripts located in the *tat/rev* intron between the D4 and A7 splice sites; this *rev* protein mediates the export of later transcripts into the cytoplasm for translation [53] (Fig. 8c). Sites A1A and D1A are involved in pre-mRNA stability [54]. Strains from the IIIB family of HIV viruses use additional A6 and D5 splice sites to generate a small exon in the *env* region; transcripts containing this exon express the tripartite *tat-env-rev* fusion protein, *tev* [55, 56].

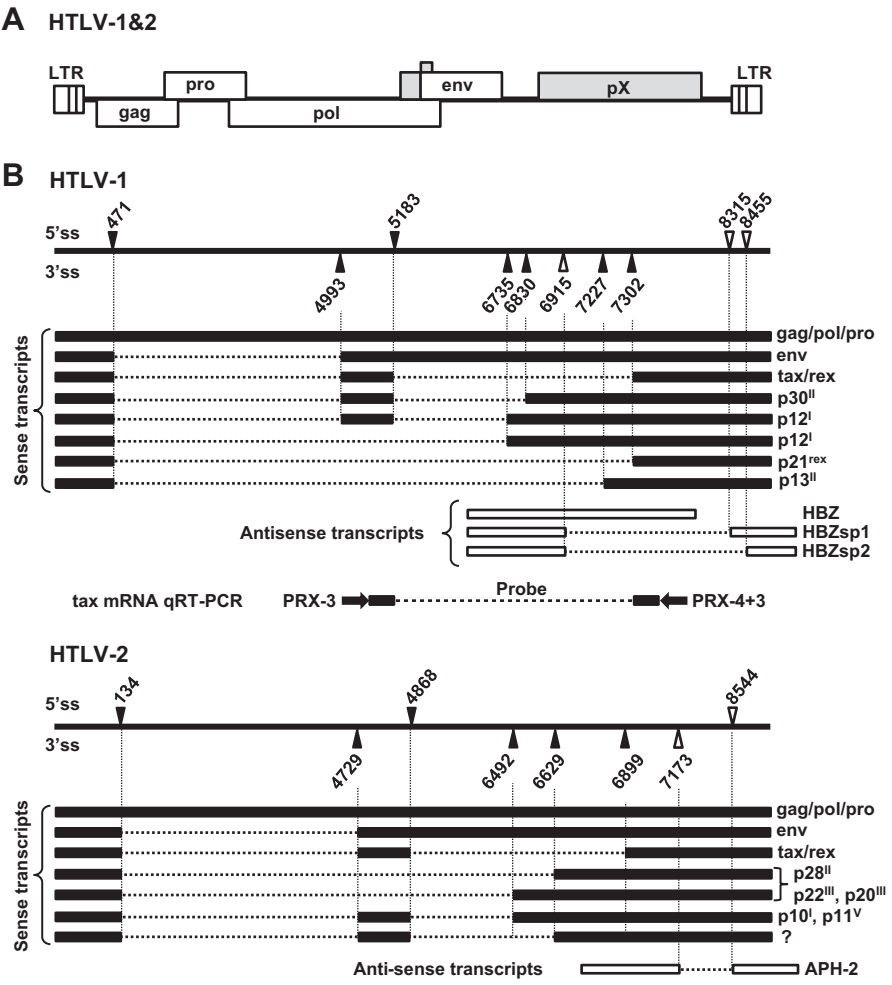
Regulation of HIV RNA splicing depends on the selection of 3' splice sites, which are, in general, weak in contrast to stronger and highly active 5' splice sites. In addition, numerous positive and negative splicing regulatory *cis*-elements identified in the HIV RNA genome bind various cellular splicing factors and thus affect the selection of individual 3' splice site (see review [49]). Comparison of nucleotide sequences between the various clades of HIV-1 has demonstrated a high level of conservation of splice sites among the different clades of HIV-1 strains (except D4a, b, c).

HTVL-1 and *HTVL-2*

Human T-cell leukemia virus type 1 (HTLV-1) and type 2 (HTLV-2) were the first two retroviruses discovered in humans [57]. HTLV-1 is etiologically linked to adult T-cell leukemia/lymphoma (ATLL), an aggressive malignancy of CD4+ T lymphocytes, as well as to a neurological disorder named *HTLV-1-associated myelopathy/tropical spastic paraparesis* (HAM/TSP) [58–60]. HTLV-1 is endemic in Japan, Africa, Caribbean basin, and South America. HTLV-2 is linked to HAM/TSP but not to ATLL. HTLV-2 infection occurs predominantly in parts of Africa and Americas [61].

HTLV-1 and HTLV-2 are two closely related complex retroviruses that share about 70% of their nucleotide sequences. Their genome organization and replication are similar to HIV. Besides essential genes (*gag/pol/env*) expressed from the

Fig. 8 (continued) ~2 kb and ~4 kb are grouped along with their protein-coding potentials. Solid boxes represent exons and dotted lines are introns. A gray box below the 9-kb mRNA illustrates a recently discovered antisense transcript [52]. (**b**) Nucleotide positions of all mapped 5' and 3' splice sites in a prototype HIV genome, pNL 4-3 (GenBank Acc. No. AF324493), starting from the 5' LTR [208]. (**c**) Alternative HIV RNA splicing is coupled with stages of HIV-1 infection. Multiple-spliced transcripts (2-kb group) are expressed in the early stage of the infection resulting in the expression of accessory proteins: *tat*, *rev*, and *nef*. During the late stage of HIV-1 infection, *rev* protein translocation to the nucleus promotes nuclear export of single-spliced (4-kb group) or unspliced HIV RNA via binding to an *rev*-responsive element (RRE) to express structural and replication proteins [51]. (**d**) Oligo primers used for RT-PCR, shown in (A), to detect spliced RNA products of HIV-1 pNL 4-3 [208]



C

Primer	Position (nt)	Sequence
RPX-3	5096-5115	5'-ATCCCGTGGAGACTCCTCAA-3'
RPX-4 + 3	7360-7338	5'-CCAAACACGTAGACTGGGTATCC-3'
Probe	5172-5183/7302-7312	5'-TCCAACACCATG/GCCCACTTCCC-3'

Fig. 9 RNA splicing of HTLV-1 and HTLV-2 transcripts. (**a**) Genomic organization of HTLV-1 and HTLV-2 with structural and replication genes (white boxes) on the 5' end and regulatory gene *pX* (gray boxes) on the 3' end of the genome flanked with long terminal repeats (LTR). (**b**) Viral transcripts generated by alternative RNA splicing in HTLV-1- and HTLV-2-infected cells. Black boxes indicate exons and dotted line are introns. Exons in antisense transcripts are represented by empty boxes. The 5' and 3' splice sites are indicated by their nucleotide positions in the virus genomes. The coding potentials of each transcript are shown on the right. (**c**) Positions and sequences of the primers and probes used for detection of HTLV-1 spliced tax mRNA as described [209]. Nucleotide positions are based on a full-length HTLV-1 cDNA (GenBank Acc. No. L03562.2)

full-length or single-spliced RNA, HTLVs also encode a number of accessory proteins in a pX region located in the 3'end of the virus genome (Fig. 9a). Transcripts for encoding accessory proteins are generated by alternative RNA splicing of a full-length primary transcript using several 5' splice sites and 3' splice sites (Fig. 9b) (reviewed in [62]). In addition, transcripts antisense to pX region have been discovered in infected cells either with HTLV-1 or HTLV-2, and these antisense transcripts encode HBZ and APH-2 proteins, respectively [63, 64]. Interestingly, these transcripts also undergo RNA splicing.

In contrast to HIV-1, the regulation of HTLV RNA splicing and the roles of cellular splicing factors in HTLV RNA splicing are poorly understood.

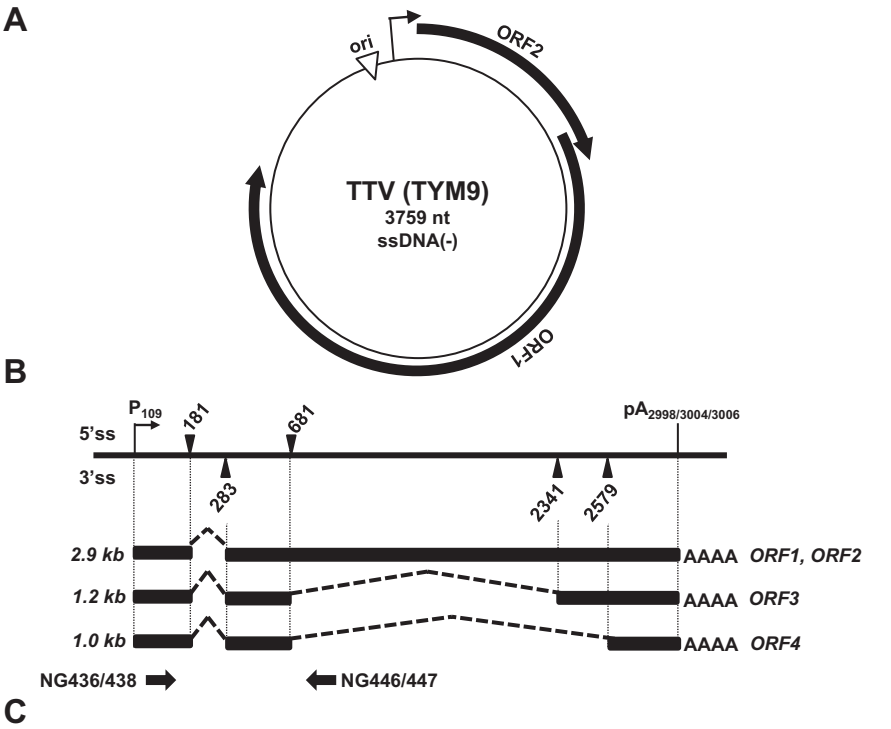
DNA Viruses

Circoviruses

Human circovirus Torque teno virus (TTV) was originally discovered in the serum of a patient with posttransfusion non-A to G hepatitis [65]. Later studies showed that TTV is present in body fluids of healthy individuals and is not associated with any pathological disorder. However, the prevalence of TTV in the general population appears highly variable and ranges from about 2% to 90% of the incidence of TTV infection. This large variation among reported studies is most likely attributable to primer selection and PCR performance [66]. TTV isolates have considerable diversity (about 30%) that can be clustered in several genotypic groups without any particular geographical distribution, indicating that TTV likely represents a ubiquitous virus. Virus replication, route of TTV infection, and association with pathological manifestations remain unclear.

TTV is small non-enveloped virus of icosahedral architecture and contains a circular single-stranded DNA genome of 3.6–3.9 kb of negative polarity (ssDNA-) [67, 68]. TTV replication is not yet fully understood, partially due to the lack of an appropriate tissue culture system. The TTV genome consists of a GC-rich noncoding region and a protein-coding region with two overlapping ORFs (Fig. 10a). Three species of RNAs of 2.9, 1.2, and 1.0 kb have been detected in infected bone marrow cells as well as in an in vitro-infected cell line [69, 70]. All transcripts originate from the same promoter and undergo alternative RNA splicing (Fig. 10b). Each RNA transcript can be translated into two different proteins by using two alternative start codons [71]. The presence of other ORFs (ORF3 and 4) in the TTV genome has been predicted, but has not yet been confirmed.

The role of TTV RNA splicing and its regulation in infection by this virus as well as TTV viral replication remained largely unknown. Another human circovirus TTV-like mini virus (TLMV) also has been identified in human sera [72]. TLMV shares the same genetic organization with TTV and other circoviruses, but its genome is only about 2.9 kb in size.



Primer	Position (nt)	Sequence	Amplicons (bp)
NG436	111–133	5'-GTTTTCTACGCCCGTCCGCAGCG-3'	
NG438	132–155	5'-CGGCAGCACCACGGAGGGTGATCC-3'	DNA: (575)
NG446	718–740	5'-CCGCGTCTA AGGTATGTGCGTCG-3'	RNA: 2.9 kb (474)
NG447	686–706	5'-TTCCCCACCGCCTCCGTCTCC-3'	

Fig. 10 Human circovirus TTV. (**a**) Single-stranded genome of Torque teno virus (TTV) with the origin of replication (ori), promotor (arrow), and two identified overlapping ORFs. (**b**) Transcripts identified in cells infected with TTV virus with nucleotide positions of two 5' splice sites (above the line) and three 3' splice sites (below the line), together with positions of promoter (arrow) and polyadenylation site (pA). The black boxes represent exons and dashes lines are introns. The size of the spliced transcripts (in kb) is indicated on the left, and the coding potentials are on the right. Arrows below the transcripts are a primer pair used to detect the 2.9-kb transcript [69] by RT-PCR as detailed in (**c**). Primer positions in (C) are based on Torque teno virus TYM9 strain (GenBank Acc. No. AB050448.1)

Hepadnaviruses

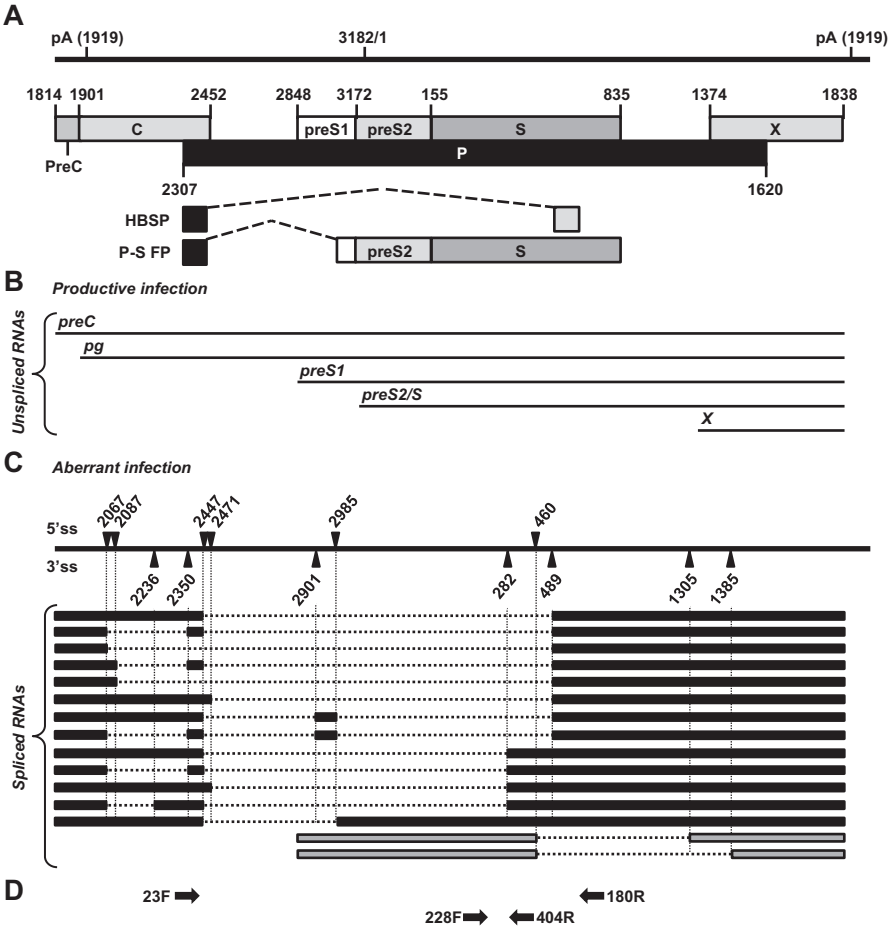
Hepatitis B virus (HBV) is a hepatotropic virus, by which chronic hepatic infection can result in the development of liver cirrhosis and hepatocellular carcinoma [73]. Despite an effective HBV vaccine that is available in many countries, HBV infection remains epidemic in many parts of the world, particularly in Asia and

sub-Saharan Africa. WHO estimates that HBV infects about 2 billion of people worldwide, with chronic infection affecting 350 million and causing death in approximately 1 million persons each year. While the vaccine can prevent HBV infection, there is no cure for already infected individuals due to the persistence of the active transcriptional template of HBV covalently closed circular DNA (cccDNA).

HBV is a non-cytopathic, hepatotropic virus. Its genome consists of a 3.2-kb-long circular and partially double-stranded DNA. HBV encodes four viral proteins: core (C), reverse transcriptase-polymerase (P), surface (S), and X proteins (Fig. 11a). In infected hepatocytes, the DNA is converted to cccDNA (covalently closed circular DNA) and serves as a template for expression of viral pre-genomic (pgRNA) and three subgenomic RNAs from several promoters. Despite the presence of several promoters, all transcripts used the same poly(A) signal for transcription termination (Fig. 11b). The core protein C and polymerase-transcriptase protein P are encoded by a bicistronic pgRNA. Subgenomic preS and S RNAs translate three surface antigens: (1) large [preS1], (2) middle [preS2], and (3) small [S] surface antigens. A short RNA of 0.7 kb encodes X protein, a nonstructural viral protein that presumably has oncogenic potential [74]. A pre-core RNA initiated upstream of pgRNA encodes HBeAg [75, 76]. Beside protein translation, pgRNA is a template for reverse transcription and results in a genomic minus DNA strand during HBV genome replication.

All HBV transcripts from cccDNA are produced by cellular RNA polymerase II. A spliced 2.2-kb RNA transcript was first identified in transfected hepatoma cells [77] and contains a single 1223-nt-long intron starting from the end of the core antigen ORF to the middle of the S antigen ORF. Subsequently, other single- and multiple-spliced forms of pgRNA have been discovered, with sizes of 2.1–2.6 kb, in both cell cultures and liver tissues of HBV patients [78–80]. So far, 13 spliced variants of pgRNA and 2 spliced isoforms of pre-S2/S RNA have been identified from HBV gene expression during infection; these spliced viral RNAs have been produced by the use of six 5' splice sites and seven 3' splice sites (Fig. 11c). A viral *cis*-element PRE (posttranscriptional regulatory element) as well as cellular splicing factors such as PTB (polypyrimidine track-binding protein) and SR proteins also may play roles in the regulation of HBV RNA splicing (see review [81]).

Approximately 30–50% of HBV RNA during HBV infection of human hepatoma cell lines Huh7 and HepG2 are spliced RNAs; Huh7 and HepG2 are two popular cell lines for in vitro HBV replication studies [82]. Huh7 cells seem to produce more spliced RNA than do Hep2G cells. The major spliced product is derived from 30% of pgRNA using nt 2447 5'ss and nt 489 3'ss in genotypes A, C, D, and E. Serum of infected patients or hepatocarcinoma tumor samples frequently contain HBV DNA originating from spliced variants [83, 84]. The level of spliced HBV RNAs in patients varies widely from no splicing to extensive splicing and is related to viral genotype [85]. The role of HBV RNA splicing in the HBV life cycle or in HBV pathogenesis remains to be elucidated. HBV spliced RNAs express two new proteins [86, 87]. A spliced mRNA derived from pgRNA with removal of a



Primers	Position (nt)	Sequence	Detected transcripts/genomes
23F	2411-2428	5′-CCGCGTCGCAGAAGATCT-3′	**spliced/defective**
180R	609-587	5′-ATGGGAATACAGGTGCAATTTCC-3′	
228F	247-271	5′-TCTAGACTCGTGGTGGACTTCTCTC-3′	**unspliced/full length**
404R	423-401	5′-CATAGCAGCAGGATGAAGAGGAA-3′	

Fig. 11 Expression and RNA splicing of HBV transcripts. (**a**) Diagram of linear HBV genome structure. Numbers indicate the beginning and the end of each ORF. (**b**) Full-length viral transcripts generated from viral genome during productive HBV infection. (**c**) Alternatively, spliced viral transcripts of *preC* (black) or *preS1* (gray) primary transcripts, with exons in black or gray boxes and introns in dotted lines. Numbers above or below the linear genome are nucleotide positions of mapped 5' and 3' splice sites. It has been assumed that alternative RNA splicing of HBV *preC* and *preS1* leads to production of defective HBV viral particles. Diagrams are modified from Sommer and Heise [81]. Arrows below the transcripts are two sets of primer pairs used to detect spliced *pgRNA* products of HBV by RT-PCR [210] as detailed in (**d**) with primer positions and sequences derived from HBV TK113 genome (GenBank Acc. No. JF754635)

454-nt intron from nt 2447 to 2901 encodes a structural polymerase-surface fusion protein p43 with potential function in the entry [86]. Another single-spliced pgRNA with removal of an intron from nt 2447 to 489 translates a 93-aa fusion protein in size of 10.4 kDa, in which the first 46-aa residues are identical to the N-terminus of the viral polymerase protein followed by the 47-aa residues generated by the frameshift from the second exon. This protein has been referred as hepatitis B splice-generated protein or HBSP [87] and seems to be associated with chronical HBV infection, HBV viral cytopathogenic effect, and HBV immune evasion (see review [88]).

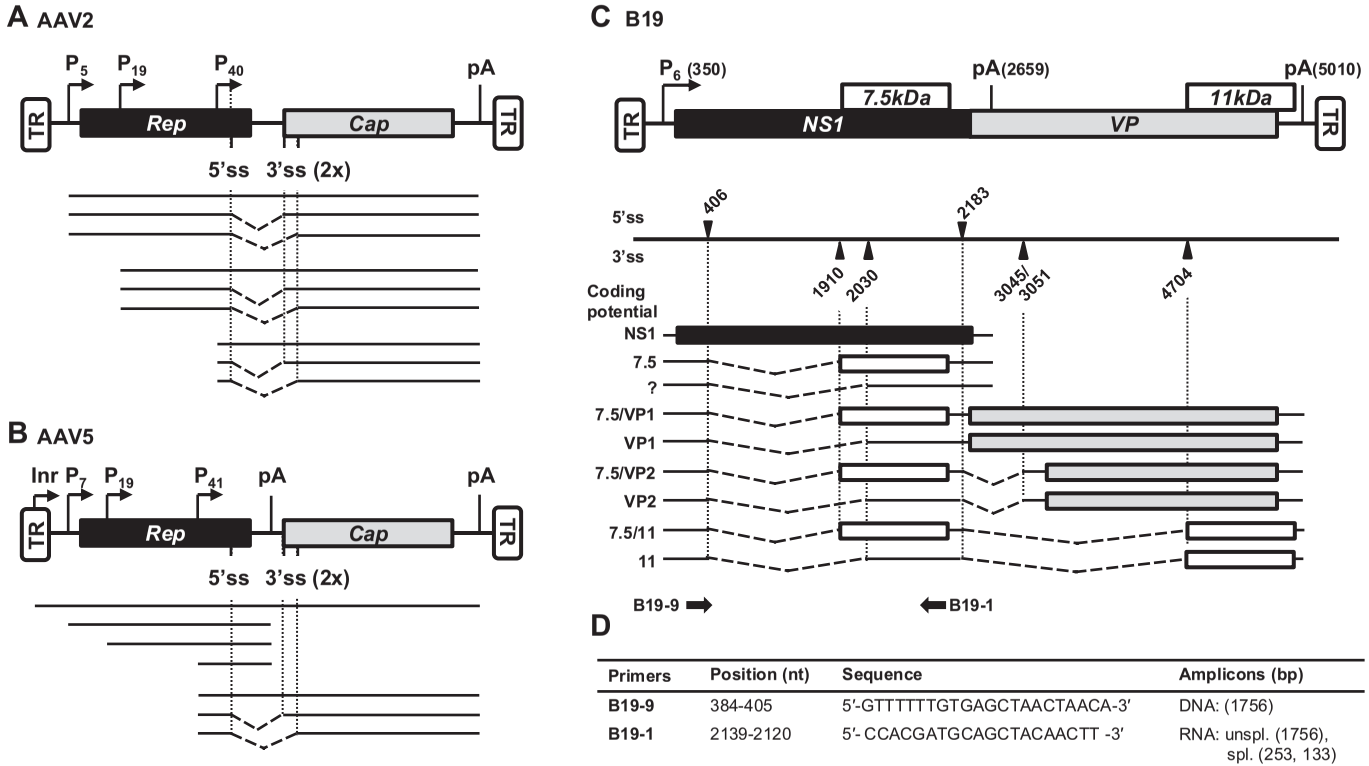
Parvoviruses

Parvoviruses are a group of small non-enveloped viruses containing a single-stranded DNA genome (ssDNA) of ~6 kb. The palindromic inverted terminal repeats at the ends of the virus genome function as an origin for replication. Parvoviruses replicate via a double-stranded DNA intermediate that serves as a template for viral transcription [89]. Replication of some parvoviruses relies on "helper" viruses such as adenovirus, herpesviruses, vaccinia virus, and human papillomaviruses [90–92].

Parvoviruses are ubiquitous viruses and infect a wide range of animals. As of today, there are at least four members of the *Parvoviridae* family that are infectious to humans: adeno-associated viruses (AAV), parvovirus B19 (B19V), human bocaviruses (HuBoV), and human Parv4 [93]. Despite structural and genetic similarity, different parvoviruses use different replication and transcription strategies during viral infection and have different host tropisms for initiating a productive infection in the presence of a helper virus.

Adeno-associated viruses (AAV), currently classified as *Dependoviruses*, were the first human parvoviruses identified in the group. AAVs infect a wide range of species with AAVs-1, AAVs-2, AAVs-3, AAVs-8, and AAVs-9 being found in human [94]. Currently no disease or pathological condition is associated with AAV infection in humans. The correlation between AAV infection and fetal loss and male infertility has been proposed due to a high prevalence of AAV DNA in placental tissues and in genital tissues of men with abnormal semen [95, 96]. This has not yet been shown to be causal. Because AAV lacks pathogenicity, induces a low immune response, and infects both dividing and nondividing cells with the capability of viral DNA integration into the host genome, AAV has gained attention as a vector for gene therapy (see review [97]).

All AAV genomes consist of two open reading frames, *Rep* and *Cap*, with *Rep* for virus replication and *Cap* for structural capsid protein. AAVs use several different strategies to produce viral products. The first group represents AAV1, AAV2, AAV3, AAV4, and AAV6, and their viral transcripts originate from one of three viral promoters on the left-hand side of the viral genome and are terminated on a single polyadenylation site on the right-hand side of the genome. The middle part of the transcripts contains a ~300-nt-long intron with a non-consensus 5' splice site and



D

Primers	Position (nt)	Sequence	Amplicons (bp)
B19-9	384-405	5′-GTTTTTTGTGAGCTAACTAACA-3′	DNA: (1756)
B19-1	2139-2120	5′- CCACGATGCAGCTACAACTT -3′	RNA: unspl. (1756), spl. (253, 133)

Fig. 12 RNA splicing of human parvoviruses. Genome organizations of AAV2 (**a**), AAV5 (**b**) and human *Erythrovirus* B19 (**c**) with terminal repeats (TR) on the ends of each genome, along with viral promoters (P), polyadenylation sites (pA), splice sites (ss), and open reading frames (boxes). As shown below each genome are viral transcripts generated by alternative transcription initiation, RNA splicing, and polyadenylation. In each panel, black boxes represent coding regions, solid lines for noncoding regions, and dashed lines for splice directions to remove the corresponding introns. (**d**) Oligo primers used for RT-PCR to detect spliced RNA products of B19 [211]. The primer positions and sequences are based on a partial genome sequence of B19-Au strain (GenBank Acc. No. M13178.1)

two 3' splice sites (Fig. 12a).The efficient splicing requires the presence of both a helper virus and a large *Rep* protein [98]. While *Rep* protein seems to be essential for AAV2 splicing, several adenovirus proteins (E1A, E1B, E2a, E4orf6, and VA RNA) as well as some products of herpes simplex virus (UL5, UL8, UL52, and UL29) also have a stimulatory effect on AAV2 splicing [99]. AAV5 and some animal AAVs are in the second group that utilizes three upstream promoters for their transcription, but their genome contains additional polyadenylation sites in the intron region. Transcripts from two upstream promoters are polyadenylated on the internal poly(A) site, whereas spliced transcripts from P41 promoter use a poly(A) site at the right side of the genome (Fig. 12b). The only spliced transcript in AAV5 infections is the *Cap* transcript which contains a smaller (~240 nt) intron. Interestingly, the splicing of AAV5 *Cap* transcript is constitutive and highly efficient even in the absence of helper virus infection [100].

Human B19 virus, a member of *Erythrovirus* genus, was first identified in the serum of blood donor [101]. Three genotypes of B19 viruses have been identified from different geographic regions [102]. After acute infection, the virus persists in host for the rest of the life. The infection by B19 virus is generally asymptomatic, but several pathological conditions have been associated with B19 infection; these include *erythema infectiosum* (the "fifth disease) [103], *polyarthropathy syndrome* [104], transient aplastic crisis (TAC) [105], and persistent anemia/pure red cell aplasia (PRCA). B19 infection during pregnancy may associate with spontaneous miscarriage and development of nonimmune *hydrops fetalis* [106].

Like other parvoviruses, B19 virus genome encodes two large open reading frames. NS1 ORF on the left-side genome translates a 77-kDa nonstructural protein, while a VP ORF on the right-side genome produces two capsid proteins (84-kDa VP1 and 58-kDa VP2). At least nine virus-specific transcripts have been detected following B19 infection [107]; all of which are transcribed solely from a single promoter P6 located upstream of NS1 gene, but are alternatively spliced and terminated at two alternative polyadenylation sites (Fig. 12c) located either in the middle or on the far right side of the genome. By using the poly(A) site in the middle of virus genome, the P6 transcript has an intron in the NS1 ORF, and splicing of this intron from NS1 transcript may create a novel ORF encoding a small accessory 7.5-kDa protein. However, if the poly(A) site on the right-side genome is used for RNA polyadenylation, the P6 transcript becomes a bicistronic (NS1 and VP) transcript with two introns. By splicing to remove the intron 1 from the bicistronic RNA, the single-spliced P6 transcript is capable of encoding both 7.5-kDa and VP1 proteins. Double RNA splicing to remove both intron 1 and intron 2 from the P6 bicistronic transcript disrupts both ORFs for NS1 and VP1, but creates either a VP2 or a novel ORF for another accessory 11-kDa protein, depending on which alternative 3' splice site is selected (Fig. 12c). Thus, all detected B19 transcripts are derived from a P6 pre-mRNA containing one or two introns with two possible alternative 3' splice sites depending on the selection of one of two possible alternative poly(A) sites. All detected B19 transcripts are alternatively spliced RNA transcripts, except the unspliced full-length NS1RNA. The *cis*-elements in the central exon and intron

2 are regulatory elements and control the alternative P6 RNA splicing, with the double-spliced P6 RNAs being the predominant species in the infected cells [107].

Adenoviruses

The two most common infection sites by adenoviruses in humans occur in the upper and lower respiratory tract and result in bronchitis and/or pneumonia. Adenovirus infection can also involve a wide range of other sites resulting in *conjunctivitis*, ear infection, *gastroenteritis*, *myocarditis*, *hemorrhagic cystitis*, *meningitis*, and *encephalitis*. There are 56 adenovirus types belonging to seven species (human adenovirus A–G). Types belonging to B and C are responsible for most respiratory infections, B and D for *conjunctivitis*, and F and G for *gastroenteritis* [108]. Adenoviruses were also found in other vertebrates.

Even though the human adenoviruses are not etiologically linked to any human cancer, some adenoviruses (types 2, 5, 12, 18, and 31) can, under special circumstances, transform rodent cells in vitro and induce tumors in small animals. Transformation activities are linked to two oncogenes: (1) E1A, which bind tumor suppressor pRB, and (2) E1B, which binds tumor suppressor p53 [109].

Adenoviruses are non-enveloped viruses with an icosahedral architecture that contains a linear, non-segmented, double-stranded DNA genome of ~26–45 kb that is capable of encoding 22–40 different gene products [109]. Adenoviruses replicate in the nucleus of infected cells (Fig. 13a). The early stage of virus infection is characterized by the expression of a nonstructural early protein, while viral structural proteins are expressed in the late stage of viral DNA replication marking a switch between the two infection phases.

Almost all adenoviral early and late transcripts undergo RNA splicing in order to produce their corresponding viral products [110]. Here, viral E1A and L1 transcripts are used as examples for the alternative RNA splicing seen in adenoviral infections.

The adenovirus E1A primary transcript contains three 5’ splice donor sites as well as two 3’ acceptor sites and is composed of three exons and two introns. The first intron from nt 636 to nt 853 is a suboptimal, minor intron. The second intron is a major intron that uses two alternative donor sites, respectively at nt 973 and nt 1111 as well as one acceptor site at nt 1226 for RNA splicing. Alternative splicing of E1A RNA through the utilization of various combinations of splice donor and acceptor sites leads to the formation of five different species (13S, 12S, 11S, 10S,

Fig. 13 (continued) (**b and c**) Alternatively spliced RNA transcripts of adenovirus early *E1A* gene (**b**) and adenovirus late *L1* gene (**c**). Black boxes, exons; white boxes, alternative exons; dashed lines, introns or splice directions. Nucleotide positions of each splice site are based on a complete genome sequence of human adenovirus type 2 (GenBank Acc. No. AC_000007.1). (**d**) Schematic exon compositions of *52*, *55K*, and *IIIa* transcripts and exon junction probes for specific detection of spliced *L1* isoforms from adenovirus type 2 by in situ hybridization as described [212]

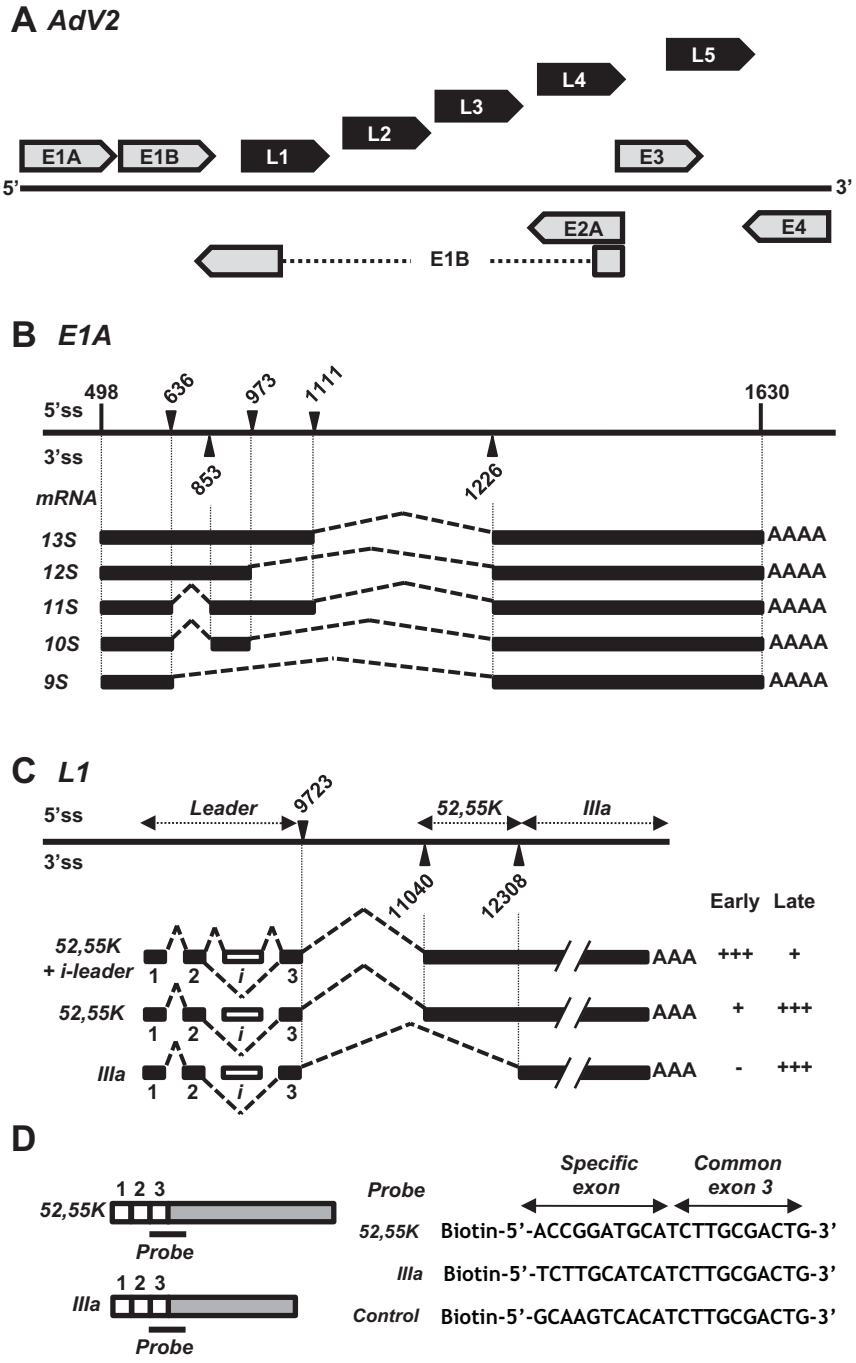


Fig. 13 Alternative RNA splicing of adenovirus early and late transcripts. (**a**) Simplified adenovirus genome with positions and orientations of viral early (gray arrows) and late (black arrows) genes.

and 9S) of E1A mRNAs based on their sedimentation coefficient (Fig. 13b) and expression of individual unique proteins [111].

The transcription of late genes starts predominantly from a major late promoter. The primary late transcript is then polyadenylated at one of five polyadenylation sites, forming five groups of late transcripts (L1–L5). Each late mRNA contains a 201-nt "leader" sequence derived from three noncoding exons that function as a translational enhancer [112]. There are two variants of leader sequence with or without an *i*-leader exon. Beside the leader sequence region, L1 transcripts also are alternatively spliced by the utilization of a common 5' splice site in combination with two alternative 3' splice sites. Selection of a proximal 3' splice site results in the formation of 52, 55K RNA, while selection of a distal 3' splice site produces IIIa mRNA (Fig. 13c).

The characteristic features of adenovirus splicing depend on the stage of virus infection. For example, E1A 13S and 12S mRNA are two major spliced products that occur during early virus infection. In contrast, 9S RNA is largely accumulated in the late stage of infection [113]. Similar phenomenon has been observed in the expression of late mRNAs. Inclusion of the *i*-leader exon is generally a signature of early transcripts, but most of the late transcripts contain a classical tripartite leader. While 52,55K L1 RNA is produced during both early and late infection, the IIIa splice site is used only in the late stage of viral infection [114]. Both cellular splicing machinery and viral products have been found to regulate alternative splicing of adenoviral transcripts during the course of viral infection (see reviews [110, 115]).

Polyomaviruses

Polyomaviruses are small non-enveloped viruses that contain a circular double-stranded DNA (dsDNA) genome of ~5000-bps. Polyomaviruses infect a wide range of mammalian and avian species, but each virus exhibits a limited host range with narrow tissue tropism. The *Polyomaviridae* family contains only one genus *Polyomavirus (PyV)*, which has nine members of the human polyomaviruses: (1) BKPyV [116], (2) JCPyV [117], (3) KI PyV [118], (4) WU PyV [119], (5) Merkel cell PyV (MCPyV) [120], (6) HPyV6, (7) HPyV7 [121], (8) *trichodysplasia spinulosa*-associated PyV (TSV) [122], and (9) HPyV9 [123]. Simian vacuolating virus 40 (SV40), a prototype virus of the family, was introduced into the human population as a contaminant in early trials of poliovirus vaccine [124]. Serological data indicate that polyomavirus infection is widespread in the general human population with initial infection occurring in childhood [125]. After the initial infection, polyomaviruses persist in the host for the rest of the life. While initial infection is usually asymptomatic, several human polyomaviruses are associated with various pathological conditions in immunocompromised patients including *nephropathy* and *cystitis* associated with BK PyV and progressive *multifocal leukoencephalopathy* associated with JCPyV [126], as well as *trichodysplasia spinulosa* presumably associated with TSV infection. Polyomaviruses expresses an oncoprotein T antigen,

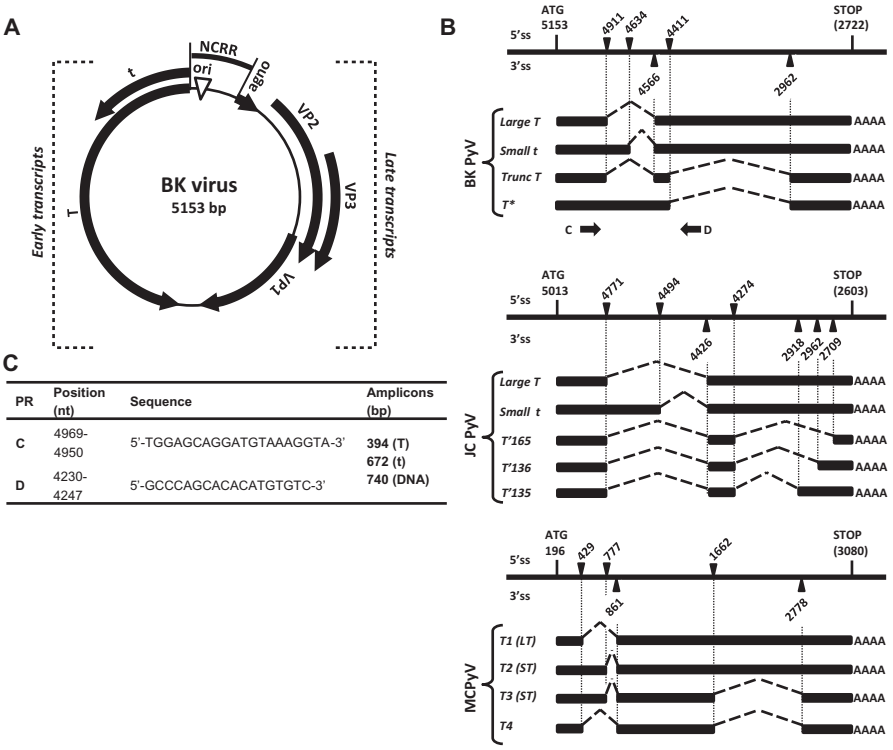


Fig. 14 RNA splicing of polyomavirus T antigen transcripts. (**a**) Genome structure of BKPyV virus, a representative of human polyomaviruses. Black arrows represent open reading frames for early viral regulatory proteins (large T (LT) and small t (ST) antigens) and late viral capsid proteins (*VP1-3*) as well as agnoprotein (agno). *NCRR*, noncoding regulatory region; *ori*, origin of replication. (**b**) Alternative RNA splicing of large *T* and small *t* antigens among BKPyV, JCPyV, and MCPyV viruses. Black boxes, exons; dashed lines, introns or splice directions; numbers, nucleotide positions of splice donor and acceptor sites. The diagrams are modified from White et al. [213] and Shuda et al. [127]. Nucleotide positions for BKPyV are strain *Dunlop* (GenBank Acc. No. V01108), for JCPyV are strain *Mad-1* (GenBank Acc. No. J02226), and for MCPyV are isolate *MKL-1* (GenBank Acc. No. FJ173815). Arrows below shows the BKPyV transcripts are oligo primers (PR) used to detect spliced large *T* and small *t* antigen transcripts of strain *Dunlop* by RT-PCR as detailed in (**c**) [131]

and this T antigen may lead to the development of human cancer by an abortive infection as recently confirmed in a rare but aggressive Merkel cell carcinoma [127–129].

The polyomavirus genome consists of three functional regions: two protein-coding regions (early and late) divided by a noncoding regulatory region (NCRR) (Fig. 14a). Early and late transcripts are expressed in two opposite directions from promoters located in the NCRR, which also contains the origin of replication. Early transcripts encode nonstructural viral regulatory proteins (T [tumor] antigens) that are important for virus replication and modulation of cell cycle. Viral DNA replica-

tion initiates the transcription of viral late genes that encode several viral capsid proteins.

In polyomavirus-infected cells, multiple isoforms of the *T* antigen are detectable because of alternative RNA splicing. The primary transcript of the *T* antigen contains two introns, but the first intron 1 has two alternative 5' splice sites. During RNA splicing, the intron 2 retention is important for production of both large *T* and small *t* antigens. However, selection of proximal 5' splice site in the intron 1 for RNA splicing leads to production of large *T* antigen, whereas selection of a distal 5' splice site in the intron 1 results in small *t* production. Because the sequence region between the proximal 5' splice site and the distal 5' splice site has a stop codon, retention of this region in small *t* RNA splicing makes the small *t* RNA larger than the large *T* RNA, but introduction of a premature stop codon in the small *t* RNA results in production of a smaller protein (Fig. 14b). In addition, a rare tiny *t* antigen of ~17 kDa has been attributed to double RNA splicing in SV40-infected cells [130]. In this case, the transcript encoding the 17-kDa antigen shows splicing of both introns, but splicing of the intron 1 by selection of the proximal 5' ss. Similar to SV40, the multiple-spliced RNA species of early transcripts also were detected in other polyomaviruses such as the truncated *T* antigen (trunc-*T* Ag) in BKPyV [131], *T'135*, *T'136*, and *T'165* in JCPyV [132] and *T3* and *T4* early transcripts in MCPyV [127] (Fig. 14b). Alternative splicing of polymavirus early transcripts allows the expression of multiple T antigens with distinguished function during the viral life cycle. In addition to the cells with actively replicating virus, the early viral transcripts also are expressed in cells with nonproductive infection or in polyomavirus-transformed cells. These cells often do not express late gene product due to integration of the viral DNA into host genome, resulting in dysregulated viral gene expression as well as cell transformation.

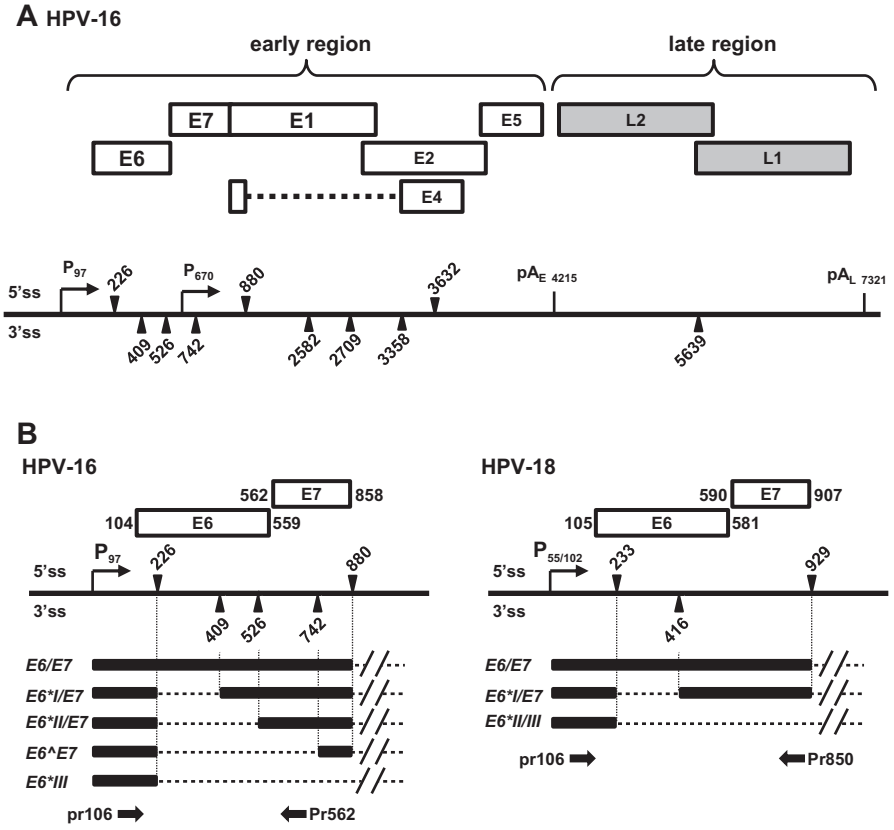
Papillomaviruses

Human papillomaviruses (HPVs) are a group of small DNA tumor viruses with a genome of ~8 kb surrounded by a viral capsid. The HPV genome consists of three regions: viral early, late, and noncoding regions and generally encodes eight viral genes (E1, E2, E4, E5, E6, E7, L1, and L2). Viral early gene products are regulatory proteins responsible for virus replication and pathogenesis during a productive infection, whereas the late L1 and L2 genes encode two viral capsid proteins for virus particle formation. Interestingly, almost all viral early genes are expressed from an early promoter upstream of the viral E6 gene and are polyadenylated at an early poly(A) signal downstream of E5 gene. Thus, viral early gene transcripts are polycistronic, with several ORFs in a single RNA molecule, and undergo extensive alternative RNA splicing during viral RNA maturation. In contrast, viral L1 and L2 are commonly transcribed from E7 ORF and polyadenylated at a late poly(A) site downstream of L1 ORF (Fig. 15a). As a result, the 5' sequences of viral L1 and L2

are part of the viral early transcript sequences. RNA splicing to remove most of these early gene sequences from the RNA is important for viral L1 and L2 expression [133].

HPVs are the etiological agent of cervical cancer and presumably of other anogenital cancers. HPV is present in >95% of all cervical cancer and is required for initiation of cervical carcinogenesis and maintenance of the cervical cancer cells. Cervical cancer is a leading cause of death for women in the developing world, with about 493,000 new cases and nearly 273,000 deaths each year. More than 200 genotypes of HPVs have been identified to date and are grouped into two major groups based on their pathogenesis and association with cervical cancer [134]. The reference genome sequences are available at https://pave.niaid.nih.gov/#home. The high-risk or oncogenic HPV types are present in cervical cancers, some anogenital cancers, as well as head-and-neck cancers. The low-risk or non-oncogenic HPVs are not associated with cancers [135]. In general, women acquire HPV infection by sexual contact. A number of epidemiology studies have demonstrated that women with repeat exposure to oncogenic HPVs as well as women with persistent cervical infection by oncogenic HPVs are at high risk for developing cervical cancer [136, 137]. Infection with oncogenic HPV-16 and HPV-18, the two most common oncogenic HPV types, leads to the development of almost 70% of all cervical and other types of anogenital cancers. Viral E6 and E7 of the oncogenic HPVs are two viral oncoproteins that inactivate, respectively, cellular p53 and pRB, which are two tumor suppressor proteins essential for cell cycle control [138, 139]. In cervical cancer tissues and cervical cancer-derived cell lines, E6 and E7 oncogenes are highly expressed; the majority of the E6/E7 bicistronic RNA are alternatively spliced as diagramed for HPV-16 and HPV-18. A major spliced RNA isoform of viral E6/E7 bicistronic RNA is E6*I derived from splicing of nt 226 5' splice site to nt 409 3' splice site for HPV-16 and of nt 233 5' splice site to nt 416 3' splice site for HPV-18 (Fig. 15b). It has been demonstrated that this RNA splicing is necessary for viral E7 translation [140] and can be easily detected by RNase protection assay (RPA) or by RT-PCR methods [140, 141].

The presence of the high-grade premalignant lesions (CIN, cervical intraepithelial neoplasia) caused by oncogenic HPV infection is a sign of increased risk for developing cervical cancer. These lesions can be detected by routine cervical examination and treated by surgery to prevent progression to cervical cancer. The Papanicolaou test (also called the Pap smear) is a screening test used in gynecology to detect premalignant and malignant cells in cervical swabs. A woman who has a Pap smear with abnormal cells may also be referred for HPV DNA testing by two FDA-approved assays: the Hybrid Capture 2 DNA test, which detects 13 high-risk HPVs (HPV-16, HPV-18, HPV-31, HPV-33, HPV-35, HPV-39, HPV-45, HPV-51, HPV-52, HPV-56, HPV-58, HPV-59, and HPV-68) and is available from Qiagen [142], or the Cobas 4800 System HPV test, which detects 14 high-risk HPVs (HPV-16, HPV-18, HPV-31, HPV-33, HPV-35, HPV-39, HPV-45, HPV-51, HPV-52, HPV-56, HPV-58, HPV-59, HPV-66, and HPV-68) and is available from Roche [143]. A few of the HPV E6/HPV E7 RNA tests also have been introduced. The APTIMA HPV Assay from Hologic was designed to detect



C

Primer	Position (nt)	Sequence	Amplicons (bp)
HPV16- Pr106	106-125	5'-GTTTCAGGACCCACAGGAGC-3'	DNA:(457), RNA:E6/E7 unspl. (457), E6*I/E7 (275), E6*II (158)
HPV16-Pr562	545-562	5'-TGATTACAGCTGGGTTTC-3'	
HPV18-Pr121	121-140	5'-ATCCAACACGGCGACCCTAC-3'	DNA: (730) RNA: E6/E7 unspl. (730), E6*I/E7 (548)
HPV18-Pr850	833-850	5'-CTGGAATGCTCGAAGGTC-3'	

Fig. 15 RNA splicing of viral oncogene E6 and E7 transcripts in high-risk human papillomavirus infections. (**a**) Genome structure of high-risk HPV-16 divided by early (genes E1–E7, open boxes) and late (L1–L2, gray boxes) regions and positions of splice sites in the HPV-16 genome. P, promoter; pA_E, early polyadenylation site; pA_L, late polyadenylation site. (**b**) Alternative splicing of HPV-16 and HPV-18 E6–E7 regions. Open boxes represent E6 and E7 ORFs with their corresponding start and stop codon positions. Transcripts derived from promoter P97 have an intron (dashes) in the E6 and E7 ORF with three alternative 3' splice sites as diagramed. Filled black boxes are exons. Coding potentials for each transcript are showed on the left. Arrows below the transcripts are the primers used for detection of spliced E6E7 transcripts detailed in (C). The diagrams are modified from Zheng and Baker [133] and Wang et al. [141]. (**c**) Sequences and nucleotide positions of two sets of primer pairs in (B) for RT-PCR amplification of alternatively spliced E6–E7 transcripts expressed in HPV-16 and HPV-18 infections [140, 141]. Primer nucleotide positions and sequences are based on corresponding HPV reference strains available on https://pave.niaid.nih.gov/

HPV E6/HPV E7 mRNA from 14 high-risk types (HPV-16, HPV-18, HPV-31, HPV-33, HPV-35, HPV-39, HPV-45, HPV-51, HPV-52, HPV-56, HPV-58, HPV-59, HPV-66, and HPV-68) [144] with a sensitivity and specificity similar to or better than the Hybrid Capture 2 DNA test [145, 146]. The PreTect HPV-Proofer from PreTect was designed to detect E6 and E7 RNA from HPV types 16, 18, 31, 33, and 45 [147, 148] and is more specific than the HC2 for identifying women with CIN 2+ but has a lower sensitivity [149]. By using the primers detailed in Fig. 15c for RT-PCR assays, the spliced E6/E7 RNAs of HPV-16 and HPV-18 can be easily detected due to an amplicon size smaller than E6/E7 DNA, without worry of any carry-over viral DNA contamination commonly encountered with HPV DNA tests.

Herpesviruses

Herpesviruses are large DNA viruses with a complex life cycle. Their relatively large linear double-stranded DNA (dsDNA) genome (~100–200 kb) is encapsulated in a capsid with icosahedral architecture. The capsid is covered with a heterogeneous layer of viral proteins and RNAs called tegument. Outside of this tegument is a lipid bilayer membrane (envelope) containing several virus-encoded glycoproteins. A hallmark of herpesvirus infection is the establishment of a lifelong "latent" infection in their host following initial infection. Latent virus is often reactivated by various stimuli and causes recurrent infections, which is a typical feature of all herpesviruses.

Currently there are more than 100 known herpesviruses infecting a wide range of animal species. All human herpesviruses belong to the *Herpesviridae* family which is further grouped into four subfamilies: *Alphaherpesvirinae*, *Betaherpesvirinae*, *Gammaherpesvirinae*, and unassigned viruses. Currently, eight herpesvirus species have been isolated from humans; these have been assigned to three subfamilies of *Herpesviridae*. These include (1) herpes simplex virus type 1 [HSV-1, also referred as human herpesvirus 1 (HHV1)], (2) herpes simplex virus type 2 (HSV-2 or HHV2), (3) varicella-zoster virus (VZV or HHV3), (4) Epstein-Barr virus (EBV or HHV4), (5) human cytomegalovirus (CMV or HCMV or HHV5), (6) human herpesvirus 6 (HHV6), (7) human herpesvirus 7 (HHV7), and (8) Kaposi sarcoma-associated herpesvirus (KSHV or HHV8).

After viral entry, the viral genome is translocated to the nucleus of the infected cell where the expression of viral genes and viral genome replication occurs. All herpesviruses have two types of viral life cycle, latent and lytic, with each having a distinctive transcriptional profile. Latent infection is characterized by the expression of a few viral genes (latent transcripts) that maintain the viral genome in latently infected cells. Lytic infection is associated with viral genome replication and production of infectious virions; this generally leads to destruction of the infected cell. In contrast to latent infection, almost all viral lytic genes in the lytic infection are expressed in a timely regulated fashion and are divided, based on their dependence on viral protein expression and viral genome replication, into three kinetic classes:

immediate early, early, and late. In some circumstances, the virus in latently infected cells may be reactivated and proceeds to lytic infection. The mechanisms controlling the establishment of latency and reactivation of herpesviruses are not yet fully understood. The human herpes viral genome encodes up to 100 different genes including a variable number of noncoding genes expressing noncoding RNAs or viral miRNAs [150–152].

Most human herpesviruses are highly prevalent in the general population. Initial infection generally occurs in childhood or early adolescence through body contact and is followed by the establishment of latent infection. Some herpesviruses are sexually transmitted. Blood transfusion, tissue transplantation, and/or congenital transmission are additional mechanisms for acquiring the virus. The primary infection often occurs in epithelia, i.e., the point of entry, followed by establishment of latent infection, which generally occurs in a specialized cell type (neurons or lymphocytes) and serves as a virus reservoir. Recurrence of infection is caused by virus reactivation from its latent state with the virus escaping from the host immunological surveillance. Overall symptoms of herpesvirus infections in healthy individuals are generally mild, but may be life threatening in immunocompromised patients. While it is clear that infections by some herpesviruses such as EBV and KSHV are etiologically linked to the development of several types of cancer, the role of other human herpesviruses in cell transformation remains unknown [153]. Several antiviral compounds are used to treat acute herpesvirus infections. The only vaccine against herpesviruses currently approved for use in clinic is varicella/chickenpox vaccine against VZV.

The infections by herpesviruses are most commonly diagnosed by the presence of specific antibodies and antigens or by detection of viral DNA by PCR. However, without quantification at multiple time points, these techniques are unable to distinguish virus carriers from patients with active virus replication. Detection of viral transcripts associated with virus lytic phase by RT-PCR provides indication of active virus replication, but often leads to a false-positive result due to viral DNA contamination. Such DNA contamination problems could be avoided by selection of an amplicon over the intron in spliced viral transcripts; a specific product of the spliced RNA could be distinguished from its corresponding DNA based on its size. The number of spliced viral transcripts varies from one herpesvirus to another, ranging from only a handful of split genes in HSV-1 to about 30% in KSHV [154]. Both latent and lytic genes may have an intron and sometimes these are alternatively spliced.

Herpes Simplex Virus Type 1

HSV-1, a member of *Alphaherpesvirinae* subfamily, is a human neurotropic herpesvirus associated with *herpes labialis*, Bell's palsy, and *vestibular neuritis* [155]. After initial infection, HSV-1 establishes a latent infection in sensory neural ganglia from where it can periodically reactivate. The viral genome consists of two unique regions (long and short) flanked with the inverted repeat regions (internal or

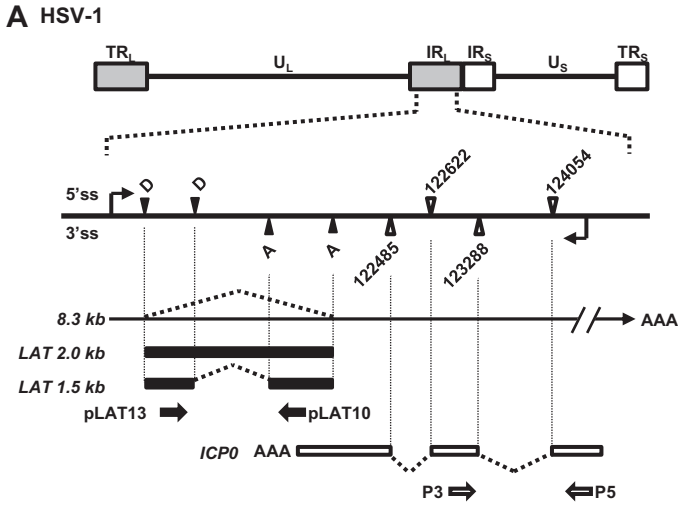
terminal) (Fig. 16a). HSV-1 encodes at least 84 genes. Most of these genes are named according to their position within a particular part of the viral genome such as UL1 (unique long region ORF1) or US3 (unique short region ORF3), while others have alternative historical names, such as ICP0 (infected cells protein 0). Only a few HSV-1 transcripts are spliced (LAT, ICP0, UL15, US1, US12/ICP47) (see review [156]), including both latent and lytic transcripts.

During latency, HSV-1 expresses LAT (latency-associated transcript) RNA using a repeat region of the viral genome called LAT-DNA [157, 158]. Two forms of LAT RNAs are detectable in latently infected neurons. A major 2.0-kb RNA is produced by splicing of a capped and polyadenylated 8.3-kb primary transcript and represents a unique stable intron while the spliced exonic RNA is unstable and quickly degraded [159]. A minor 1.5-kb RNA is generated by further splicing of the 2.0-kb RNA by removal of an internal intron of 559 or 556-bp, depending on the virus strain [160] (Fig. 6a). Both LAT RNAs are uncapped without a poly(A) tail and accumulate in the nucleus of infected cells. HSV-1 LAT RNA is a noncoding regulatory RNA for establishment and maintenance of viral latency by inhibiting the expression of viral lytic genes and thus interfering with the cellular apoptosis pathway [161]. Recent studies have demonstrated that LAT transcript functions as a precursor for the generation of virus-encoded miRNAs [162]. The expression of LAT-DNA also has been observed during lytic infection. Lytic LAT transcripts differ from latent LAT RNA by the presence of a poly(A) tail [163].

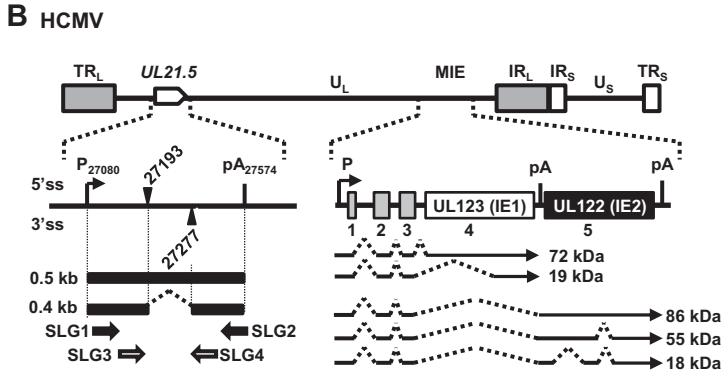
ICP0 (IE110) is encoded by a gene located in the viral genome repeat region and partially overlaps with LAT transcripts. Antisense expression of LAT transcripts inhibits the expression of ICP0 during latency. ICP0 is an immediate-early gene expressed in the early stage of lytic infection. ICP0 functions as a non-specific transactivator and a cofactor of another viral transactivator ICP4 [164]. ICP0 initiates lytic replication in both newly infected cells as well as after reactivation in cells with latent infection. ICP0 is transcribed in reverse orientation from viral genome, and its pre-mRNA contains three exons separated by two introns [165] (Fig. 16a). After splicing, the mature mRNA encodes ICP0 protein with 775-aa residues. An alternatively spliced ICP0 transcript retaining intron 2 is detectable in the infected cells [166] and encodes a truncated ICP0R in size of 262-aa residues due to the presence of a stop codon in the intron 2. Thus, both ICP0 and ICP0R have the same aa sequences in the N-terminal part. ICP0R functions as a repressor of viral expression [167].

HSV-2 represents another important human pathogen belonging to the alphaherpesvirus subfamily. Genital infection with HSV-2 causes genital herpes, which is considered to be a sexually transmitted disease. HSV-2 is also neurotropic and establishes latent infection in sacral ganglia. HSV-1 and HSV-2 are two closely related viruses with similar genomes and gene structures, including their LAT and ICP0 regions [168].

In general, the infection of HSV-1 and HSV-2 is controlled by the host immune system. Thus, initial or recurrent infections are usually associated with only mild symptoms. Infection in immunocompromised patients can cause several severe diseases including encephalitis [169]. Genital infections or reactivation of HSV-2 during pregnancy can lead to congenital infection [170]. Detection of viral DNA may



Primer	Position (nt)	Sequence	Amplicons (bp)
pLAT10	120408-120389	5'-ATGGAGCCAGAACCACAGTG-3'	DNA: (947)
pLAT13	119464 -119484	5'-GTAGGTTAGACACCTGCTTCT-3'	RNA: LAT 2.0 kb (947), LAT 1.5 kb (297)
P3	123214-123233	5'-TTCGGTCTCCGCCTGAGAGT-3'	DNA: (922)
P5	124116-124135	5'-GACCCTCCAGCCGCATACGA-3'	RNA: ICP0 unspl. (922), ICP0 spl. 1.5 kb (157)



Primer	Position (nt)	Sequence	Amplicons (bp)
SLG1	27120-27139	5'-CTATGGATCTTGAGCTTACT-3'	
SLG2	27410-27429	5'-TCGCTGCCATCTCCGTCTGT-3'	DNA: (258)
SLG3	27144-27163	5'-GTGACCTTGACGGTGGCTTT-3'	RNA: unsliced (258) spliced (175)
SLG4	27382-27401	5'-CGTCATACTCCCCGGAGTAA-3'	

Fig. 16 RNA splicing of LAT and ICP0 transcripts in HSV-1 infections and UL21.5, UL122, and UL123 transcripts in HCMV infections. (**a**) Genome of herpes simplex virus type 1 (HSV-1)

not provide sufficient information about virus replication status due to the permanent presence of viral DNA in the infected cells. Detection of viral lytic products, such as spliced ICP0 RNA, may be a better predictor of virus reactivation and may be seen even before the occurrence of clinical symptoms allowing early diagnosis and enabling early treatment. Disappearance of the detectable lytic products could be a sign of treatment efficiency since the viral transcripts disappear earlier than viral DNA.

Human Cytomegalovirus

Human cytomegalovirus (HCMV) together with HHV-6 and HHV-7 belongs to the *Betaherpesvirinae*. A high prevalence of CMV infection has been noted in 50–80% of the human population. In a majority of healthy individuals, the primary CMV infection occurs asymptomatically but, in some cases, can be associated with sore throat, prolong fever, or a syndrome similar to infectious mononucleosis. After initial infection, the virus usually remains latent in T cells for the rest of the host life without apparent symptoms. In contrast, CMV infections in immunocompromised individuals, such as newborns, transplant recipients, persons with AIDS, or cancer patients, can lead to severe disease and even death. The symptoms include *hepatitis*, *retinitis*, *colitis*, *pneumonia*, *encephalitis*, and others.

CMV has a large genome of about 220 kb capable of encoding approximately 200 genes (reviewed in [171]). While the majority of CMV transcripts do not have introns, the presence of several split genes has been identified in all kinetic classes of the viral genes [172]. A major immediate-early region (MIE) located within a unique long (UL) region of the CMV genome contains several genes that are highly expressed during the early stage of viral lytic infection. These include UL123 (IE1), UL122 (IE2), and UL119-115. MIE transcripts contain multiple introns and undergo complex alternative RNA splicing. MIE transcripts IE1 and IE2 are expressed from the same promoter but are alternatively polyadenylated. These transcripts have five major exons and can be alternatively spliced to express additional isoforms of IE1

Fig. 16 (continued) consists of two unique regions (long U_L and short U_S) flanked with terminal (TR) and internal (IR) repeats. Shown below are two representatives of HSV-1 RNA splicing, latency-associated transcript (*LAT*) and immediate-early *ICP0* transcripts with positions of 5' and 3' splice sites, and primer pairs (arrows below) used for detection of spliced transcripts as detailed in the table further below. Boxes (full for *LAT* and empty for *ICP0*) represent exons divided by introns (dashed lines). The primer pairs in the table were described by Tanaka et al. [214], with nucleotide positions in HSV-1 genome (GenBank Acc. No. X14112.1). (**b**) Genome structure of human cytomegalovirus. See other details in (**a**). Shown below the genome structure are spliced transcripts of late *UL21.5* gene and two immediate-early genes *UL123 (IE1)* and *UL122 (IE2)* expressed from a major immediate-early region (MIE). Boxes or solid lines are exons and dashes are introns or splicing directions. P, promoter; pA, polyadenylation site. Arrows below *UL21.5* transcripts are primer pairs used for detection of spliced *UL21.5* transcripts by nested RT-PCR as described [174] and detailed in the table below, with the nucleotide positions in HCMV genome (strain AD169, GenBank Acc. No. X17403.1)

and IE2 proteins (Fig. 16b). Splicing also was detected in transcripts from other CMV genes such as TRL4, UL89, US3, R160461, and R27080 [172]. Gene UL21.5 (previously named as R27080) is one of the known CMV split late genes (SLG). The UL21.5 transcript that encodes viral glycoprotein is expressed from the UL region posited from nt 27080 to nt 27574 of CMV genome [173] and has a short intron of 83 nts. Removal of this intron leads to production of a mature mRNA in size of ~0.4 kb. Both spliced and unspliced UL21.5 RNAs are easily detectable by RT-PCR from infected cells [174].

Allogenic bone marrow transplant recipients are at high risk for developing CMV diseases. Historically, viremia has been used as an indicator of CMV disease as well as to guide preemptive treatment. Multiple approaches have been developed to detect CMV viremia in circulating lymphocytes by direct virus isolation with cultivation, by detection of viral antigens in polymorphonuclear cells, or by quantitative viral DNA [175, 176]. However, the detection of viremia is not sufficient for disease prediction since many viremic patients never develop symptoms. However, active CMV replication in peripheral blood lymphocytes can be verified by analyzing viral mRNAs [177]. Amplification of spliced viral transcripts has some advantages in comparison to intronless transcripts and is not affected by DNA contamination. Detection of spliced immediate-early transcripts had been reported to have a good correlation with the detection of viral DNA or viral antigen [178–180]. Detection of late gene UL21.5 has a better prediction value and has a significant correlation with disease progression [174, 181, 182].

Epstein-Barr Virus

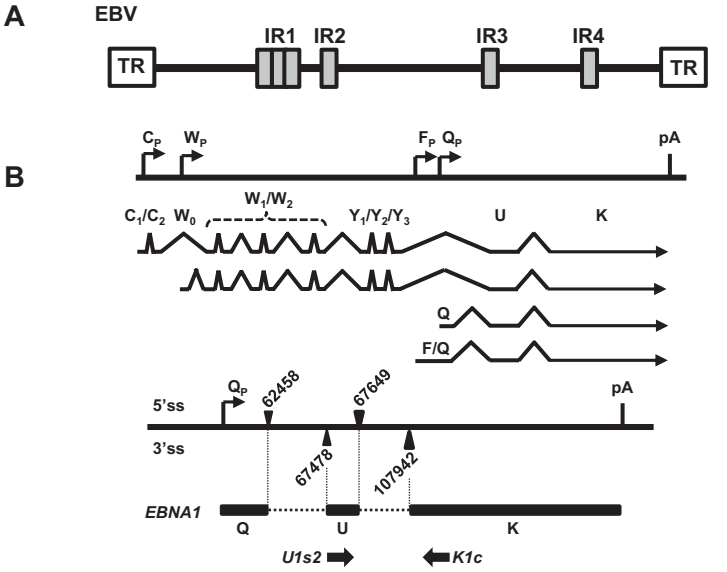
Epstein-Barr virus (EBV), a well-characterized member of the *Gammaherpesvirinae* subfamily, is an important human pathogen. EBV infection is highly prevalent, with more than 95% of the human population becoming seropositive in early life. While primary infection during childhood is usually unremarkable, the virus acquisition in adolescence and adulthood is often associated with the development of the *infectious mononucleosis syndrome*. In healthy individuals, the EBV infection is well controlled by the immune system. However, EBV remains in long-living memory B cells where it establishes a latent infection. EBV is an oncogenic virus capable of transforming the infected B cells [183]. EBV infections have been associated with the development of a number of human malignancies, including *nasopharyngeal carcinoma*, *Burkitt's lymphoma*, *Hodgkin's lymphoma*, *gastric carcinoma*, and others (see review [184]). Active EBV replication due to immunosuppression may cause *posttransplant lymphoproliferative disease* [185]. During latent infection, EBV expresses six nuclear antigens (EBNA-1, EBNA-2, EBNA-3A, EBNA-3B, EBNA-3C, and EBNA-LP), three latent membrane proteins (LMP-1, LMP-2A, and LMP-2B), and several noncoding transcripts (EBER-1 and EBER-2 and BARTs). Many EBV latent products are defined as oncogenes and are responsible for EBV-mediated cell transformation [115]. Several types of EBV latency have been defined by variable expression of latent genes in a number of malignancies [186].

The EBV genome (~172 kb) is flanked by multiple terminal repeats (TR) and can be divided into a long and a short unique region (U_L and U_S) by internal repeat 1 (IR1) (Fig. 17a). EBV encodes at least 80 viral proteins [187] as well as several noncoding RNAs including viral miRNAs [188, 189]. The ORF names are derived from their positions in the *Bam*HI fragment (from A to Z) by orientation (L, left; or R, right) and a digital number representing the frame (F) order (e.g., BZLF1). Other genes retain their historical names based on the gene product function. The number of split genes in EBV is significantly higher than that found in alpha- and beta-herpesviruses. Extensive alternative RNA splicing is prominent, particularly for almost all of the EBV latent transcripts, but the transcripts of many lytic genes are also spliced.

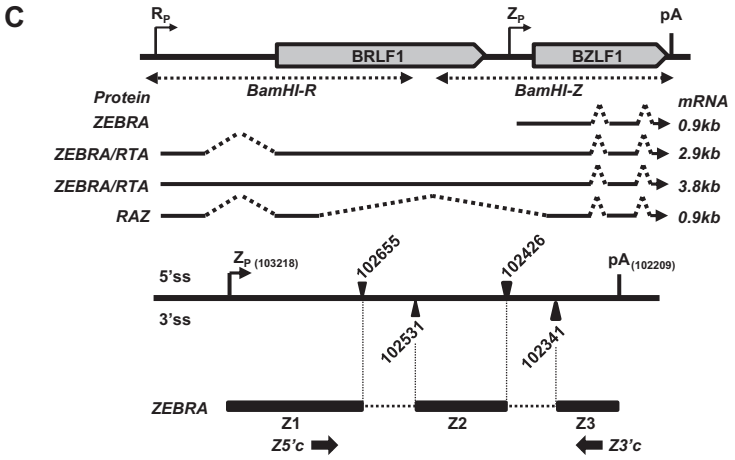
EBNA-1 is a multifunctional viral protein critical for establishing and maintaining EBV latency and for regulation of viral promoter activities [190]. In infected cells, EBNA-1 is expressed from a spliced mRNA derived from a primary transcript of ~100 kb. This transcript originates from one of two alternative promoters, *Cp* or *Wp*, which are named by their localization in different *Bam*HI fragments of the viral genome (Fig. 17b). At the early stage of latent infection, *Wp* is initially used, but EBNA-1 and EBNA-2 expressed from *Wp* transactivate the *Cp* promoter and cause a switch of transcription from *Wp* to *Cp* [191]. Usage of the *Cp* promoter is associated with EBV "latency type III." In Burkitt's lymphoma and Burkitt's lymphoma-derived cell lines, EBNA-1 expression is initiated from the distal *Qp* promoter rather than from *Cp* and *Wp* and is associated with "latency type I" [192]. EBNA-1 is also expressed in the lytic phase from an additional *Fp* promoter located closely upstream to the *Qp* promoter [193].

The establishment of active EBV replication after virus reactivation from latency is dependent on the expression of two immediate-early genes, BZLF1 and BRLF1 [194], and encodes viral transactivators ZEBRA (BZLF1) and RTA (BRLF1). Although BRLF1 and BZLF1 are transcribed separately from a different promoter with the *Rp* for BRLF1 and the *Zp* for BZLF1, both gene transcripts utilize the same polyadenylation site for RNA polyadenylation [195] (Fig. 17c). Thus, the *Zp* promoter-derived transcript is a monocistronic ZEBRA RNA containing two constitutive introns; splicing of these two introns results in production of a 0.9-kb mRNA that encodes ZEBRA protein. Transcription from the *Rp* promoter leads to production of a 3.8-kb bicistronic transcript, ZEBRA/RTA, which contains two additional introns and the ZEBRA RNA. Splicing of the intron 1 in the 5' noncoding region of ZEBRA/RTA transcript leads to production of a 2.9-kb RNA as a major RNA isoform. However, both isoforms of ZEBRA/RTA RNA have the potential to encode ZEBRA and RTA proteins. A third minor isoform of ZEBRA/RTA transcript is derived from the splicing of an additional internal intron spanning from BRLF1 ORF to BZLF1 ORF; this splicing produces a RAZ transcript of ~0.9 kb encoding a RTA-ZEBRA fusion protein, RAZ. RAZ may function as an inhibitor to ZEBRA during EBV infection [196].

Transcripts for EBNA-1 are believed to be expressed in all forms of EBV latent infection, except latently infected nondividing B cells having "latency type 0." This makes detection of EBNA-1 expression a good marker for the presence of EBV in tumors. The expression of ZEBRA during lytic infection could be used to monitor productive EBV infection as well as EBV reactivation.



Primer	Position (nt)	Sequence	Amplicons (bp)
U1s2	67595-67614	5'-ATAGACCGCCAGTAGACCTG-3'	DNA: (~40kb) RNA: EBNA1 unspliced (~40kb), EBNA1 spliced (178)
K1c	108064-108045	5'-GGTTATCACCCCCTCTTCTT-3'	



Primer	Position (nt)	Sequence	Amplicons (bp)
Z5'c	102676-102657	5'-GAAACCACAACAGCCAGAAT-3'	DNA: (457) RNA: ZEBRA unspliced (457), ZEBRA spliced (250)
Z3'c	102220-102239	5'-AAGAGATCCTCGTGTAAAAC-3'	

Fig. 17 RNA splicing of EBNA1 and ZEBRA transcripts in EBV infections. (**a**) Organization of EBV genome with terminal repeats (TR) and internal repeats (IR1-4). (**b**) Multiple transcripts of

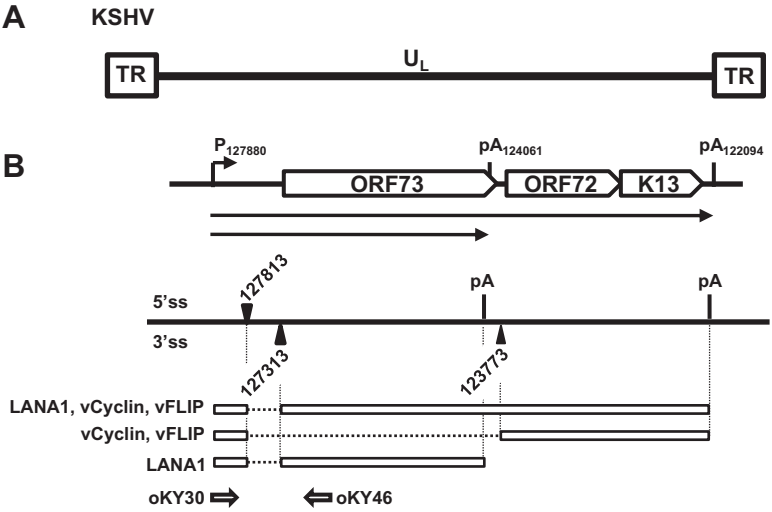
Kaposi Sarcoma-Associated Herpesvirus

Kaposi sarcoma-associated herpesvirus (KSHV) is the latest human herpesvirus to be discovered [197]. After primary infection, KSHV establishes latent infection in endothelial cells as well as B cells [198]. In healthy individuals, both primary and latent KSHV infections are generally asymptomatic. Suppression of the immune system in KSHV-positive individuals, such as in AIDS patients or tissue transplant recipients, is associated with the development of several cancers, including all forms of *Kaposi sarcoma* (a solid tumor of endothelial origin) or B-cell lymphomas [*primary effusion lymphoma* (PEL) and *multicentric Castleman's disease* (MCD)] (see review [199]). The presence of the viral genome as well as expression of viral-encoded products in all cancer cells strongly suggests the active role of KSHV in cell transformation.

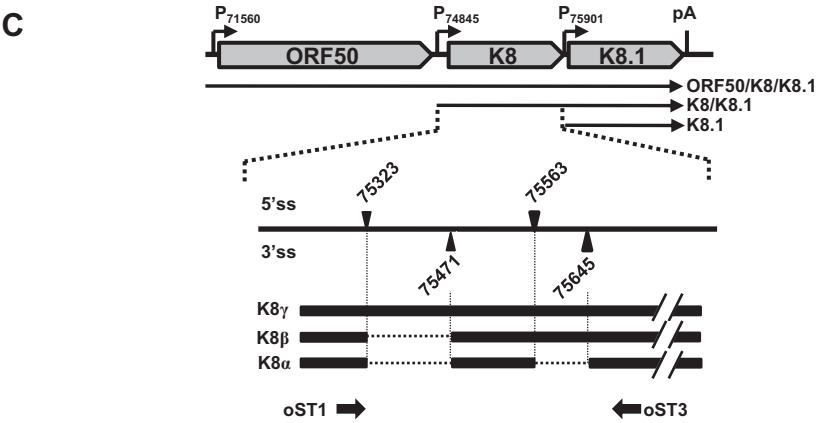
KSHV belongs to the *Gammaherpesvirinae* and has a similar genome organization as does EBV, with a long unique region flanked with terminal repeats (Fig. 18a). The KSHV genome (~165 kb) encodes up to 90 genes that are named by their position in the viral genome from left to right (e.g., ORF47) [200]. Some KSHV unique genes are designated with a digital K number, like K2, while some have alternative names based on their function (ORF57 or MTA-*mRNA transcript accumulation*). KSHV transcripts derived from ~30% of the viral genes, including both latent and lytic genes, undergo RNA splicing [154].

During latency, the KSHV genome expresses a latency-associated nuclear antigen-1 (LANA-1) [201] from ORF73. The gene ORF73 posits along with ORF72 and K13 in a larger latent locus of the virus genome. The latter two genes encode viral homologues of cellular proteins vCyclin (ORF72) and vFLICE (K13). ORF73/72/K13 are transcribed from a single promoter (P_{127880}) as a tricistronic RNA containing an intron with two alternative 3' splice sites. Alternative RNA splicing and alternative RNA polyadenylation of the tricistronic pre-mRNA result in production of three mature mRNAs (5.4, 3.3, and 1.7 kb) [202] (Fig. 18b). The 5.4-

Fig. 17 (continued) EBV latency-associated *EBNA-1* transcribed from several alternative promoters (C_p-Q_p). EBNA-1 RNA contains multiple exons (C_1-*K*, lines) and introns (half triangles). A detailed *EBNA-1* transcript derived from the Q_p promoter is shown below with splice sites (black triangles) and splice directions (dashed lines for introns). A primer pair used to detect the spliced *EBNA-1* transcript from exon *U* to exon *K* [185] is detailed in the table below. (**c**) Gene structures of *BRLF1 (RTA)* and *BZLF1 (ZEBRA)* (two EBV-immediate-early genes) and their spliced RNA products. EBV *RTA* and *ZEBRA* are transcribed by two alternative promoters, R_P and Z_P, but polyadenylated by using the same polyadenylation signal downstream of *ZEBRA* ORF. Thus, the bicistronic *RTA* transcript derived from *Rp* promoter contains multiple introns and has potentials to encode *RTA*, *ZEBRA*, and *RAZ* proteins by alternative RNA splicing, whereas the monocistronic *ZEBRA* transcript derived from the *Zp* promoter encodes only *ZEBRA* protein and also contains multiple introns as detailed further below with nucleotide positions of splice sites, exons (*Z1–Z3*, black boxes), and introns (dashed lines). Arrows below exons *Z1* and *Z3* are a primer pair used for detection of spliced *ZEBRA* mRNA [185] and detailed in the table, with nucleotide positions in EBV genome (strain B95-8, GenBank Acc. No. V01555.2)



Primer	Position (nt)	Sequence	Amplicons (bp)
oKY30	127828-127846	5'-AGAGCAGCAGCTTGGTCCG-3'	DNA: (952)
oKY46	126877- 126895	5'-CTGGAGACTGCGTGGGTGG-3'	RNA: LANA1 unspliced (952), LANA1 spliced (453)



Primer	Position (nt)	Sequence	Amplicons (bp)
oST1	75182-75200	5'-ACCACCAAGAGGACCACACATTTC-3'	DNA:(657)
oST3	75813-75838	5'-CACACAAAGTCTGGCATGGTTCTCCC-3'	RNA: K8γ (657) K8β (510), K8α (428)

Fig. 18. RNA splicing of representative latent transcripts and early transcripts in KSHV infections. (**a**) Genome of KSHV contains a long unique region (U_L) flanked by terminal repeats (TR).

kb transcript most likely responsible for LANA-1 expression is produced by usage of the proximal 3' splice site, whereas usage of the distal 3' splice site leads to expression of 1.7-kb transcripts for vCyclin and vFLICE. Both transcripts are polyadenylated at the same distal polyadenylation site. The minor 3.3-kb transcript uses the proximal splice site for RNA splicing but is polyadenylated at a proximal non-canonical polyadenylation signal (Fig. 18b).

KSHV lytic replication is controlled by a major viral transactivator, ORF50 (also referred as Rta) [203, 204]. Like LANA-1, ORF50 posits along with K8 and K8.1 in a larger gene locus (ORF50/K8/K8.1 cluster) (Fig. 18c) and is expressed as an immediately early transcript during lytic virus replication. K8 encoding a viral k-bZIP protein is an early gene and K8.1 encoding a glycoprotein is a late gene. Although each of the three genes bears its own promoter, all of their RNA transcripts use a single polyadenylation site located downstream of K8.1 gene and undergo alternative RNA splicing (see review [154]). Thus, the 3' portion of ORF50 transcript is homologous to K8 and K8.1 and has the same intron and exon structures as seen in the K8 and K8.1 transcripts. The ORF50 transcript is tricistronic, K8 is bicistronic, and K8.1 is monocistronic in nature. The bicistronic K8 transcript is composed of four exons separated by three introns (Fig. 18c). A functional K8α protein is expressed from a fully spliced mRNA, but retention of the intron 2 in K8β mRNA results in the expression of a minor form K8β protein [205]. An unspliced K8 RNA, K8γ, is also detectable, but rare in lytically infected cells.

In summary, LANA-1 expression is a hallmark of KSHV latent infection. Transcripts originated from the ORF73/72/K13 gene cluster are expressed in latently infected, KSHV-transformed cells and are detectable by RT-PCR. Active virus replication is associated with the expression of viral lytic genes. Amplification of the spliced K8 region that detects the expression of both ORF50 tricistronic and K8 bicistronic transcripts could be used to monitor viral lytic replication [206].

Fig. 18 (continued) (**b**) Gene structure of a latent gene locus containing *ORF73/ORF72/K13* genes. Three genes are transcribed from a single promoter (P) as a polycistronic RNA. This polycistronic primary transcript is processed by alternative RNA splicing and alternative polyadenylation (pA, see arrows below). Nucleotide positions of splice sites (triangles) and polyadenylation site in the KSHV genome (GenBank Acc. No. U75698.1) are diagramed further below with coding potentials of each spliced product on the left. Open boxes are exons and dash lines are introns or splicing directions. oKY30 and oKY46 are a primer pair used to detect spliced LANA transcripts [215] as detailed in the table below. (**c**) Gene structure of a KSHV lytic locus consisting of immediate-early *RTA (ORF50)*, early *K8*, and late *K8.1* genes. Three genes are expressed from three separate promoters (P), but all of their transcripts (full lines with arrows immediately below to the right) are polyadenylated at the same polyadenylation site (pA), resulting in *RTA* as a tricistronic, *K8* as a bicistronic, and *K8.1* as a monocistronic transcript. The enlarged *K8* coding region contains three exons (filled boxes) and two introns (dashed lines), with nucleotide positions of each splice site in the KSHV genome (GenBank Acc. No. U75698.1). Names of three common forms of *K8* transcripts from alternative RNA splicing are shown on the left. oST1 and oST2 are two primers used to detect spliced *K8* RNAs as described [215, 216], and are detailed in the table below

Conclusion

The major aim of this chapter is to provide readers with knowledge of viral RNA splicing during viral infection as well as how the detection of these spliced viral RNA transcripts can be used as a new approach in diagnostic virology. In the first part, basic information about the mechanisms of RNA splicing and the methodological approaches for specific detection of these spliced RNA molecules is provided. The core of these techniques represents an amplification and detection of nucleic acids. The advantage of nucleic acid-based techniques is the application of the same platform for detection of various viral pathogens, often at the same time, by multiplexing. The rapid setup of these methods is especially important for a rapid response to emerging viruses as has been successfully proven in the case of severe acute respiratory syndrome (SARS), avian influenza, Zika virus, and Ebola virus outbreaks where nucleic acid amplification was rapidly deployed to detect and to confirm these infections. The low material requirement and their simplicity make these detection methods suitable for applications in low resources setting such as laboratories where the first contact is seen as well as field laboratories. Because the genomic sequences for many viruses can be detected by amplification of the nucleic acid molecules as a routine procedure in many diagnostic laboratories, the detection of spliced viral transcripts could be performed simultaneously using already existing methods.

The second part of this chapter summarizes the current knowledge of viral RNA splicing events for the majority of known human viruses. Some unique viral agents, such as human circoviruses and adeno-associated viruses, where a direct link between infection and pathological manifestation remains to be determined, have been included. In addition, examples of each virus, where the detection of spliced viral RNA could bring additional benefit to current techniques to improve the disease prognosis or better monitoring of efficiency of therapeutic intervention, have been provided. Systematic study of RNA splicing events during viral infection is likely to lead to better viral diagnostics and better management of viral therapy and will eventually lead to a better understanding of the pathogenesis of these human viral pathogens.

References

1. Wiedbrauk DL. Nucleic acid amplification and detection methods. In: Sprcter S, Hodinka RL, Young SA, Wiedbrauk DL, editors. Clinical virology manual. 4th ed. Washington, DC: ASM Press; 2009. p. 156–84.
2. Berget SM, Moore C, Sharp PA. Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A. 1977;74(8):3171–5.
3. Chow LT, Gelinas RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell. 1977;12(1):1–8.
4. Fong N, Bentley DL. Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. 2001;15(14):1783–95.

5. Maniatis T, Reed R. An extensive network of coupling among gene expression machines. Nature. 2002;416(6880):499–506.
6. Tarn WY, Steitz JA. A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell. 1996;84(5):801–11.
7. Wu Q, Krainer AR. AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. Mol Cell Biol. 1999;19(5):3225–36.
8. Wachtel C, Manley JL. Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Mol Biosyst. 2009;5(4):311–6.
9. Toor N, Rajashankar K, Keating KS, Pyle AM. Structural basis for exon recognition by a group II intron. Nat Struct Mol Biol. 2008;15(11):1221–2.
10. Chasin LA. Searching for splicing motifs. Adv Exp Med Biol. 2007;623:85–106.
11. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J. 2009;417(1):15–27.
12. Han SP, Tang YH, Smith R. Functional diversity of the hnRNPs: past, present and perspectives. Biochem J. 2010;430(3):379–92.
13. Kornblihtt AR. Coupling transcription and alternative splicing. Adv Exp Med Biol. 2007;623:175–89.
14. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T. Epigenetics in alternative pre-mRNA splicing. Cell. 2011;144(1):16–26.
15. Zheng ZM. Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. J Biomed Sci. 2004;11(3):278–94.
16. Wang ET, Sandberg R, Luo S, et al. Alternative isoform regulation in human tissue transcriptomes. Nature. 2008;456(7221):470–6.
17. Croft L, Schandorff S, Clark F, Burrage K, Arctander P, Mattick JS. ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat Genet. 2000;24(4):340–1.
18. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl. 1995;4(6):357–62.
19. Tyagi S, Kramer FR. Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol. 1996;14(3):303–8.
20. Whitcombe D, Theaker J, Guy SP, Brown T, Little S. Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotechnol. 1999;17(8):804–7.
21. Hughes TR, Hiley SL, Saltzman AL, Babak T, Blencowe BJ. Microarray analysis of RNA processing and modification. Methods Enzymol. 2006;410:300–16.
22. Cuperlovic-Culf M, Belacel N, Culf AS, Ouellette RJ. Microarray analysis of alternative splicing. OMICS. 2006;10(3):344–57.
23. Pinzani P, Orlando C, Pazzagli M. Laser-assisted microdissection for real-time PCR sample preparation. Mol Aspects Med. 2006;27(2-3):140–59.
24. Esposito G. Complementary techniques: laser capture microdissection--increasing specificity of gene expression profiling of cancer specimens. Adv Exp Med Biol. 2007;593:54–65.
25. Niedobitek G, Herbst H. In situ detection of Epstein-Barr virus and phenotype determination of EBV-infected cells. Methods Mol Biol. 2006;326:115–37.
26. Knoll JH, Lichter P, Bakdounes K, Eltoum IE. In situ hybridization and detection using non-isotopic probes. Curr Protoc Mol Biol 2007; Chapter 14:Unit 14.7.
27. Speel EJ, Hopman AH, Komminoth P. Tyramide signal amplification for DNA and mRNA in situ hybridization. Methods Mol Biol. 2006;326:33–60.
28. Yamamichi N, Shimomura R, Inada K, et al. Locked nucleic acid in situ hybridization analysis of miR-21 expression during colorectal cancer development. Clin Cancer Res. 2009;15(12):4009–16.
29. Bonvicini F, Mirasoli M, Gallinella G, Zerbini M, Musiani M, Roda A. PNA-based probe for quantitative chemiluminescent in situ hybridisation imaging of cellular parvovirus B19 replication kinetics. Analyst. 2007;132(6):519–23.

30. Kang JG, Majerciak V, Uldrick TS, et al. Kaposi's sarcoma-associated herpesviral IL-6 and human IL-6 open reading frames contain miRNA binding sites and are subject to cellular miRNA regulation. J Pathol. 2011;225(3):378–89.
31. Blanco AM, Artero R. A practical approach to FRET-based PNA fluorescence in situ hybridization. Methods. 2010;52(4):343–51.
32. de Planell-Saguer M, Rodicio MC, Mourelatos Z. Rapid in situ codetection of noncoding RNAs and proteins in cells and formalin-fixed paraffin-embedded tissue sections without protease treatment. Nat Protoc. 2010;5(6):1061–73.
33. Su Z, Ning B, Fang H, et al. Next-generation sequencing and its applications in molecular diagnostics. Expert Rev Mol Diagn. 2011;11(3):333–43.
34. Nakamura S, Yang CS, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One. 2009;4(1):e4219.
35. Davis CA, Grate L, Spingola M, Ares M Jr. Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res. 2000;28(8):1700–6.
36. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860–921.
37. Bouvier NM, Palese P. The biology of influenza viruses. Vaccine. 2008;26(Suppl 4):D49–53.
38. Lamb RA, Horvath CM. Diversity of coding strategies in influenza viruses. Trends Genet. 1991;7(8):261–6.
39. Lamb RA, Choppin PW. Segment 8 of the influenza virus genome is unique in coding for two polypeptides. Proc Natl Acad Sci U S A. 1979;76(10):4908–12.
40. Lamb RA, Lai CJ. Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell. 1980;21(2):475–85.
41. Briedis DJ, Lamb RA. Influenza B virus genome: sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS1 and NS2 proteins. J Virol. 1982;42(1):186–93.
42. Nakada S, Graves PN, Palese P. The influenza C virus NS gene: evidence for a spliced mRNA and a second NS gene product (NS2 protein). Virus Res. 1986;4(3):263–73.
43. Lamb RA, Lai CJ, Choppin PW. Sequences of mRNAs derived from genome RNA segment 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc Natl Acad Sci U S A. 1981;78(7):4170–4.
44. Shih SR, Suen PC, Chen YS, Chang SC. A novel spliced transcript of influenza A/WSN/33 virus. Virus Genes. 1998;17(2):179–83.
45. Chiang C, Chen GW, Shih SR. Mutations at alternative 5′ splice sites of M1 mRNA negatively affect influenza A virus viability and growth rate. J Virol. 2008;82(21):10873–86.
46. Yamashita M, Krystal M, Palese P. Evidence that the matrix protein of influenza C virus is coded for by a spliced mRNA. J Virol. 1988;62(9):3348–55.
47. Pekosz A, Lamb RA. Influenza C virus CM2 integral membrane glycoprotein is produced from a polypeptide precursor by cleavage of an internal signal sequence. Proc Natl Acad Sci U S A. 1998;95(22):13233–8.
48. Hongo S, Sugawara K, Muraki Y, et al. Influenza C virus CM2 protein is produced from a 374-amino-acid protein (P42) by signal peptidase cleavage. J Virol. 1999;73(1):46–50.
49. Stoltzfus CM. Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res. 2009;74:1–40.
50. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol. 1990;64(6):2519–29.
51. Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C. Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. FEBS J. 2010;277(4):867–76.
52. Landry S, Halin M, Lefort S, et al. Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology. 2007;4:71.

53. Fischer U, Pollard VW, Luhrmann R, et al. Rev-mediated nuclear export of RNA is dominant over nuclear retention and is coupled to the Ran-GTPase cycle. Nucleic Acids Res. 1999;27(21):4128–34.
54. Lutzelberger M, Reinert LS, Das AT, Berkhout B, Kjems J. A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. J Biol Chem. 2006;281(27):18644–51.
55. Benko DM, Schwartz S, Pavlakis GN, Felber BK. A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol. 1990;64(6):2505–18.
56. Salfeld J, Gottlinger HG, Sia RA, Park RE, Sodroski JG, Haseltine WA. A tripartite HIV-1 tat-env-rev fusion protein. EMBO J. 1990;9(3):965–70.
57. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc Natl Acad Sci U S A. 1980;77(12):7415–9.
58. Yoshida M, Miyoshi I, Hinuma Y. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc Natl Acad Sci U S A. 1982;79(6):2031–5.
59. Gessain A, Barin F, Vernant JC, et al. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet. 1985;2(8452):407–10.
60. Osame M, Usuku K, Izumo S, et al. HTLV-I associated myelopathy, a new clinical entity. Lancet. 1986;1(8488):1031–2.
61. Gessain A, de Thé G. Geographic and molecular epidemiology of primate T lymphotropic retroviruses: HTLV-I, HTLV-II, STLV-I, STLV-PP, and PTLV-L. Adv Virus Res. 1996;47:377–426.
62. Bindhu M, Nair A, Lairmore MD. Role of accessory proteins of HTLV-1 in viral replication, T cell activation, and cellular gene expression. Front Biosci. 2004;9:2556–76.
63. Cavanagh MH, Landry S, Audet B, et al. HTLV-I antisense transcripts initiating in the 3′ LTR are alternatively spliced and polyadenylated. Retrovirology. 2006;3:15.
64. Halin M, Douceron E, Clerc I, et al. Human T-cell leukemia virus type 2 produces a spliced antisense transcript encoding a protein that lacks a classic bZIP domain but still inhibits Tax2-mediated transcription. Blood. 2009;114(12):2427–38.
65. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M. A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem Biophys Res Commun. 1997;241(1):92–7.
66. Springfeld C, Bugert JJ, Schnitzler P, Tobiasch E, Kehm R, Darai G. TT virus as a human pathogen: significance and problems. Virus Genes. 2000;20(1):35–45.
67. Miyata H, Tsunoda H, Kazi A, et al. Identification of a novel GC-rich 113-nucleotide region to complete the circular, single-stranded DNA genome of TT virus, the first human circovirus. J Virol. 1999;73(5):3582–6.
68. Mushahwar IK, Erker JC, Muerhoff AS, et al. Molecular and biophysical characterization of TT virus: evidence for a new virus family infecting humans. Proc Natl Acad Sci U S A. 1999;96(6):3177–82.
69. Okamoto H, Nishizawa T, Tawara A, et al. TT virus mRNAs detected in the bone marrow cells from an infected individual. Biochem Biophys Res Commun. 2000;279(2):700–7.
70. Kamahora T, Hino S, Miyata H. Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. J Virol. 2000;74(21):9980–6.
71. Qiu J, Kakkola L, Cheng F, et al. Human circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. J Virol. 2005;79(10):6505–10.
72. Takahashi K, Iwasa Y, Hijikata M, Mishiro S. Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. Arch Virol. 2000;145(5):979–93.
73. Yang JD, Roberts LR. Hepatocellular carcinoma: A global view. Nat Rev Gastroenterol Hepatol. 2010;7(8):448–58.
74. Studach LL, Rakotomalala L, Wang WH, et al. Polo-like kinase 1 inhibition suppresses hepatitis B virus X protein-induced transformation in an in vitro model of liver cancer progression. Hepatology. 2009;50(2):414–23.

75. Seeger C, Mason WS. Hepatitis B virus biology. Microbiol Mol Biol Rev. 2000;64(1):51–68.
76. Fang ZL, Sabin CA, Dong BQ, et al. The association of HBV core promoter double mutations (A1762T and G1764A) with viral load differs between HBeAg positive and anti-HBe positive individuals: a longitudinal analysis. J Hepatol. 2009;50(2):273–80.
77. Su TS, Lai CJ, Huang JL, et al. Hepatitis B virus transcript produced by RNA splicing. J Virol. 1989;63(9):4011–8.
78. Suzuki T, Masui N, Kajino K, Saito I, Miyamura T. Detection and mapping of spliced RNA from a human hepatoma cell line transfected with the hepatitis B virus genome. Proc Natl Acad Sci U S A. 1989;86(21):8422–6.
79. Chen PJ, Chen CR, Sung JL, Chen DS. Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. J Virol. 1989;63(10):4165–71.
80. Suzuki T, Kajino K, Masui N, Saito I, Miyamura T. Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA. Virology. 1990;179(2):881–5.
81. Sommer G, Heise T. Posttranscriptional control of HBV gene expression. Front Biosci. 2008;13:5533–47.
82. Abraham TM, Lewellyn EB, Haines KM, Loeb DD. Characterization of the contribution of spliced RNAs of hepatitis B virus to DNA synthesis in transfected cultures of Huh7 and HepG2 cells. Virology. 2008;379(1):30–7.
83. Rosmorduc O, Petit MA, Pol S, et al. In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology. 1995;22(1):10–9.
84. Lin X, Ma ZM, Yao X, Zhang YP, Wen YM. Replication efficiency and sequence analysis of full-length hepatitis B virus isolates from hepatocellular carcinoma tissues. Int J Cancer. 2002;102(5):487–91.
85. Lee GH, Wasser S, Lim SG. Hepatitis B pregenomic RNA splicing--the products, the regulatory mechanisms and its biological significance. Virus Res. 2008;136(1-2):1–7.
86. Huang HL, Jeng KS, Hu CP, Tsai CH, Lo SJ, Chang C. Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. Virology. 2000;275(2):398–410.
87. Soussan P, Garreau F, Zylberberg H, Ferray C, Brechot C, Kremsdorf D. In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. J Clin Invest. 2000;105(1):55–60.
88. Assrir N, Soussan P, Kremsdorf D, Rossignol JM. Role of the hepatitis B virus proteins in pro- and anti-apoptotic processes. Front Biosci. 2010;15:12–24.
89. Berns KI. Parvovirus replication. Microbiol Rev. 1990;54(3):316–29.
90. Myers MW, Laughlin CA, Jay FT, Carter BJ. Adenovirus helper function for growth of adeno-associated virus: effect of temperature-sensitive mutations in adenovirus early gene region 2. J Virol. 1980;35(1):65–75.
91. Handa H, Carter BJ. Adeno-associated virus DNA replication complexes in herpes simplex virus or adenovirus-infected cells. J Biol Chem. 1979;254(14):6603–10.
92. Geoffroy MC, Salvetti A. Helper functions required for wild type and recombinant adeno-associated virus growth. Curr Gene Ther. 2005;5(3):265–71.
93. Brown KE. The expanding range of parvoviruses which infect humans. Rev Med Virol. 2010;20(4):231–44.
94. Gao G, Vandenberghe LH, Alvira MR, et al. Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol. 2004;78(12):6381–8.
95. Erles K, Rohde V, Thaele M, Roth S, Edler L, Schlehofer JR. DNA of adeno-associated virus (AAV) in testicular tissue and in abnormal semen samples. Hum Reprod. 2001;16(11):2333–7.
96. Tobiasch E, Rabreau M, Geletneky K, et al. Detection of adeno-associated virus DNA in human genital tissue and in material from spontaneous abortion. J Med Virol. 1994;44(2):215–22.
97. Blechacz B, Russell SJ. Parvovirus vectors: use and optimisation in cancer gene therapy. Expert Rev Mol Med. 2004;6(16):1–24.
98. Qiu J, Pintel DJ. The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template. Mol Cell Biol. 2002;22(11):3639–52.

99. Weindler FW, Heilbronn R. A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. J Virol. 1991;65(5):2476–83.
100. Qiu J, Nayak R, Tullis GE, Pintel DJ. Characterization of the transcription profile of adeno-associated virus type 5 reveals a number of unique features compared to previously characterized adeno-associated viruses. J Virol. 2002;76(24):12435–47.
101. Cossart YE, Field AM, Cant B, Widdows D. Parvovirus-like particles in human sera. Lancet. 1975;1(7898):72–3.
102. Candotti D, Etiz N, Parsyan A, Allain JP. Identification and characterization of persistent human erythrovirus infection in blood donor samples. J Virol. 2004;78(22):12169–78.
103. Anderson MJ, Lewis E, Kidd IM, Hall SM, Cohen BJ. An outbreak of erythema infectiosum associated with human parvovirus infection. J Hyg (Lond). 1984;93(1):85–93.
104. Woolf AD, Campion GV, Chishick A, et al. Clinical manifestations of human parvovirus B19 in adults. Arch Intern Med. 1989;149(5):1153–6.
105. Potter CG, Potter AC, Hatton CS, et al. Variation of erythroid and myeloid precursors in the marrow and peripheral blood of volunteer subjects infected with human parvovirus (B19). J Clin Invest. 1987;79(5):1486–92.
106. Jordan JA. Identification of human parvovirus B19 infection in idiopathic nonimmune hydrops fetalis. Am J Obstet Gynecol. 1996;174:37–42.
107. Ozawa K, Ayub J, Hao YS, Kurtzman G, Shimada T, Young N. Novel transcription map for the B19 (human) pathogenic parvovirus. J Virol. 1987;61(8):2395–406.
108. Lynch JP III, Fishbein M, Echavarria M. Adenovirus. Semin Respir Crit Care Med. 2011;32(4):494–511.
109. Berk AJ. Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. Oncogene. 2005;24(52):7673–85.
110. Akusjarvi G. Temporal regulation of adenovirus major late alternative RNA splicing. Front Biosci. 2008;13:5006–15.
111. Stephens C, Harlow E. Differential splicing yields novel adenovirus 5 E1A mRNAs that encode 30 kd and 35 kd proteins. EMBO J. 1987;6(7):2027–35.
112. Cuesta R, Xi Q, Schneider RJ. Adenovirus-specific translation by displacement of kinase Mnk1 from cap-initiation complex eIF4F. EMBO J. 2000;19(13):3465–74.
113. Akusjarvi G, Stevenin J. Remodelling of the host cell RNA splicing machinery during an adenovirus infection. Curr Top Microbiol Immunol. 2003;272:253–86.
114. Akusjarvi G, Persson H. Controls of RNA splicing and termination in the major late adenovirus transcription unit. Nature. 1981;292(5822):420–6.
115. Zheng ZM. Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. Int J Biol Sci. 2010;6(7):730–55.
116. Gardner SD, Field AM, Coleman DV, Hulme B. New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet. 1971;1(7712):1253–7.
117. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH. Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet. 1971;1(7712):1257–60.
118. Allander T, Andreasson K, Gupta S, et al. Identification of a third human polyomavirus. J Virol. 2007;81(8):4130–6.
119. Gaynor AM, Nissen MD, Whiley DM, et al. Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog. 2007;3(5):e64.
120. Feng H, Shuda M, Chang Y, Moore PS. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science. 2008;319(5866):1096–100.
121. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB. Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe. 2010;7(6):509–15.
122. van der ME JRW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC. Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromised patient. PLoS Pathog. 2010;6(7):e1001024.

123. Scuda N, Hofmann J, Calvignac-Spencer S, et al. A novel human polyomavirus closely related to the african green monkey-derived lymphotropic polyomavirus. J Virol. 2011;85(9):4586–90.
124. Sweet BH, Hilleman MR. The vacuolating virus, S.V. 40. Proc Soc Exp Biol Med. 1960;105:420–7.
125. Knowles WA, Pipkin P, Andrews N, et al. Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40. J Med Virol. 2003;71(1):115–23.
126. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R. Phylogenetics, evolution, and medical importance of polyomaviruses. Infect Genet Evol. 2009;9(5):784–99.
127. Shuda M, Feng H, Kwun HJ, et al. T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A. 2008;105(42):16272–7.
128. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS. Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. J Clin Invest. 2011;121(9):3623–34.
129. Houben R, Adam C, Baeurle A, et al. An intact retinoblastoma protein-binding site in Merkel cell polyomavirus large T antigen is required for promoting growth of Merkel cell carcinoma cells. Int J Cancer. 2012;130(4):847–56.
130. Zerrahn J, Knippschild U, Winkler T, Deppert W. Independent expression of the transforming amino-terminal domain of SV40 large I antigen from an alternatively spliced third SV40 early mRNA. EMBO J. 1993;12(12):4739–46.
131. Abend JR, Joseph AE, Das D, Campbell-Cecen DB, Imperiale MJ. A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. J Gen Virol. 2009;90:1238–45.
132. Trowbridge PW, Frisque RJ. Identification of three new JC virus proteins generated by alternative splicing of the early viral mRNA. J Neurovirol. 1995;1(2):195–206.
133. Zheng ZM, Baker CC. Papillomavirus genome structure, expression, and post-transcriptional regulation. Front Biosci. 2006;11:2286–302.
134. Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology. 2010;401(1):70–9.
135. Munoz N, Bosch FX, de Sanjosé S, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med. 2003;348(6):518–27.
136. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. Lancet. 2007;370(9590):890–907.
137. Chen HC, Schiffman M, Lin CY, et al. Persistence of type-specific human papillomavirus infection and increased long-term risk of cervical cancer. J Natl Cancer Inst. 2011;103(18):1387–96.
138. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell. 1990;63(6):1129–36.
139. Dyson N, Howley PM, Munger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science. 1989;243(4893):934–7.
140. Tang S, Tao M, McCoy JP Jr, Zheng ZM. The E7 oncoprotein is translated from spliced E6*I transcripts in high-risk human papillomavirus type 16- or type 18-positive cervical cancer cell lines via translation reinitiation. J Virol. 2006;80(9):4249–63.
141. Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM. Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. J Virol. 2011;85(16):8080–92.
142. Khan MJ, Castle PE, Lorincz AT, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst. 2005;97(14):1072–9.
143. Stoler MH, Wright TC Jr, Sharma A, Apple R, Gutekunst K, Wright TL. High-risk human papillomavirus testing in women with ASC-US cytology: results from the ATHENA HPV study. Am J Clin Pathol. 2011;135(3):468–75.

144. Getman D, Aiyer A, Dockter J, Giachetti C, Zhang F, Ginocchio CC. Efficiency of the APTIMA HPV Assay for detection of HPV RNA and DNA targets. J Clin Virol. 2009;45(Suppl 1):S49–54.
145. Ratnam S, Coutlee F, Fontaine D, et al. Aptima HPV E6/E7 mRNA test is as sensitive as Hybrid Capture 2 Assay but more specific at detecting cervical precancer and cancer. J Clin Microbiol. 2011;49(2):557–64.
146. Clad A, Reuschenbach M, Weinschenk J, Grote R, Rahmsdorf J, Freudenberg N. Performance of the Aptima high-risk human papillomavirus mRNA assay in a referral population in comparison with Hybrid Capture 2 and cytology. J Clin Microbiol. 2011;49(3):1071–6.
147. Molden T, Kraus I, Skomedal H, Nordstrom T, Karlsen F. PreTect HPV-Proofer: real-time detection and typing of E6/E7 mRNA from carcinogenic human papillomaviruses. J Virol Methods. 2007;142(1-2):204–12.
148. Benevolo M, Vocaturo A, Caraceni D, et al. Sensitivity, specificity, and clinical value of human papillomavirus (HPV) E6/E7 mRNA assay as a triage test for cervical cytology and HPV DNA test. J Clin Microbiol. 2011;49(7):2643–50.
149. Ratnam S, Coutlee F, Fontaine D, et al. Clinical performance of the PreTect HPV-Proofer E6/E7 mRNA assay in comparison with that of the Hybrid Capture 2 test for identification of women at risk of cervical cancer. J Clin Microbiol. 2010;48(8):2779–85.
150. Steitz J, Borah S, Cazalla D, et al. Noncoding RNPs of viral origin. Cold Spring Harb Perspect Biol. 2011;3(3):a005165.
151. Cullen BR. Herpesvirus microRNAs: Phenotypes and functions. Curr Opin Virol. 2011;1(3):211–5.
152. Plaisance-Bonstaff K, Renne R. Viral miRNAs. Methods Mol Biol. 2011;721:43–66.
153. Taylor GS, Blackbourn DJ. Infectious agents in human cancers: lessons in immunity and immunomodulation from gammaherpesviruses EBV and KSHV. Cancer Lett. 2011;305(2):263–78.
154. Zheng ZM. Split genes and their expression in Kaposi's sarcoma-associated herpesvirus. Rev Med Virol. 2003;13(3):173–84.
155. Arduino PG, Porter SR. Herpes Simplex Virus Type 1 infection: overview on relevant clinicopathological features. J Oral Pathol Med. 2008;37(2):107–21.
156. Rajcani J, Andrea V, Ingeborg R. Peculiarities of herpes simplex virus (HSV) transcription: an overview. Virus Genes. 2004;28(3):293–310.
157. Javier RT, Stevens JG, Dissette VB, Wagner EK. A herpes simplex virus transcript abundant in latently infected neurons is dispensable for establishment of the latent state. Virology. 1988;166(1):254–7.
158. Croen KD, Ostrove JM, Dragovic LJ, Smialek JE, Straus SE. Latent herpes simplex virus in human trigeminal ganglia. Detection of an immediate early gene “anti-sense” transcript by in situ hybridization. N Engl J Med. 1987;317(23):1427–32.
159. Farrell MJ, Dobson AT, Feldman LT. Herpes simplex virus latency-associated transcript is a stable intron. Proc Natl Acad Sci U S A. 1991;88(3):790–4.
160. Spivack JG, Woods GM, Fraser NW. Identification of a novel latency-specific splice donor signal within the herpes simplex virus type 1 2.0-kilobase latency-associated transcript (LAT): translation inhibition of LAT open reading frames by the intron within the 2.0-kilobase LAT. J Virol. 1991;65(12):6800–10.
161. Henderson G, Peng W, Jin L, et al. Regulation of caspase 8- and caspase 9-induced apoptosis by the herpes simplex virus type 1 latency-associated transcript. J Neurovirol. 2002;8(Suppl 2):103–11.
162. Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR. MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. Nature. 2008;454(7205):780–3.
163. Devi-Rao GB, Goodart SA, Hecht LM, Rochford R, Rice MK, Wagner EK. Relationship between polyadenylated and nonpolyadenylated herpes simplex virus type 1 latency-associated transcripts. J Virol. 1991;65(5):2179–90.

164. Chen J, Silverstein S. Herpes simplex viruses with mutations in the gene encoding ICP0 are defective in gene expression. J Virol. 1992;66(5):2916–27.
165. Perry LJ, Rixon FJ, Everett RD, Frame MC, McGeoch DJ. Characterization of the IE110 gene of herpes simplex virus type 1. J Gen Virol. 1986;67:2365–80.
166. Everett RD, Cross A, Orr A. A truncated form of herpes simplex virus type 1 immediate-early protein Vmw110 is expressed in a cell type dependent manner. Virology. 1993;197(2):751–6.
167. Spatz SJ, Nordby EC, Weber PC. Mutational analysis of ICP0R, a transrepressor protein created by alternative splicing of the ICP0 gene of herpes simplex virus type 1. J Virol. 1996;70(11):7360–70.
168. Tang S, Bertke AS, Patel A, Wang K, Cohen JI, Krause PR. An acutely and latently expressed herpes simplex virus 2 viral microRNA inhibits expression of ICP34.5, a viral neurovirulence factor. Proc Natl Acad Sci U S A. 2008;105(31):10931–6.
169. Berger JR, Houff S. Neurological complications of herpes simplex virus type 2 infection. Arch Neurol. 2008;65(5):596–600.
170. Desselberger U. Herpes simplex virus infection in pregnancy: diagnosis and significance. Intervirology. 1998;41(4-5):185–90.
171. Bankier AT, Beck S, Bohni R, et al. The DNA sequence of the human cytomegalovirus genome. DNA Seq. 1991;2(1):1–12.
172. Rawlinson WD, Barrell BG. Spliced transcripts of human cytomegalovirus. J Virol. 1993;67(9):5502–13.
173. Bresnahan WA, Shenk T. A subset of viral transcripts packaged within human cytomegalovirus particles. Science. 2000;288(5475):2373–6.
174. Nelson PN, Rawal BK, Boriskin YS, et al. A polymerase chain reaction to detect a spliced late transcript of human cytomegalovirus in the blood of bone marrow transplant recipients. J Virol Methods. 1996;56(2):139–48.
175. Gerna G, Revello MG, Percivalle E, Morini F. Comparison of different immunostaining techniques and monoclonal antibodies to the lower matrix phosphoprotein (pp65) for optimal quantitation of human cytomegalovirus antigenemia. J Clin Microbiol. 1992;30(5):1232–7.
176. Fernando S, Booth J, Boriskin Y, et al. Association of cytomegalovirus infection with post-transplantation cardiac rejection as studied using the polymerase chain reaction. J Med Virol. 1994;42(4):396–404.
177. Greijer AE, Adriaanse HM, Kahl M, et al. Quantitative competitive NASBA for measuring mRNA expression levels of the immediate early 1, late pp67, and immune evasion genes US3, US6 and US11 in cells infected with human cytomegalovirus. J Virol Methods. 2001;96(2):133–47.
178. Randhawa PS, Manez R, Frye B, Ehrlich GD. Circulating immediate-early mRNA in patients with cytomegalovirus infections after solid organ transplantation. J Infect Dis. 1994;170(5):1264–7.
179. Velzing J, Rothbarth PH, Kroes AC, Quint WG. Detection of cytomegalovirus mRNA and DNA encoding the immediate early gene in peripheral blood leukocytes from immunocompromised patients. J Med Virol. 1994;42(2):164–9.
180. Meyer T, Reischl U, Wolf H, Schuller C, Arndt R. Identification of active cytomegalovirus infection by analysis of immediate-early, early and late transcripts in peripheral blood cells of immunodeficient patients. Mol Cell Probes. 1994;8(4):261–71.
181. Boriskin YS, Fuller K, Powles RL, et al. Early detection of cytomegalovirus (CMV) infection in bone marrow transplant patients by reverse transcription-PCR for CMV spliced late gene UL21.5: a two site evaluation. J Clin Virol. 2002;24(1-2):13–23.
182. Andre E, Imbert-Marcille BM, Cantarovich D, Besse B, Ferre-Aubineau V, Billaudel S. Use of reverse transcription polymerase chain reaction with colorimetric plate hybridization to detect a cytomegalovirus late spliced mRNA in polymorphonuclear leukocytes from renal transplant patients. Diagn Microbiol Infect Dis. 1999;34(4):287–91.
183. Yao QY, Ogan P, Rowe M, Wood M, Rickinson AB. Epstein-Barr virus-infected B cells persist in the circulation of acyclovir-treated virus carriers. Int J Cancer. 1989;43(1):67–71.

184. Kutok JL, Wang F. Spectrum of Epstein-Barr virus-associated diseases. Annu Rev Pathol. 2006;1:375–404.
185. Qu L, Green M, Webber S, Reyes J, Ellis D, Rowe D. Epstein-Barr virus gene expression in the peripheral blood of transplant recipients with persistent circulating virus loads. J Infect Dis. 2000;182(4):1013–21.
186. Young LS, Arrand JR, Murray PG. EBV gene expression and regulation. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K, editors. Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge, MA: Cambridge University Press; 2007. Chapter 27.
187. Baer R, Bankier AT, Biggin MD, et al. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature. 1984;310(5974):207–11.
188. Swaminathan S. Noncoding RNAs produced by oncogenic human herpesviruses. J Cell Physiol. 2008;216(2):321–6.
189. Barth S, Meister G, Grasser FA. EBV-encoded miRNAs. Biochim Biophys Acta. 2011;1809(11-12):631–40.
190. Kennedy G, Sugden B. EBNA-1, a bifunctional transcriptional activator. Mol Cell Biol. 2003;23(19):6901–8.
191. Woisetschlaeger M, Yandava CN, Furmanski LA, Strominger JL, Speck SH. Promoter switching in Epstein-Barr virus during the initial stages of infection of B lymphocytes. Proc Natl Acad Sci U S A. 1990;87(5):1725–9.
192. Nonkwelo C, Ruf IK, Sample J. The Epstein-Barr virus EBNA-1 promoter Qp requires an initiator-like element. J Virol. 1997;71(1):354–61.
193. Nonkwelo C, Henson EB, Sample J. Characterization of the Epstein-Barr virus Fp promoter. Virology. 1995;206(1):183–95.
194. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L. Lytic cycle switches of oncogenic human gammaherpesviruses. Adv Cancer Res. 2007;97:81–109.
195. Manet E, Gruffat H, Trescol-Biemont MC, et al. Epstein-Barr virus bicistronic mRNAs generated by facultative splicing code for two transcriptional trans-activators. EMBO J. 1989;8(6):1819–26.
196. Furnari FB, Zacny V, Quinlivan EB, Kenney S, Pagano JS. RAZ, an Epstein-Barr virus transdominant repressor that modulates the viral reactivation mechanism. J Virol. 1994;68(3):1827–36.
197. Chang Y, Cesarman E, Pessin MS, et al. Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science. 1994;266(5192):1865–9.
198. Ganem D. KSHV infection and the pathogenesis of Kaposi's sarcoma. Annu Rev Pathol. 2006;1:273–96.
199. Ablashi DV, Chatlynne LG, Whitman JE Jr, Cesarman E. Spectrum of Kaposi's sarcoma-associated herpesvirus, or human herpesvirus 8, diseases. Clin Microbiol Rev. 2002;15(3):439–64.
200. Russo JJ, Bohenzky RA, Chien MC, et al. Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A. 1996;93(25):14862–7.
201. Verma SC, Lan K, Robertson E. Structure and function of latency-associated nuclear antigen. Curr Top Microbiol Immunol. 2007;312:101–36.
202. Dittmer D, Lagunoff M, Renne R, Staskus K, Haase A, Ganem D. A cluster of latently expressed genes in Kaposi's sarcoma-associated herpesvirus. J Virol. 1998;72(10):8309–15.
203. Sun R, Lin SF, Gradoville L, Yuan Y, Zhu F, Miller G. A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. Proc Natl Acad Sci U S A. 1998;95(18):10866–71.
204. Lukac DM, Kirshner JR, Ganem D. Transcriptional activation by the product of open reading frame 50 of Kaposi's sarcoma-associated herpesvirus is required for lytic viral reactivation in B cells. J Virol. 1999;73(11):9348–61.

205. Yamanegi K, Tang S, Zheng ZM. Kaposi's sarcoma-associated herpesvirus K8beta is derived from a spliced intermediate of K8 pre-mRNA and antagonizes K8alpha (K-bZIP) to induce p21 and p53 and blocks K8alpha-CDK2 interaction. J Virol. 2005;79(22):14207–21.
206. Majerciak V, Pripuzova N, McCoy JP, Gao SJ, Zheng ZM. Targeted disruption of Kaposi's sarcoma-associated herpesvirus ORF57 in the viral genome is detrimental for the expression of ORF59, K8alpha, and K8.1 and the production of infectious virus. J Virol. 2007;81(3):1062–71.
207. Cheung TK, Guan Y, Ng SS, et al. Generation of recombinant influenza A virus without M2 ion-channel protein by introduction of a point mutation at the 5′ end of the viral intron. J Gen Virol. 2005;86:1447–54.
208. Purcell DF, Martin MA. Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol. 1993;67(11):6365–78.
209. Yamano Y, Nagai M, Brennan M, et al. Correlation of human T-cell lymphotropic virus type 1 (HTLV-1) mRNA with proviral DNA load, virus-specific CD8(+) T cells, and disease severity in HTLV-1-associated myelopathy (HAM/TSP). Blood. 2002;99(1):88–94.
210. Preiss S, Littlejohn M, Angus P, et al. Defective hepatitis B virus DNA is not associated with disease status but is reduced by polymerase mutations associated with drug resistance. Hepatology. 2008;48(3):741–9.
211. Nguyen QT, Wong S, Heegaard ED, Brown KE. Identification and characterization of a second novel human erythrovirus variant, A6. Virology. 2002;301(2):374–80.
212. Gama-Carvalho M, Condado I, Carmo-Fonseca M. Regulation of adenovirus alternative RNA splicing correlates with a reorganization of splicing factors in the nucleus. Exp Cell Res. 2003;289(1):77–85.
213. White MK, Safak M, Khalili K. Regulation of gene expression in primate polyomaviruses. J Virol. 2009;83(21):10846–56.
214. Tanaka S, Minagawa H, Toh Y, Liu Y, Mori R. Analysis by RNA-PCR of latency and reactivation of herpes simplex virus in multiple neuronal tissues. J Gen Virol. 1994;75:2691–8.
215. Majerciak V, Yamanegi K, Allemand E, Kruhlak M, Krainer AR, Zheng ZM. Kaposi's sarcoma-associated herpesvirus ORF57 functions as a viral splicing factor and promotes expression of intron-containing viral lytic genes in spliceosome-mediated RNA splicing. J Virol. 2008;82(6):2792–801.
216. Tang S, Zheng ZM. Kaposi's sarcoma-associated herpesvirus K8 exon 3 contains three 5′ -splice sites and harbors a K8.1 transcription start site. J Biol Chem. 2002;277(17):14547–56.