Chapter 38
Detection of Viral RNA Splicing in Diagnostic Virology

Vladimir Majerciak and Zhi-Ming Zheng

Introduction

Diagnostic virology is to identify the etiologic cause of infection from patient's samples. In the past the diagnostic virology relied on three classical techniques to make a diagnosis of viral infection: (a) virus isolation by direct virus cultivation, (b) viral antigen detection, (c) indirect detection of virus-specific antibodies. While being important tools in the diagnostic virology today, these techniques are time-consuming and require specific tools such as cultivation media, cell or tissue cultures, antibodies, purified antigens. In the past decade the number of new molecular-based methods grew rapidly and gained more popularity in diagnostic labs. The core of these techniques constitutes of techniques based on nucleic acid detection by specific amplification, hybridization, and/or sequencing (reviewed in ref. [1]). The most nucleic acid-based diagnostic methods are simple, speed, sensitive and specific and thus meet the gold four-S-standard for their application in any diagnostic laboratories. The methods are simple and speed because only a specific primer pair and a PCR machine are needed in a lab setting and identification of a viral pathogen takes within few hours. They are sensitive and specific and require only a small amount of patients' materials to detect a specific nucleotide sequence region. In general, these techniques can be used to detect almost all types of viral pathogens and even to identify multiple viral pathogens or their variants at the same time. In this chapter we focus on detection of viral RNA splicing as a new tool for diagnostic virology.

V. Majerciak • Z.-M. Zheng (✉)
Tumor Virus RNA Biology Section, HIV and AIDS Malignancy Branch,
Center for Cancer Research, National Cancer Institute, National Institutes of Health,
10 Center Dr., Rm. 10/6 N106, Bethesda, MD 20892-1868, USA
e-mail: zhengt@exchange.nih.gov

Principle of RNA Splicing

Definition of RNA Splicing

RNA splicing was discovered more than 30 years ago by Berget [2] and Chow [3] by mapping adenovirus transcription and identification of intervening sequences (introns) in type 2 adenovirus primary transcripts. Subsequently, RNA splicing is recognized as an essential nuclear event for mammalian gene expression and for virus multiplication of almost all DNA viruses and some RNA viruses. The majority of mammalian genes consist of multiple segments called exons which 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 non-coding primary transcripts. RNA splicing is a posttranscriptional event; however, recent studies showed that it often occurs cotranscriptionally [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. 38.1a). These *cis*-elements allow cellular splicing machinery to recognize and remove the intron from pre-mRNA. Most of mammalian introns start with GU dinucleotide on its 5′ end and a AG dinucleotide on its 3′ end ("GU-AG" introns). The GU-AG pairs are conserved sequences defining 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 AU-AC introns [6, 7]. The presence of splice sites is not sufficient for intron definition. All introns must contain an additional element called "branch point" located 20–50 nts upstream from 3′ splice site with consensus sequence CU(A/C)A(C/U) where A is a most conserved base. Sequence between the branch point and acceptor site is a run of 15–40 pyrimidines (mostly U) and is referred as a "polypyrimidine track."

Fig. 38.1 (continued) 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 1st transesterification reaction between the branch point and the 5′ donor site, resulting in 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 joints with the 5′ end of exon 2 via the 2nd transesterification reaction to form a mature mRNA. Intron is removed in the form of lariat structure and quickly degraded
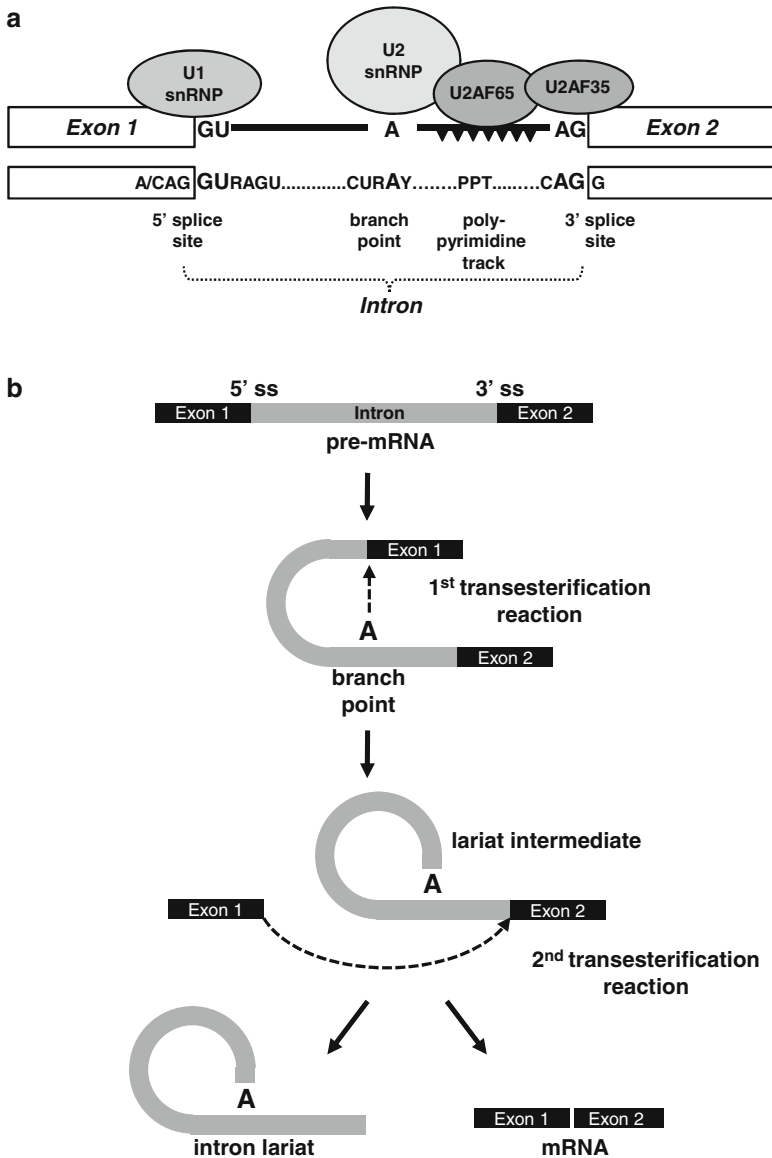
Fig. 38.1 Pre-mRNA structure and splicing reactions. (**a**) Structure of a pre-mRNA containing an intron (*solid line*) and two exons (*empty boxes*). An intron is defined by several specific sequence motifs which allow intron recognition by cellular splicing machinery. A 5′ splice site or donor site 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. 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 (*grey*) between

RNA splicing is catalyzed by cellular splicing machinery consisting of several components: small nuclear ribonucleoproteins (snRNPs, U1, U2, U4, U5, and U6) and splicing factors. During initial step snRNP U1 and U2 recognized intron sequences at 5′ splice site and branch point via complementary base-pairing, along with U2 accessory proteins U2AF65 and U2AF35 association with polypyrimidine track and 3′ splice site, respectively (Fig. 38.1a). Intron recognition is a signal for formation of a large protein complex called "spliceosome" where intron removal takes place [8, 9] by two transesterification reactions (Fig. 38.1b). First, the primary transcript is cleaved at the intron 5′ end (5′ splice site) to leave the upstream exon free and 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 bound 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 regulation of RNA splicing. As we described above, the level of conservation of the sequences at splice sites and branch point will affect the strength of binding of core splicing factors and 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 differentially expressed in a specified tissues and/or 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 type of splicing factors, positions of the binding sites and overall spliceosome composition [11, 12]. Current studies showed that beside 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 cases showing that an RNA splice site may be not selected constitutively, but rather skipped, during RNA splicing. Consequently, this alternative RNA splicing leads to production of RNA isoforms with different exon composition and sizes often resulting in production of different protein isoforms. In general, there are four major classes of alternative RNA splicing (Fig. 38.2), 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 because of the presence of weak or suboptimal splice signals in the RNA or lack of a particular splicing factor in that cell or tissue. It has been noticed that usage of weak
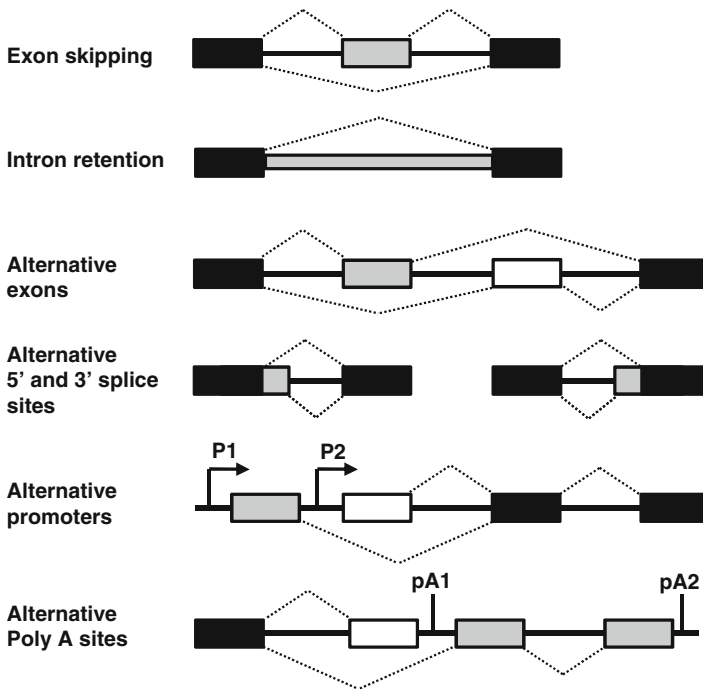
Fig. 38.2 Alternative RNA splicing. Alternative RNA splicing allows production of multiple splicing isoforms from a single pre-mRNA species. Constitutive exons (*black boxes*) are included in all splicing products while alternative exons (*grey* and *white boxes*) are either included or excluded in various isoforms of mature mRNAs. Major forms of alternative RNA splicing includes: 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

splice sites is highly dependent on auxiliary splicing factors binding to the regulatory *cis*-elements. Since the expression of these factors is variable from cells to cells and from tissues to tissues, alternative RNA splicing is often associated with a specific cell type, tissue, stage of cell differentiation [16] and results in various isoforms of transcripts from one gene [17]. Current analysis also revealed that 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 proteome in higher organisms other than total number of the mapped genes.

Molecular Methods for Detection of RNA Splicing

A 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 not only is 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 to detect RNA splicing. First, RNA molecules isolated from samples are separated based on 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. 38.3a). Constitutive exon-based probes would detect all spliced RNA isoforms and the remaining, unspliced pre-mRNA and 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 could be used to specifically detect a spliced product. In addition, a specific probe from an alternative exon or intron could be also designed for detection of individual splicing isoforms derived from exon/intron inclusion.

RNase Protection Assay

RNase protection assay (RPA) requires most often ^{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 a partial intron region which will be digested from the probe due to lack of the intron sequence in the spliced mRNA. As a result, the protected probe by the corresponding exon regions of the detecting mRNA is shorter and will run faster in the gel (Fig. 38.3b). In general, RPA is more sensitive than northern Northern blot in detection of RNA splicing.

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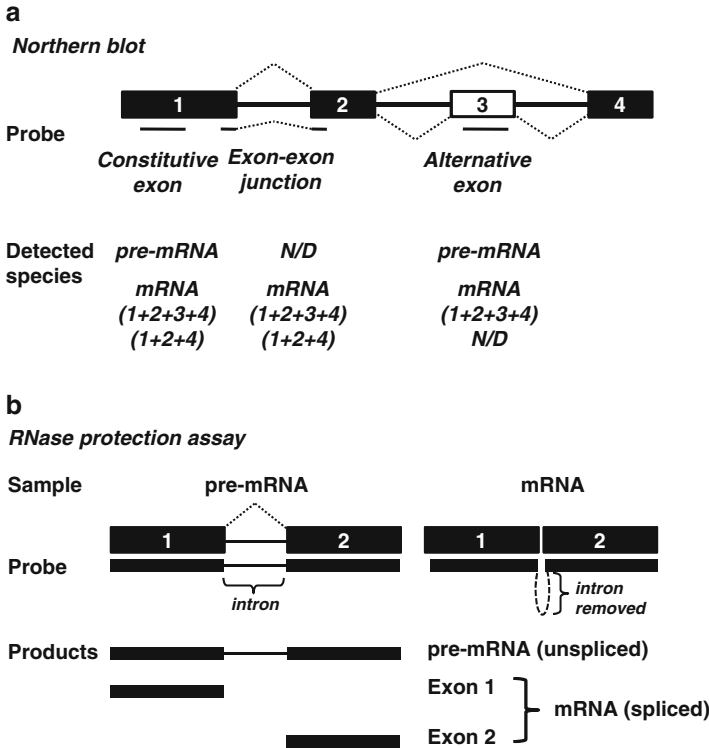
a

Northern blot

Probe

Constitutive   Exon-exon   Alternative
exon   junction   exon

Detected species	*pre-mRNA*	*N/D*	*pre-mRNA*
	mRNA	*mRNA*	*mRNA*
	(1+2+3+4)	*(1+2+3+4)*	*(1+2+3+4)*
	(1+2+4)	*(1+2+4)*	*N/D*

b

RNase protection assay

Sample   pre-mRNA   mRNA

Probe

intron   *intron removed*

Products   **pre-mRNA (unspliced)**

Exon 1
                                          mRNA (spliced)
Exon 2

Fig. 38.3 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, dashed loop) resulting in production of smaller protected products corresponding to each exon

RT-PCR

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 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. A 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 predicted size may represent an unspliced pre-mRNA or the contaminating genomic DNA. The later 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. 38.4a). Another approach is to specifically amplify a spliced product by using an exon junction primer because the sequence at exon–exon junction occurs neither in pre-mRNA nor in genomic DNA. Similarly, a primer based on an alternative exon would amplify only transcript with the inclusion of that exon (Fig. 38.4b). After amplification the size and amount of RT-PCR products are analyzed by gel electrophoresis. Because of nonlinear nature of PCR amplification the classical PCR only provides semi-quantitative data on the abundance of various spliced RNA isoforms.

Introduction of real-time quantitative RT-PCR (qRT-PCR) with a broad (10^7) dynamic range has significantly improved the sensitivity of RT-PCR. Because of high sensitivity real-time RT-PCR enables to detect and amplify RNA directly from 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 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 different chemistries are described in Fig. 38.5. As of today, SYBR green and *Taq*Man™ probes represent the most common 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 fluorescence resonance energy transfer or Förster resonance energy transfer (FRET, also known as resonance energy transfer [RET] or electronic energy transfer [EET]) to generate fluorescent signal. The FRET couples donor fluorescent dye (fluorophore) with nonfluorescent quenching moiety (quencher). When 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 which are labeled with a reporter fluorophore at the 5′ end and a quencher at the 3′ end and thus in close proximity. Each probe is complementary to a region in the middle of the detecting target between two primers during PCR reaction. When *Taq* polymerase extends the primer to synthesize the nascent strand, the 5′–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 entire length of RT-PCR reaction enabling to calculation of the amount of PCR product after each amplification cycle. Similarly to *Taq*Man™ the Molecular Beacon and Scorpions™
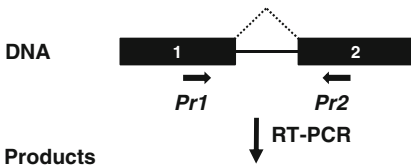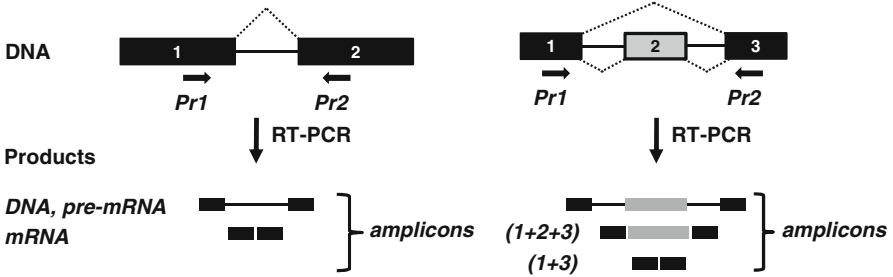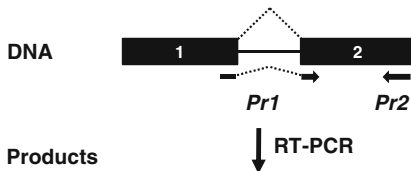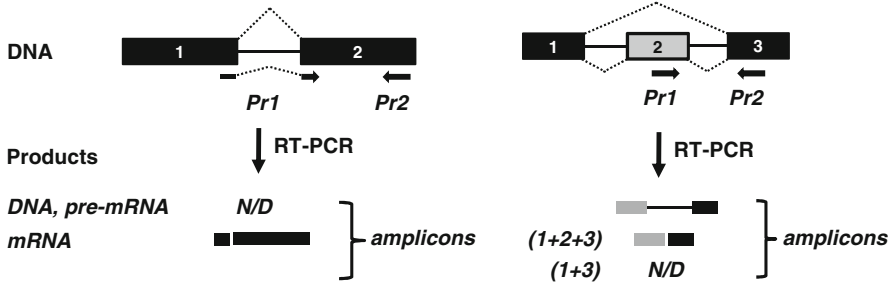
Fig. 38.4 Detection of spliced RNA products by RT-PCR. Intron removal from a pre-mRNA could be detected by RT-PCR based on the size of a product (**a**) or in a selective amplification (**a** and **b**). (**a**) A pair of primers (Pr1 and Pr2) used for amplification are derived from two constitutive 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 primes is derived from exon 1 and exon 3 spanning over alternative exon 2 (*grey 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 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

use probes to detect specific PCR product. However, instead of probe degradation the signal is generated by physical separation of fluorophore and quenching moieties after hybridization of specific probe to PCR product during amplification (Fig. 38.5).

Each chemistry has its own advantages and disadvantages which need to be considered during experimental design. SYBR green represents a simple, easy-to-use, and the most economical real-time RT-PCR method. Disadvantage of SYBR green
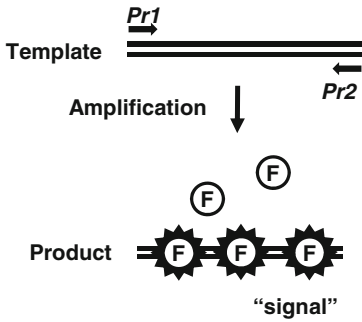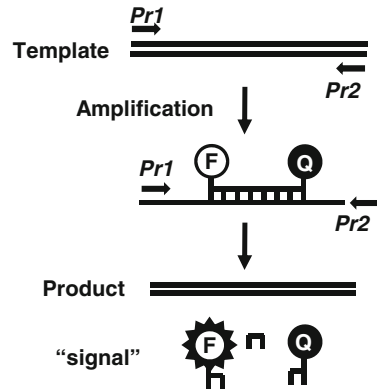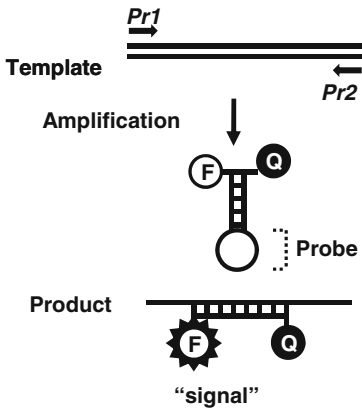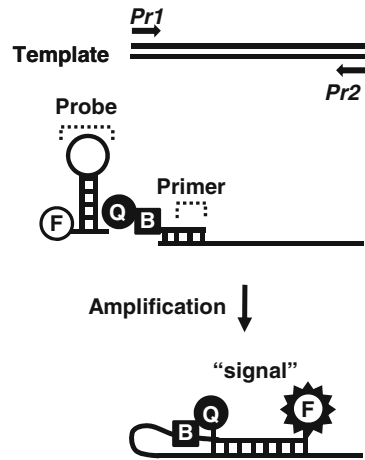
Fig. 38.5 Major chemistries of real-time PCR. PCR product detection by four major commercially available chemistries in real-time PCR: SYBR green (*I*), *Taq*Man™ (*II*), Molecular Beacon (*III*), and Scorpions™ probes (*IV*). The mechanism of each chemistry is described in details in other chapter. *Double lines* represent a double-stranded DNA template generated by RT-PCR while a *single line* is correspondent to single-strand 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." Scorpions probes contain amplification stop sequences named "blocker" (B in *black box*)

is its binding nonspecifically to any DNA including primer dimers and nonspecific PCR products and 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 nonspecific products. Disadvantage of these 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 spectrum allows multiplexing and simultaneous detection of several products and reduces the cost and labor.

The usage of real-time RT-PCR for splicing detection requires special considerations. Since real-time RT-PRC techniques omit electrophoretic separation, the spliced product could not be distinguished based on size. Therefore it is important that only desired product is amplified. In this case, the usage of SYBR green chemistry is the most challenging because of the lack of specificity. Probe-based methods provide higher specificity due to probe hybridization to selected sequences which are not present in nonspecific products. To detect only desired spliced product the probe and primer should be complementary to a specific exon–exon junction or alternatively spliced region. Several other factors must be considered when using RT-PCR-based techniques in diagnostics including RNA sample quality and preparation, *Taq* polymerase inactivating contaminants in clinical samples, and 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 in several thousands, spotted on very small area in 2D format on solid surface (glass or plastic). Probes represent DNA oligos of various length and chemistry. Each probe has specific DNA sequence allowing detection of corresponding DNA with complementary sequence. Currently, there are two major technologies of DNA arrays manufacturing of microarray: (a) direct synthesis of probes on the array or (b) printing arrays from 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 microarray readers (Fig. 38.6a). The intensity of fluorescent signal corresponding to the number of bound molecules allows determination of the level of RNA in the original sample by mathematical algorithm.

There are two different approaches in probe design to study of RNA splicing using DNA microarrays: tiling and exon arrays [22] (Fig. 38.6b). In tiling arrays the set of overlapping probes cover the full-length of nascent primary transcript including exons and introns. The analysis of florescence for each probe allows to identify exons and introns based on the difference in signal intensity (Fig. 38.6b*i*). The advantage of tiling arrays is their ability to identify known as well as new splice events. Therefore, the tilling arrays are often used as discovery tools. The disadvantages are
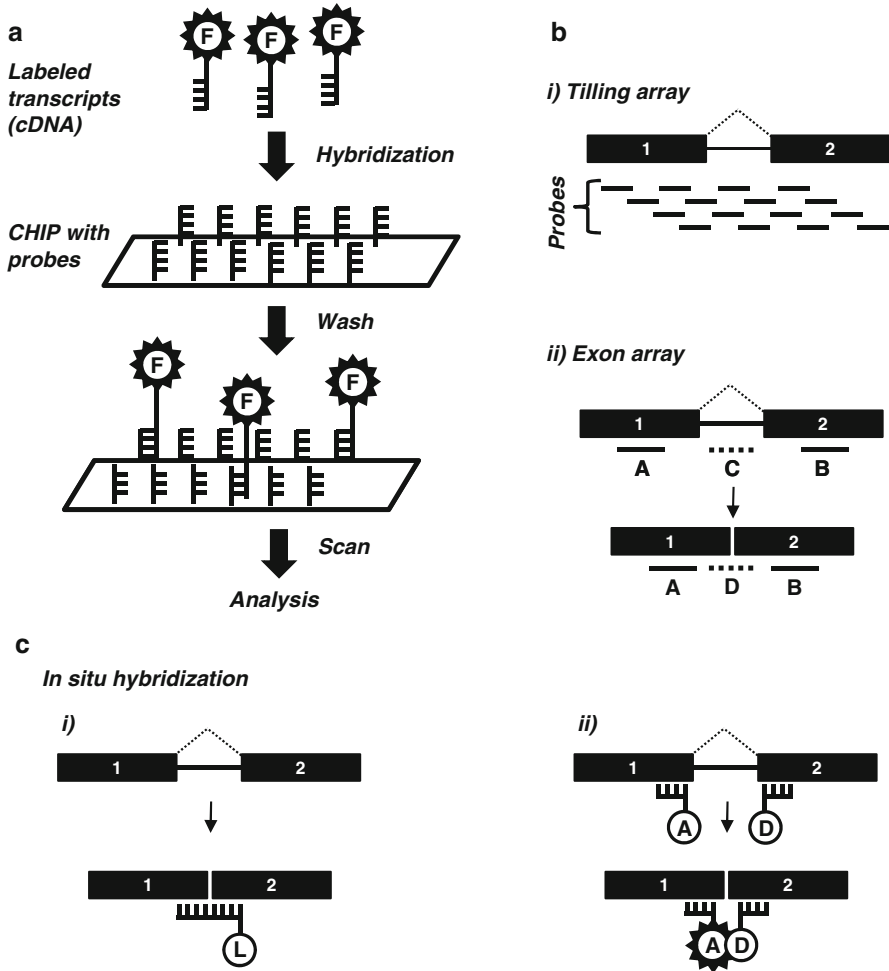
Fig. 38.6 Splicing microarrays and in situ hybridization. (**a**) A work flow for microarray assay. First, sample RNAs are converted into cDNAs simultaneously labeled with specific fluorophore (F). The labeled cDNAs are hybridized with oligonucleotide probes attached to a solid surface. The unbound cDNAs 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 lies*). 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. (ii) Detection of RNA splicing by cohybridization of two probes labeled with acceptor (A in *cirle*) 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 close 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]

requirement of large amount of probes and are 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. 38.6b*ii*). The fluorescence intensity is detected for each probe and mathematical model is applied to determine the occurrence of splicing event. The advantages of exon arrays are: the requirement of smaller number of probes with simpler data analysis. However, the exon arrays detect only known or predicted splicing variants. Due to their large capacities the exon arrays could be designed to detect splicing in multiple viral pathogens.

In Situ Hybridization

Tissue sections historically represent an important tool in diagnostic of pathological changes during viral infection and detection of viral pathogens on cellular level. There are two major types of tissue sections: frozen and formalin-fixed, paraffin-embedded (FFPE). Both are routinely used for detection of viral antigens by various types of staining, but their use in detection of nucleic acids including spliced transcripts is still rather rare. 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 cells of interest by laser capture microdissection to bring additional level of specificity [23, 24]. However, detection of nucleic acids by in situ hybridization (ISH) directly on tissue sections can provide additional benefits about gene expression linked with spatial distribution of specific RNA transcripts at cellular and often even at 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 enzymes-labeled antibodies (CISH) [26]. Labeling probes specific for different transcripts with different fluorophores (FISH) allow 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 usage of DNA probes which suffer from low affinity to complementary RNA targets and 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 came from by introduction of locked nucleic acids (LNA) and peptide nucleic acids (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 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. 38.6c*i*). Another approach in detection of spliced transcripts by IHS represents the methods by cohybridization of two probes labeled with donor and acceptor fluorophore and 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 of

two probes carries a fluorophore acceptor while the second probe is labeled by fluorophore donor. When probes bind to genomic DNA or unspliced nascent transcript their binding sites are separated by intron regions resulting in distance between donor and acceptor too big for two fluorophores to engage in FRET. However, intron removal by splicing brings probe binding sites to proximity close enough for FRET to occur resulting in generation of measurable fluorescence [31] (Fig. 38.6c*ii*). This results in high specificity and low background. Using a set of probes with different fluorophores allows to detect multiple spliced transcripts or various spliced isoforms of the transcript. In summary, ISH hybridization methods provide us with a necessary tool to investigate the distribution not only of protein-encoding transcripts but also of rapidly growing number of virus-encoded noncoding RNAs, of which their role in viral pathogenesis remains often 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]. It bases on generation of a large amount of short sequences in parallel sequenced 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 detecting sequence. Currently there are several platforms including 454 Life Sciences (*Roche*), Genome Analyzer (*Illumina*), ABI Solid Sequencing (*ABI*), and many others. Each platform uses different technology to generate the data, but all provide the same information.

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 is costing and requires sophisticated data analysis, which makes it less suitable for clinical diagnostics. However, *RNA-seq* does not require any prior knowledge of detecting sequence composition and therefore allows to detect 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 of any type of cells or tissues.

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 multiplication including RNA splicing. By using constitutive and/or alternative RNA splicing, most of DNA viruses and some of RNA viruses increase complexity of their proteome without 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 would indicate virus infection, the result does not provide information about the stage and dynamic of virus infection. In many cases the progress of viral replication could be assumed from changes in viral load, but this approach requires multiple sampling in the course of 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 active viral infection, providing important information about the status of infection without requiring multiple sampling. The production of viral transcripts and their RNA splicing 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 especially important for the early viral diagnosis of the infection 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 could provide essential information earlier enough for initiation of antiviral therapy. A rapid shut-off 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 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 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 RNA splicing detection results is straight forward without worrying carryover DNA contamination as a spliced RNA is smaller than its corresponding DNA template. As we describe above, there are many techniques available today for RNA splicing assay. These techniques are not only easy to set up with a low cost comparing to virus isolations and immunological methods, but could be quickly applied to detect new emerging viruses which cultivation of the virus is impossible and/or no immunological method is available. This is particularly true in combination of RNA-*seq* which can rapidly provide sequence information about a viral transcriptome and RNA splicing of 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 *Orthomyxoviridae* family. Influenza viruses are enveloped RNA viruses with a segmented, single-stranded RNA genome of negative polarity (ssRNA−). 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 viral RNAs are produced by viral RNA-depended 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 served as a template for viral genome replication (see review ref. [38]).

RNA splicing in influenza viruses was first detected in an RNA transcript from the smallest segment 8 in influenza A and B and their corresponding segment 7 in influenza C. This transcript encodes two nonstructural viral proteins: larger NS1 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 leads influenza A NS1 and NS2 proteins sharing the same AUG start codon and first nine amino acid residues. Translation of NS2 protein continues in +1 frame after RNA splicing and this leads the C-terminal NS2 partially overlapping the NS1 by 70 amino acid residues [40] (Fig. 38.7a). Similar splicing event for production of NS1 and NS2 proteins has been detected from infections of influenza B [41] and influenza C [42] viruses.

Influenza A segment 7 which encodes M1 and M2 proteins produces three RNA species by alternative RNA splicing. The unspliced RNA which is collinear with the genome encodes M1 nucleoprotein with 252 amino acid residues and 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. 38.7b). 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. M2 protein bears ion channel activity and shares nine amino acid residues with the M1 N-terminus. The mRNA3 contains a short open reading frame in its exon 2 with potential to encode a short peptide of nine amino acid residues. However, the expression of this peptide was never experimentally confirmed. The role of this transcript during virus replication remains unknown.

While M2 and mRNA3 transcripts are detectable in cells infected with all influenza A viruses, some strains, like A/WSN/33, produce additional spliced transcript named mRNA4 [44]. mRNA4 is generated by usage of additional 5′ splice site at position nt 146 and share the same 3′ splice site with M2 and mRNA3 at position nt 740 (Fig. 38.7b). mRNA4 has 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 6,000 influenza strains revealed that about 20 influenza A strains has conserved mRNA4 splice site [45]. Sequence information of all influenza viruses could be found at http://www.flu.lanl.gov/.

The primary RNA transcript of segment 7 in influenza virus B does not undergo alternative splicing to produce M2 protein as in case of influenza A. RNA splicing in M transcript takes place with segment 6 of influenza C [46]. Two transcripts generated from segment 6 in the infected cells are the full-length primary and single spliced transcript by removal of an intron located at the 3′ end of the primary transcript. The primary transcript contains a 374-aa ORF (P42), but the spliced message contains a shorter ORF encoding 242-aa residues (CM1) due to generation of STOP codon after RNA splicing from nt 751 5′ splice site to nt 982 3′ splice site
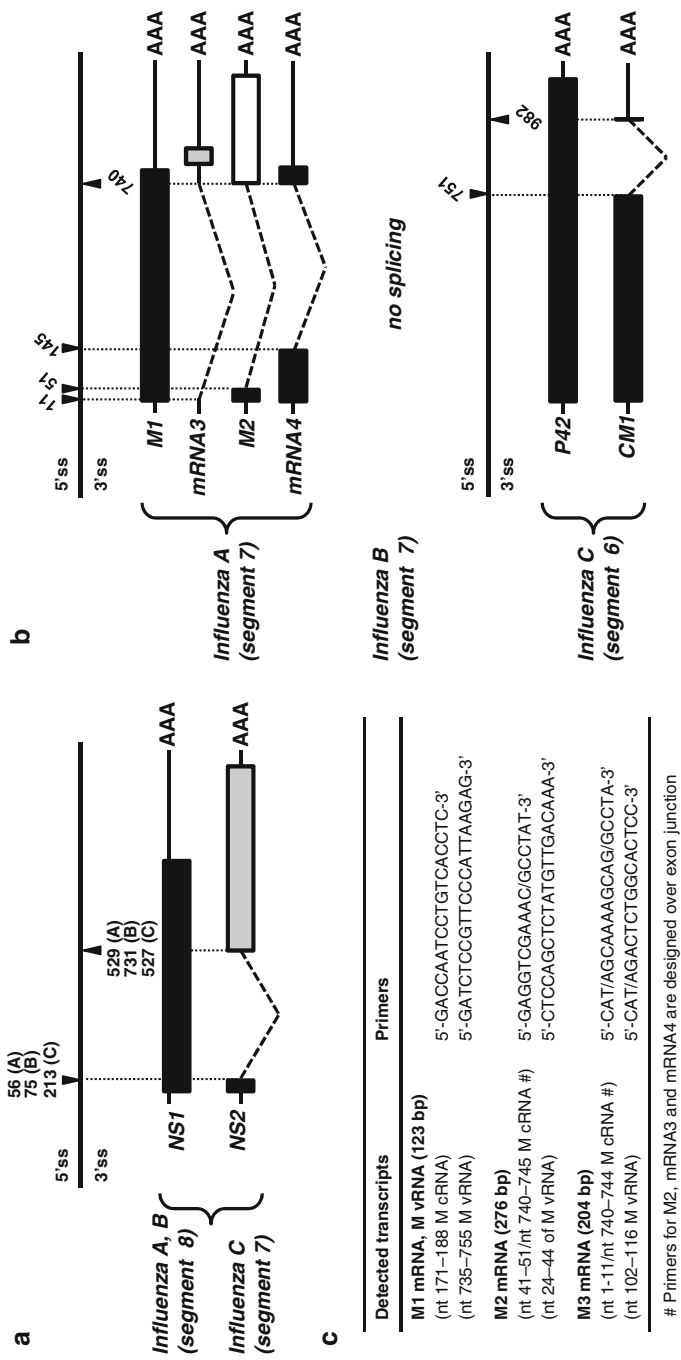
Fig. 38.7 Diagrams of RNA splicing of influenza virus NS and M transcripts. *Thin lines* represent noncoding sequences. *Dashed lines* indicate splicing directions. Transcripts are polyadenylated (AAA) at the 3′ end. *Black boxes* mark ORF in primary transcripts. *Small black boxes* and *white boxes* 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 are 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]. cRNA, complementary RNA; vRNA, virion RNA

(Fig. 38.7b). 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 highly variable in each species (M). Thus, combination of NS1 and M RNA splicing assays would reveal not only active influenza virus infection, but also specify the infection with a specific influenza virus species.

Human Retroviruses

HIV-1 and -2

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

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

In contrast to simple retroviruses, HIV genome has high coding capacity. Besides encoding viral structural and replication proteins (*gag*, *pol*, *env*) HIV expresses a large numbers of accessory proteins. This is achieved by production of over 40 RNA isoforms derived from single RNA transcript by extensive alternative RNA splicing [49] (Fig. 38.8a).

Three groups of HIV transcripts could be observed by size in Northern blot analysis. The first group represents an unspliced 9-kb transcript which serves a template for

Fig. 38.8 (continued) encodes gag and pol proteins. Alternatively spliced HIV-1 transcripts in size of ~2-kb and ~4-kb are grouped along with their protein coding potentials. *Solid boxes* represent exons and *dotted lines* are introns. A *grey 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 Accs. 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]
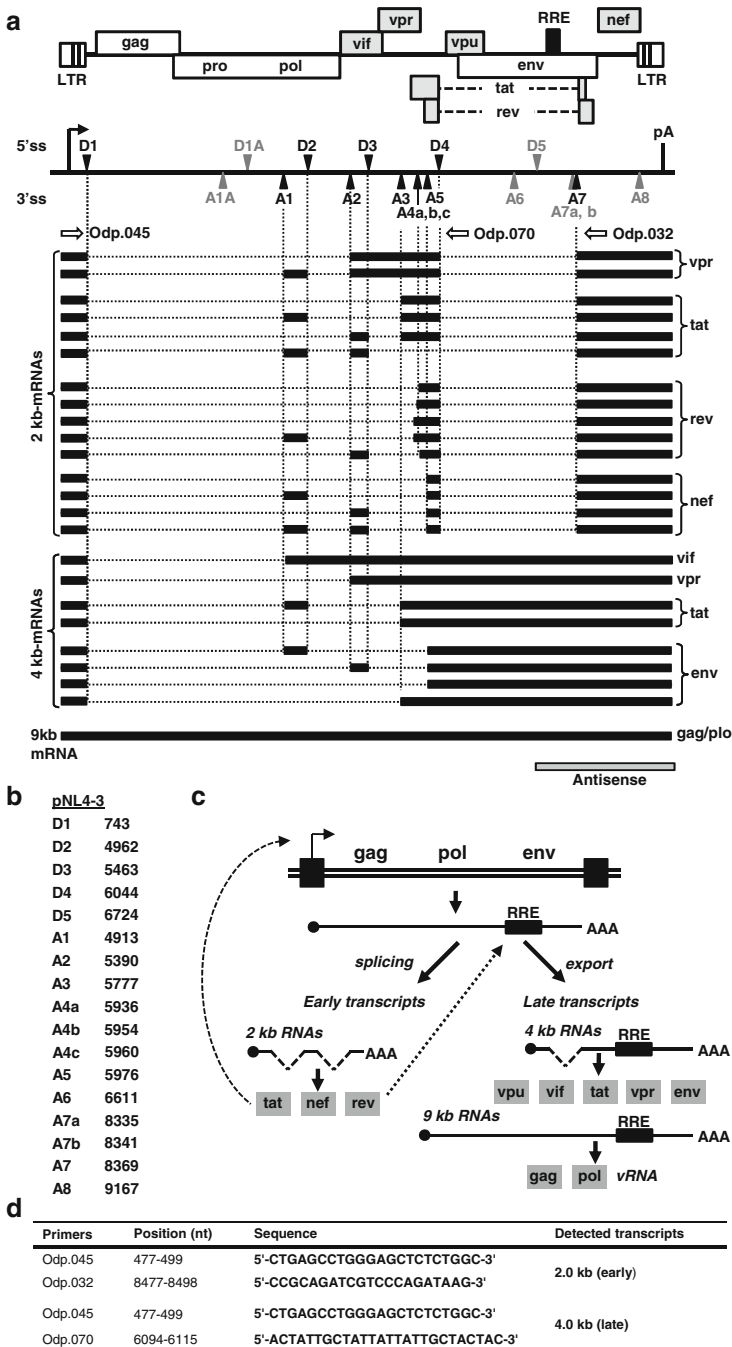
Fig. 38.8 RNA splicing of HIV-1 transcripts. (**a**) Schematic diagram of HIV-1 genome with structural and replicatory protein ORFs marked with *empty boxes* and accessory protein ORFs with *grey boxes*. *LTR* long terminal repeats. Shown below the diagram are positions of 5′ splice sites (D1–D4) and 3′ splice sites (A1–A7) identified in a 9-kb full-length primary transcript which

expression of *gag* and *gag/pol* as well as a genomic RNA for newly formed virions. The second group represents single spliced RNA transcripts in size of ~4-kb which encode *env*, *vif*, *vpr*, and *vpu* proteins. The third group in size of ~2-kb consists of multiple spliced RNA transcripts to encode accessory proteins *tat*, *rev*, *nef*, and *vpr*. During virus infection, HIV generates such a wide variety of RNA transcripts by usage of at least five alternative 5′ splice sites and 8–9 alternative 3′ splice sites [50, 51] (Fig. 38.8b). In addition, several antisense transcripts initiating from several 3′ long terminal repeats (3′ LTR) were also detected in HIV-1-infected cells [52].

Recent studies showed that HIV RNA splicing is largely regulated by viral RNA *cis*-elements and cellular splicing factors and is orchestrated for completion of HIV life cycle in the course of virus infection. Multiple spliced transcripts of 2-kb family are expressed in the early stage of virus infection to express *tat*, *rev* and *nef*. This group of spliced RNAs is produced by using 3′ splice site A3–A5 located in the central part of the viral genome with A3 site for expression of *tat*, A4a–c for *rev*, and A5 for *nef* proteins. During late stage of HIV infection, nuclear import and accumulation of *tat* together with *rev* protein allows the *rev* protein bind to a *rev-responsive element* (RRE) in partially spliced 4-kb and unspliced 9-kb RNA transcripts located in *tat/rev* intron between D4 and A7 splice sites to mediate the later transcripts export into the cytoplasm for translation [53] (Fig. 38.8c). Sites A1A and D1A are involved in pre-mRNA stability [54]. Strains from IIIB family of HIV viruses use additional A6 and D5 to generate small exon in *env* region and the transcripts containing this exon express 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 strong and highly active 5′ splice sites. In addition, numerous positive and negative splicing regulatory *cis*-elements identified in HIV RNA genome bind various cellular splicing factors and affect the selection of individual 3′ splice site (see review ref. [49]). Comparison of nucleotide sequences between various clades of HIV-1 has shown a high level of conservation of splice sites among different clades of HIV-1 strains (except D4a, b, c).

HTVL-1 and -2

Human T-cell leukemia virus type 1 (HTLV-1) and type 2 (HTLV-2) were first two retroviruses discovered in human [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-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 -2 are two closely related complex retroviruses sharing about 70 % of nucleotide sequences. Their genome organization and replication is similar to HIV. Besides essential genes (*gag/pol/env*) expressed from the full-length or single spliced RNA, HTLVs also encode a number of accessory proteins from a pX region
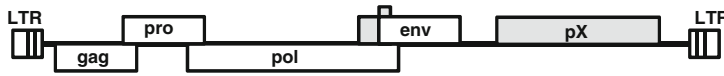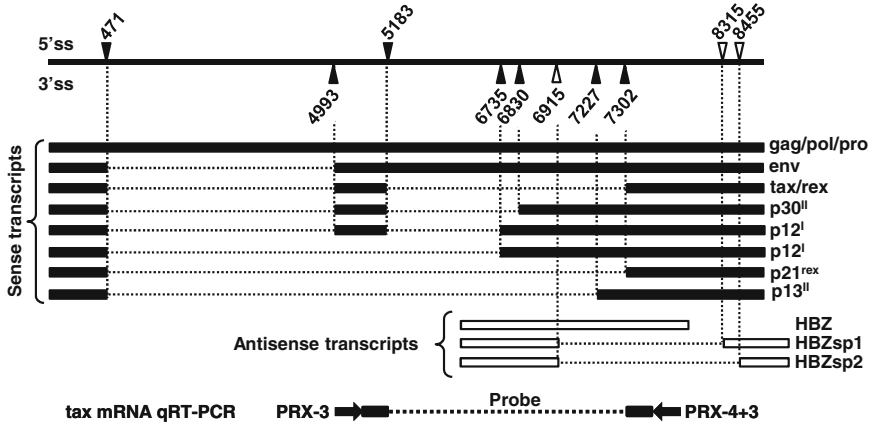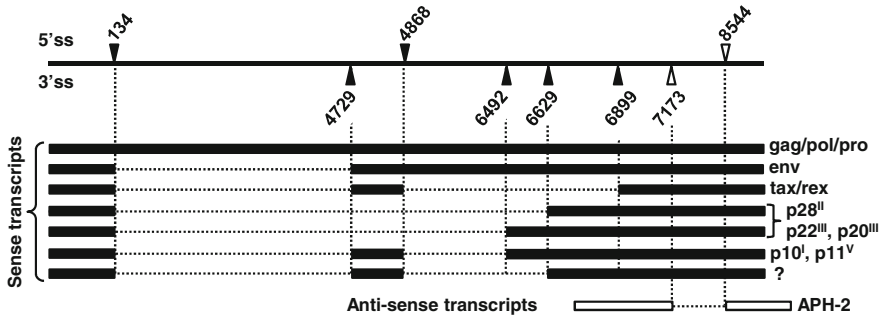
Fig. 38.9 RNA splicing of HTLV-1 & -2 transcripts. (**a**) Genomic organization of HTLV-1 and -2 with structural and replication genes (*white boxes*) on the 5′ end and regulatory gene pX (*grey boxes*) on the 3′ end of the genome flanked with long terminal repeats (LTR). (**b**) Viral transcripts generated 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 Accs. No. L03562.2)

located in the 3′ end of virus genome (Fig. 38.9a). 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. 38.9b) (reviewed in ref. [62]). Recently transcripts antisense to pX region have been discovered in the 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 HTLVs RNA splicing and the roles of cellular splicing factors in HTVL RNA splicing are poorly understood.

RNA Splicing in DNA Viruses

Circoviruses

Human circovirus TT virus (TTV) was originally discovered in the serum of a patient with posttransfusion non-A-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 general population appears highly variable from about 2 to 90 % of the incidence of TTV infection. These large variables among the reported studies are most likely attributable to primer selection and PCR performance [66]. TTV isolates have considerable diversity (about 30 %) which could be clustered in several genotypes groups without any particular geographical distribution, indicating that TTV represents a rather ubiquitous virus. Virus replication, route of TTV infection, and association with pathological manifestations remain unclear.

TTV is small nonenveloped virus of icosahedral architecture and contains a circular single-stranded DNA genome in size 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 supportive tissue culture system. TTV genome consists of a GC-rich noncoding region and a protein coding region with two overlapping ORFs (Fig. 38.10a). Three species of RNAs in sizes of 2.8-, 1.2-, and 1.0-kb were 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. 38.10b). Each RNA transcript could be translated into two different proteins by using two alternative start codons [71]. The presence of other ORFs (ORF3 and 4) in TTV genome was predicted, but has not confirmed yet.

The role of TTV RNA splicing and its regulation in virus infection and virus production remained largely unknown. Another human circovirus TTV-like mini virus (TLMV) has recently been identified in human sera [72]. TLMV shares same genetic organization with TTV and other circoviruses, but its genome is only about 2.9-kb in size.
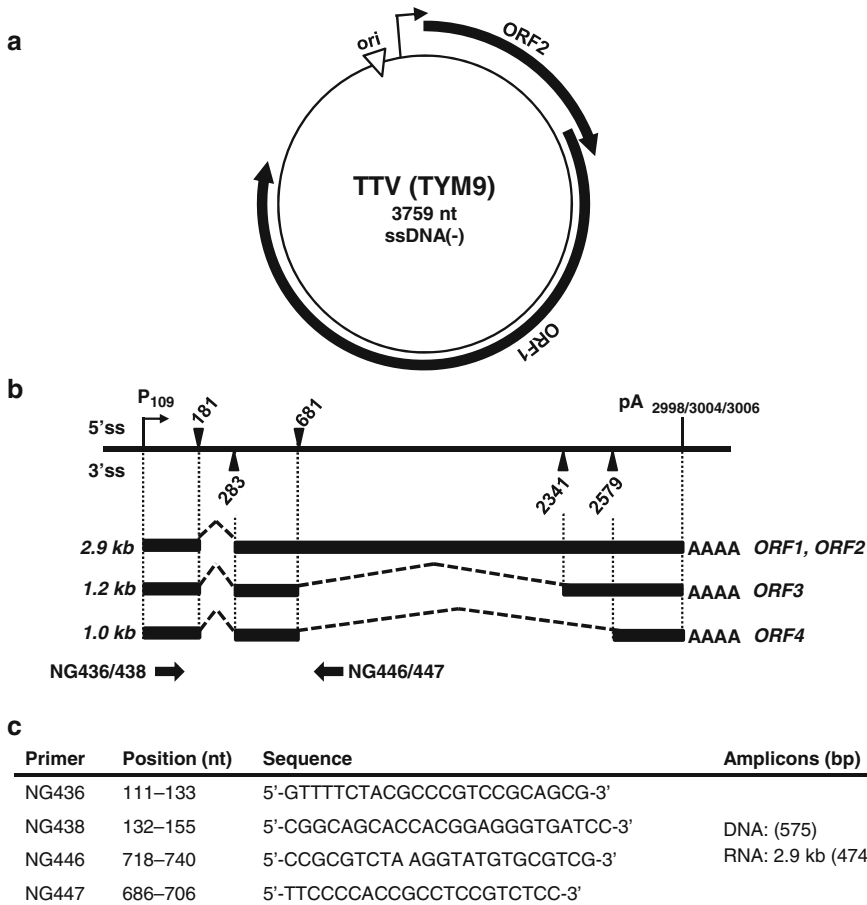
Fig. 38.10 Human circovirus TTV. (**a**) Single-stranded genome of TTV virus with origin of replication (ori), promoter (*arrow*) and two identified overlapping ORFs. (**b**) Transcripts identified in cells infected with TTV virus with nucleotide positions of two 5′ splice sites (*above the line*) and 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**). Primers positions in (**c**) are based on Torque teno virus TYM9 strain (GenBank Accs. No. AB050448.1)

Hepadnaviruses

Hepatitis B virus (HBV) chronic infection leads to development of liver cirrhosis and hepatocellular carcinoma [73]. Despite an effective HBV vaccine 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 two

billions of people worldwide, with chronic infection affecting 350 millions of people and causing death approximately one million people every year. While the vaccine can prevent HBV infection, there is no cure for already infected individuals.

HBV is a noncytopathic, 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-protein (Fig. 38.11a). In infected hepatocytes viral DNA is converted to cccDNA (covalently closed circular DNA) and serves as a template for expression of viral pregenomic (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. 38.11b). The core protein C and polymerase-transcriptase protein P are encoded by a bicistronic pgRNA. Subgenomic preS and S RNAs translate three surface antigens (large [preS1], middle [preS2], and small [S] surface antigens). A short RNA in size of 0.7 kb encodes X protein, a nonstructural viral protein presumably with oncogenic potentials [74]. A precore RNA initiated upstream of the pgRNA encodes HBeAg [75, 76]. Beside protein translation pgRNA is a template for reverse transcription into 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 1,223-nt long intron starting from the end of core antigen ORF to the middle of S antigen ORF. Subsequently, other single and multiple spliced forms of pgRNA were discovered in sizes of 2.1–2.6-kb from cell cultures and liver tissues of HBV patients [78–80]. So far, 13 spliced variants of pgRNA and two spliced isoforms of pre-S2/S RNA have been identified during HBV gene expression infection and these spliced viral RNAs are produced by using six 5′ splice sites and seven 3′ splice sites (Fig. 38.11c). A viral *cis*-element PRE (posttranscriptional regulatory element) and cellular splicing factors such as PTB (polypyrimidine track-binding protein) and SR-proteins may play roles in regulation of HBV RNA splicing (see review ref. [81]).

Approximate 30–50 % of HBV RNA are spliced RNAs during HBV infection of human hepatoma cell lines Huh7 and HepG2, two popular cell lines for in vitro HBV replication studies [82]. Huh7 cells seem to produce more spliced RNA than Hep2G cells. The major spliced product is derived from 30 % of pgRNA using nt 2,447 5′ ss and nt 489 3′ ss in genotypes A, C, D, E. Serum of infected patients or hepatocarcinoma tumor samples frequently contain HBV DNA originated from spliced variants [83, 84]. Level of spliced HBV RNAs in patients varies widely from no splicing to extensive splicing and is related to viral genotype [85]. Role of HBV RNA splicing in HBV life cycle or pathogenesis remains to be elucidated. HBV spliced RNAs express two new proteins [86, 87]. A spliced mRNA derived from pgRNA with removal of a 454-nt intron from nt 2,447 to 2,901 encodes a structural polymerase-surface fusion protein (P-S FP) p43 with potential function in the entry [86]. Another single spliced pgRNA with removal of an intron from nt 2,447 to 489 translates a 93-aa fusion protein in size of 10.4-kDa, of which its first 46-aa residues are identical to the N-terminus of viral polymerase protein followed by the 47-aa residues generated by the frameshift from the second exon. This protein has been
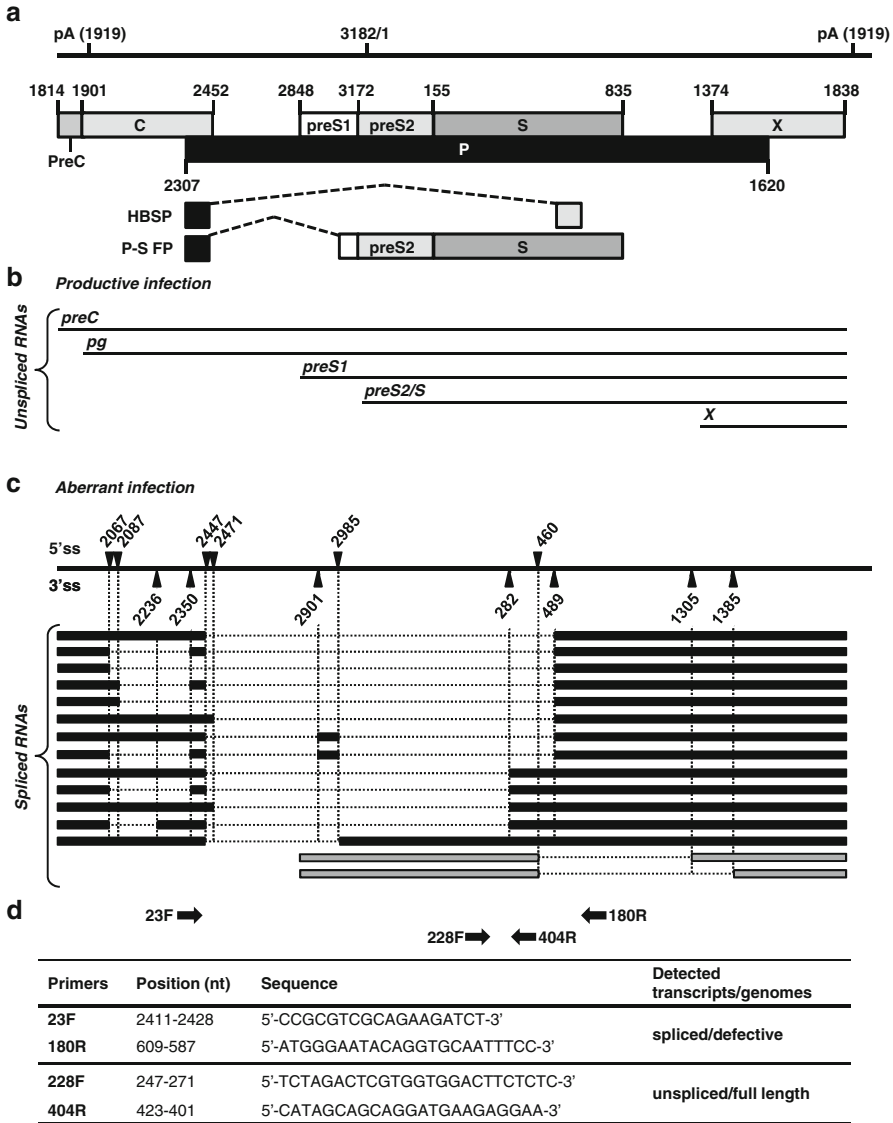
Primers	Position (nt)	Sequence	Detected transcripts/genomes
23F	2411-2428	5'-CCGCGTCGCAGAAGATCT-3'	spliced/defective
180R	609-587	5'-ATGGGAATACAGGTGCAATTTCC-3'	
228F	247-271	5'-TCTAGACTCGTGGTGGACTTCTCTC-3'	unspliced/full length
404R	423-401	5'-CATAGCAGCAGGATGAAGAGGAA-3'	

Fig. 38.11 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 (*grey*) primary transcripts, with exons in *black* or *grey 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 preS1leads 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 Accs. No. JF754635)

referred as hepatitis B splice-generated protein or HBSP [87] and could be associated with HBV chronic infection, viral cytopathogenic effect, and immune evasion (see review ref. [88]).

Parvoviruses

Parvoviruses are a group of small nonenveloped viruses containing a single-stranded DNA genome (ssDNA) in size of ~6-kb. The palindromic inverted terminal repeats at the ends of virus genome function as an origin of replication. Parvoviruses replicate via a double-stranded DNA intermediate which serves as a template for viral transcription [89]. Replication of some parvoviruses relies on "helper" virus 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 *Parvoviridae* family that are infectious to humans: adeno-associated viruses (AAV), parvovirus B19 (B19V), human bocaviruses (HuBoV), and human Parv4 [93]. Despite of structural and genetic similarity, different parvoviruses use different replication and transcription strategies during virus infection and have different host tropism to initiate a productive infection in the presence of a helper virus.

Adeno-associated viruses (AAV), currently classified as *Dependoviruses*, were first human parvoviruses identified in the group. AAVs infect wide range of species with AAVs-1, -2, -3, -8, and -9 being found in human [94]. Currently no disease or pathological condition is associated with AAV infection in human. The correlation between AAV infection and fetal loss and male infertility was proposed due to high prevalence of AAV DNA in placental tissues and in genital tissues of men with abnormal semen [95, 96]. Because AAV lacks pathogenicity, induces low immune response, infects both dividing and nondividing cells with capability of viral DNA integration into the host genome, AAV has gained attention as a vector for gene therapy (see review ref. [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. 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 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 nonconsensus 5′ splice site and two 3′ splice sites (Fig. 38.12a).The efficient splicing requires the presence of both helper virus and large *Rep* protein [98]. While *Rep* protein seems to be essential for AAV2 splicing, several adenovirus proteins (E1A, E1B, E2a, E4 or f6, and VA RNA) as well as some products of herpes simplex virus (UL5, UL8, UL52, and UL29) have stimulatory effect on AAV2 splicing [99]. AAV5 and some animal AAVs are in the second group which utilizes three upstream promoters for their transcription, but their genome contains additional polyadenylation sites in the intron region. Transcripts
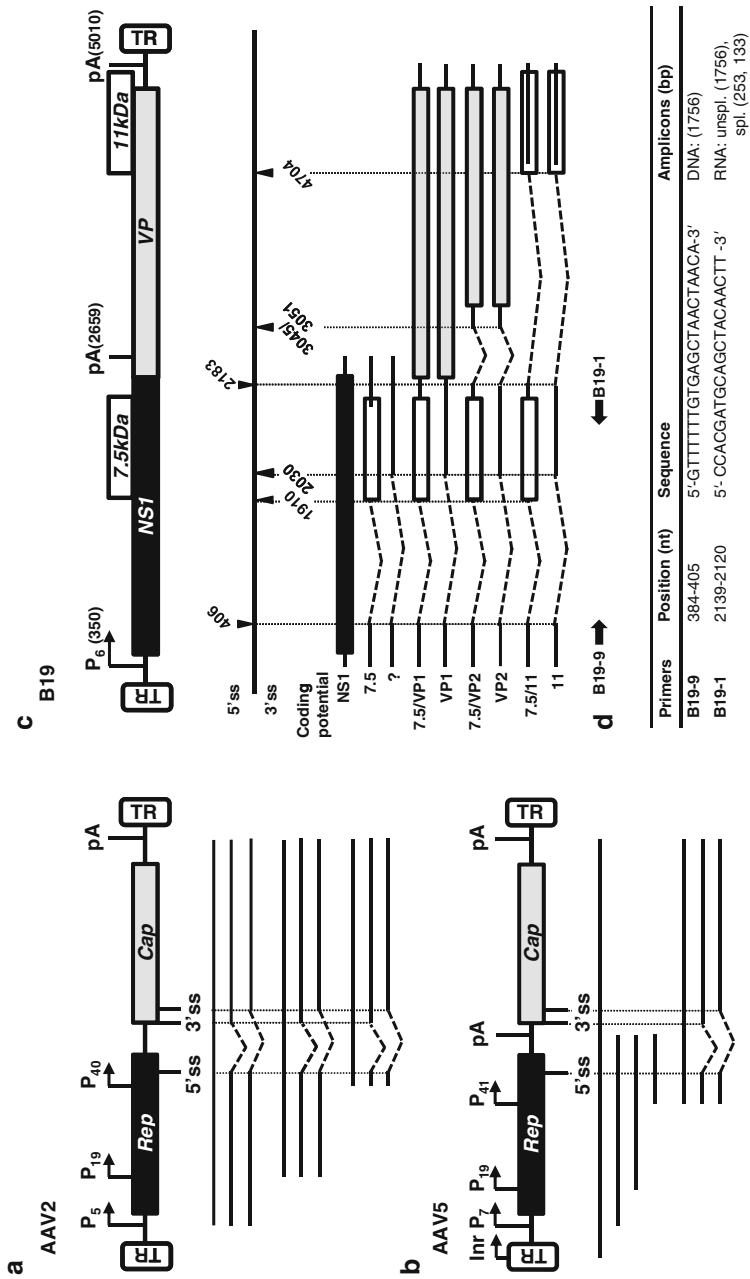
Fig. 38.12 RNA splicing of human parvoviruses. Genome organizations of AAV2 (**a**), AAV5 (**b**), and human *Erythrovirus* B19 (**c**) with terminal repeats (TR) on the ends of each genome, along with viral promoters (P), polyadenylation sites (pA), splice sites (ss), and open reading frames (*boxes*). 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 Accs. No. M13178.1)

from two upstream promoters are polyadenylated on the internal polyA site, whereas spliced transcripts from P41 promoter use a poly A site at the right side of the genome (Fig. 38.12b). 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, member of *Erythrovirus* genus, was first identified in the serum of blood donor [101]. Three 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 in general asymptomatic, but several pathological conditions are associated with B19 infection and 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].

Similar to other parvoviruses, B19 virus genome encodes two large open reading frames. NS1 ORF on the left-side genome translates a 77-kDa nonstructural protein and 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] which are all transcribed solely from a single promoter P6 located upstream of NS1 gene, but are alternatively spliced and terminated at two alternative polyadenylation sites (Fig. 38.12c) 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 to encode 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 become 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 to encode 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 a VP2 or a novel ORF for another accessory 11-kDa protein, depending on which alternative 3′ splice site being selected (Fig. 38.12c). Thus, all detected B19 transcripts are derived from a P6 pre-mRNA containing one or two introns with two alternative 3′ splice site, depending on the selection of which one of two alternative poly A sites, and are alternatively spliced RNA transcripts, except the unspliced full-length NS1RNA. The *cis*-elements in the central exon and intron 2 are regulatory elements to control the alternative P6 RNA splicing, with the double spliced P6 RNAs being the predominant species in the infected cells [107].

Adenoviruses

The most common infection by adenoviruses in humans occurs in upper respiratory tract causing bronchitis and pneumonia. Adenovirus infection can also induce a wide range of other symptoms including 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 induces tumors in small animals. Transformation activities are linked to two oncogenes E1A, which bind tumor suppressor pRB, and E1B, which binds tumor suppressor p53 [109].

Adenoviruses are nonenveloped viruses with icosahedral architecture and contain a linear, nonsegmented, double-stranded DNA genome in size of ~26–45-kb which is capable to encode 22–40 different gene products [109]. Adenoviruses replicates in the nucleus of infected cells (Fig. 38.13a). The early stage of virus infection is characteristic with the expression of nonstructural early protein, while viral structural proteins are expressed in the late stage of viral DNA replication marking switch between two infection phases.
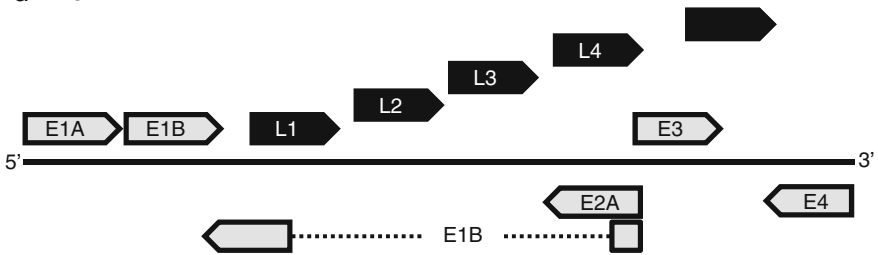
Almost all adenoviral early and late transcripts undergo RNA splicing in order to produce their corresponding viral products [110]. Viral E1A and L1 transcripts are exemplified in this chapter for alternative RNA splicing seen in adenovirus infections.

Adenovirus E1A primary transcript contains three 5′ splice donor sites and two 3′ acceptor sites and is composed of three exons and two introns. The first intron between by D1 and A1 is a suboptimal, minor intron. The second intron is a major intron which uses two alternative donor sites D2 and D3 and one acceptor site A2 for RNA splicing. Alternative splicing of E1A RNA through usage of various combinations of splice donor and acceptor sites leads to formation of five species (13S, 12S, 11S, 10S, and 9S) of E1A mRNAs according to their sedimentation coefficient (Fig. 38.13b) and expression of individual unique protein [111].
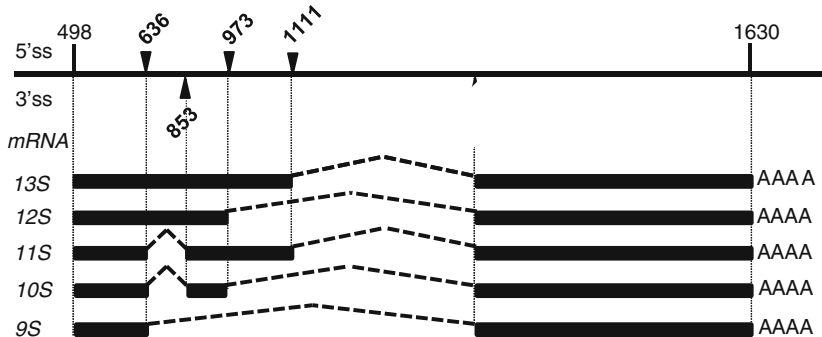
Transcription of late genes starts predominantly from a major late promoter. The primary late transcript is then polyadenylated at one of five polyadenylation sites, forming five groups of late transcripts (L1–L5). Each late mRNA contains a 201-nt "leader" sequence derived from three noncoding exons which function as a translational enhancer [112]. There are two variants of leader sequence with or without i-leader exon. Beside the leader sequence region, L1 transcripts are also alternatively spliced by using a common 5′ splice site in combination with two alternative 3′ splice sites. Selection of a proximal 3′ splice site results in formation of 52, 55K RNA and selection of a distal 3′ splice site produces IIIa mRNA (Fig. 38.13c).

Characteristic feature of adenovirus splicing is depending on the stage of virus infection. For example, E1A 13S and 12S mRNA are two major spliced products during early virus infection. In contrary, 9S RNA is highly accumulated in the late stage of infection [113]. Similar phenomenon was 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 the classical tripartite leader. While 52,55K L1 RNA is produced during both early and late infection, the IIIa splice site is used
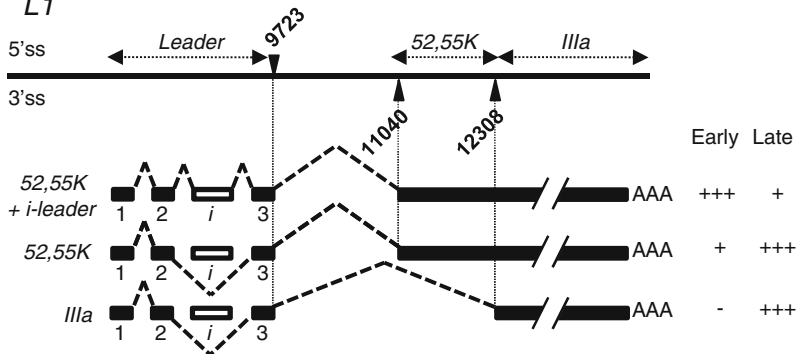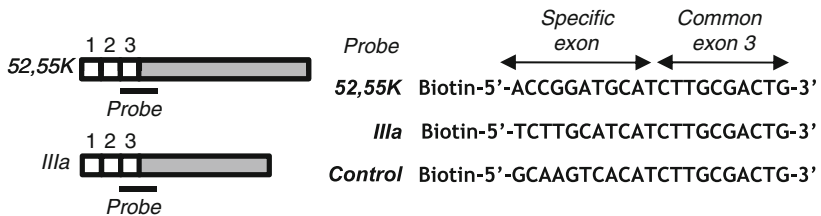
Fig. 38.13 Alternative RNA splicing of adenovirus early and late transcripts. (**a**) Simplified adenovirus genome with positions and orientations of viral early (*grey arrows*) and late (*black arrows*) genes. (**b** and **c**) Alternatively spliced RNA transcripts of adenovirus early E1A gene (**b**) and 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 Acss. 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]

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 in the course of viral infection (see reviews refs. [110, 115]).

Polyomaviruses

Polyomaviruses are small nonenveloped viruses which contains a circular double-stranded DNA (dsDNA) genome in size of ~5,000-bps. Polyomaviruses infect wide range of mammalian and avian species, but each virus exhibits a limited host range and narrow tissue tropism. The *Polyomaviridae* family contains only one genus *Polyomavirus* (*PyV*) which has nine members of human polyomaviruses: BKPyV [116], JCPyV [117], KI PyV [118], WU PyV [119], Merkel cell PyV (MCPyV) [120], HPyV6, HPyV7 [121], trichodysplasia spinulosa-associated PyV (TSV) [122], and most recently discovered HPyV9 [123]. Simian vacuolating virus 40 (SV40), a prototype virus of the family, was introduced into human population as a contaminant in early trials of poliovirus vaccine [124]. Serological data indicate that polyomavirus infection is widespread in general human population with initial infection occurring in childhood [125]. After infection polyomaviruses persist in host for the rest of the life. While initial infection is mostly 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 JC PyV [126], or trichodysplasia spinulosa presumably associated with TSV infection. Polyomaviruses express an oncoprotein T antigen and may lead to development of human cancer by abortive infection as has recently been confirmed in a rare but aggressive Merkel cell carcinoma [127–129].

Polyomavirus genome consists of three functional regions: two protein-coding regions (early and late) divided by a noncoding control region (NCCR) (Fig. 38.14a). Early and late transcripts are expressed in an opposite direction from promoters located in the NCCR which also contains origin of replication. Early transcripts encode nonstructural viral regulatory proteins (T [tumor] antigens) important for virus replication and modulation of cell cycle. Viral DNA replication initiates transcription of viral late genes to encode several viral capsid proteins.

In polyomavirus-infected cells, multiple isoforms of T antigen are detectable as a result of alternative RNA splicing. The primary transcript of T antigen contains two introns, but its 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 a 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
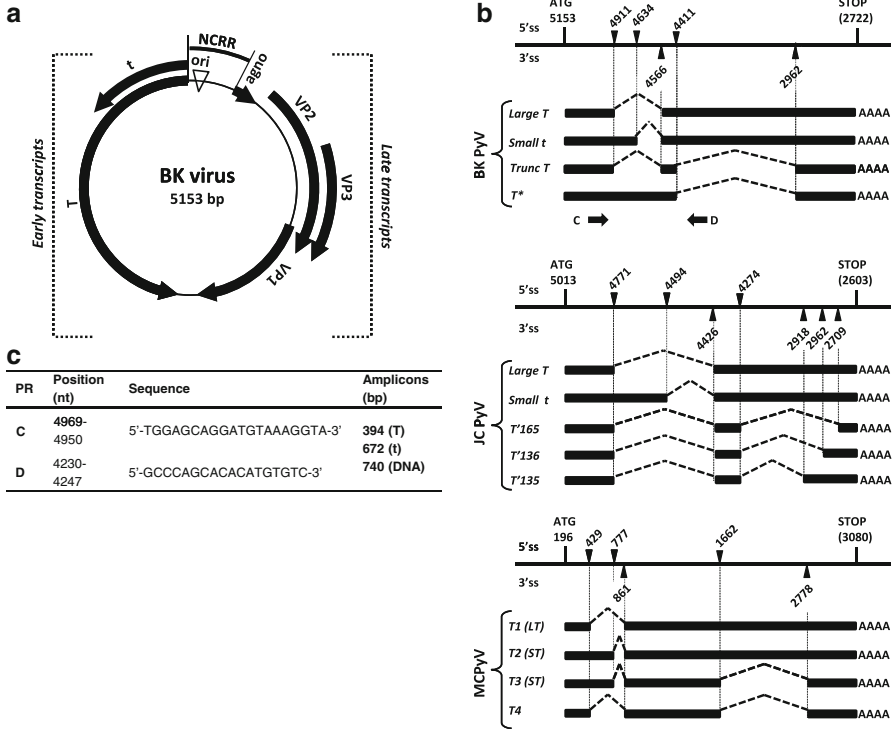
Fig. 38.14 RNA splicing of polyomavirus T antigen transcripts. (**a**) Genome structure of BKPyV virus, a representative of human polyomaviruses. *Black arrows* represent open reading frames for early viral regulatory proteins (large and small T antigens) and late viral capsid proteins (VP1-3). *NCRR* noncoding repeat region, *ori* origin of replication. Agno, auxiliary agnoprotein. (**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 Accs. No. V01108), for JCPyV are strain Mad-1 (GenBank Accs. No. J02226) and for MCPyV are strain TSK (GenBank Accs. No. FJ173815). *Arrows* below the BKPyV transcripts are oligo primers used to detect spliced large T and small t antigen transcripts of strain Dunlop by RT-PCR as detailed in (**c**) [131]

of a smaller protein (Fig. 38.14b). In addition, a rare tiny-*t* antigen in size 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 were detected also in other polyomaviruses such as truncated T-antigen (truncTAg) in BKPyV [131], T′135, T′136 and T′165 in JCPyV [132] and T3 and T4 early transcripts in MCPyV [127] (Fig. 38.14b). Alternative splicing of polyomavirus early transcripts allows expression of multiple T-antigens with distinguished function in viral life cycle. Beside the cells with actively replicating virus the early viral transcripts are expressed also in the cells

with nonproductive infection or in the polyomavirus-transformed cells. These cells often do not expressed late gene due to integration of viral DNA into host genome resulting in disregulated viral gene expression and cell transformation.

Papillomaviruses

Human papillomaviruses (HPVs) are a group of small DNA tumor viruses and has a genome in size of ~8-kb surrounded by a viral capsid. HPV genome consists of three regions: viral early, late, and noncoding regions and in general encodes eight viral genes (E1, E2, E4, E5, E6, E7, L1, and L2). Viral early gene products are regulatory proteins responsible for virus multiplication and pathogenesis during a productive infection, whereas 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 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. 38.15a). As a result, the 5′ sequences of viral L1 and L2 are part of the viral early transcript sequences. RNA splicing to removal the most of these early gene sequences from the RNA by RNA splicing 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 for women death in the developing world, with about 493,000 new cases and nearly 273,000 deaths each year (www.who.int/hpvcentre). More than 120 genotypes of HPVs have been identified to date and are grouped into two major groups according to their pathogenesis and association with cervical cancer [134]. The high-risk or oncogenic HPV types are present in cervical cancers, whereas low-risk or nononcogenic HPVs are not found in cervical or other anogenital cancers [135]. In general, women acquire HPV infection by sexual contact. Various epidemiology studies indicate that women with repeat exposure to oncogenic HPVs and with persistent cervical infection of oncogenic HPVs are in high risk to develop cervical cancer [136, 137]. Infection with oncogenic HPV-16 and HPV-18, two most common oncogenic HPV types, leads to development of almost 70 % of all cervical and other anogenital cancers. Viral E6 and E7 of oncogenic HPVs are two viral oncoproteins that inactivate, respectively, cellular p53 and pRB, 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 and majority of the E6/E7 bicistronic RNA are alternatively spliced as diagramed for HPV-16 and HPV-18 (Fig. 38.15b). 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
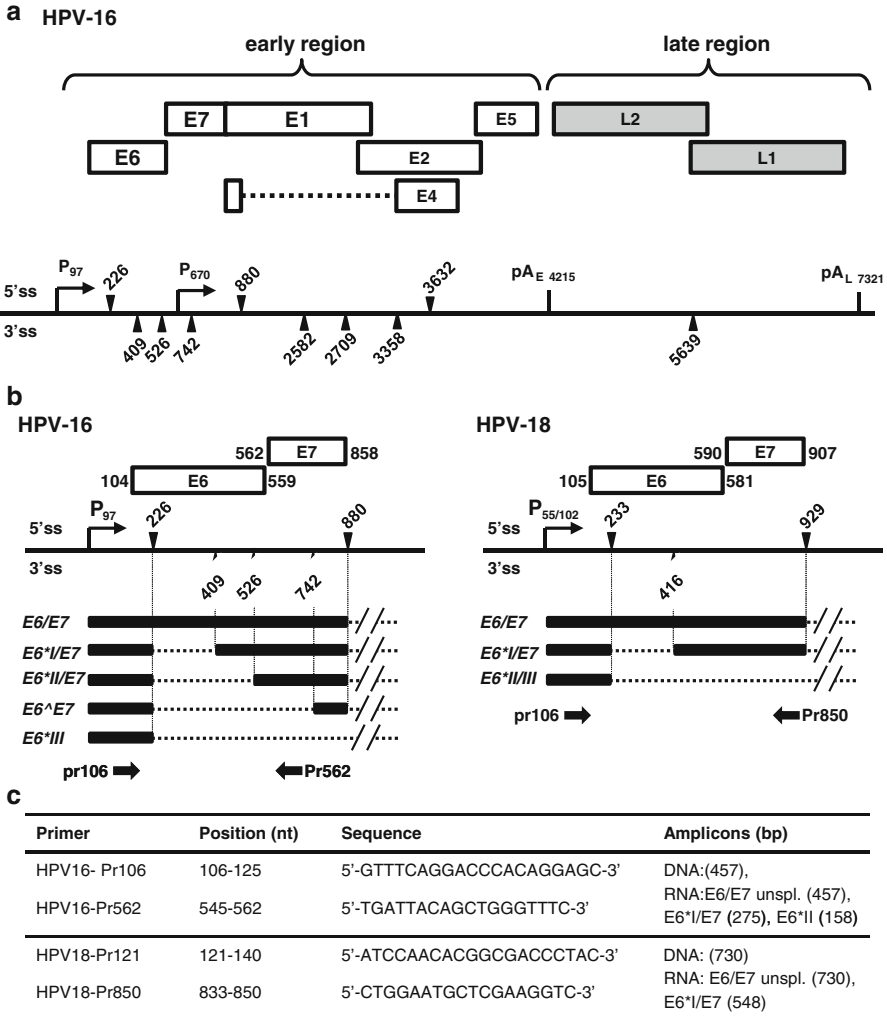
Fig. 38.15 RNA splicing of viral oncogene E6 and E7 transcripts in high-risk human papillomavirus infections. (**a**) Genome structure of high-risk human papillomavirus type 16 divided by early (genes E1–E7, *open boxes*) and late (L1–L2, *grey boxes*) regions and positions of splice sites in HPV-16 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 -18 infections [140, 141]. Primer nucleotide positions and sequences are based on corresponding HPV reference strains available on http://pave.niaid.nih.gov

for HPV-18 (Fig. 38.15b). 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 RT-PCR methods [140, 141].

The presence of high-grade premalignant lesions (CIN, cervical intraepithelial neoplasia) caused by oncogenic HPV infection is a sign of increased risk of developing cervical cancer. These lesions can be detected by routine cervical examination and treated by surgery to prevent progression to cervical cancer. Papanicolaou test (also called Pap smear) is a screening test used in gynecology to detect premalignant and malignant cells in cervical swabs. A woman who has Pap smear with abnormal cells may also be referred for HPV DNA testing by two FDA-approved assays: Hybrid Capture 2 DNA test to detect 13 high-risk HPVs (HPV-16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) from Qiagen [142] or Cobas 4800 System to detect 14 high-risk HPVs (HPV-16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68) from Roche [143]. More recently, a few of HPV E6/E7 RNA tests have been introduced. APTIMA HPV Assay from Gen-Probe was designed to detect HPV E6/E7 mRNA from 14 high-risk types (HPV-16, -18, -31, -33, -35, -39, -45, -51, -52, -56, -58, -59, -66, and -68) [144] with a sensitivity and specificity similar to or better than Hybrid Capture 2 DNA test [145, 146]. PreTect HPV-Proofer from NorChip was designed to detect E6 and E7 RNA from HPV types 16, 18, 31, 33, and 45 [147, 148], with more specific than HC2 in identifying women with CIN 2+ but has a lower sensitivity [149]. By using the primers detailed in Fig. 38.15c for RT-PCR assays, the spliced E6/E7 RNAs of HPV-16 and HPV-18 can be detected easily based on the amplicon size smaller than E6/E7 DNA, without worry of carryover viral DNA contamination encountered by HPV DNA tests.

Herpesviruses

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

Currently there are more than 100 known herpesviruses infecting wide range of animal species. All human herpesviruses belong to *Herpesviridae* family which is further grouped into four subfamilies: *Alphaherpesvirinae*, *Betaherpesvirinae*, *Gammaherpesvirinae*, and unassigned viruses. As of today, eight herpesvirus species have been isolated from humans and they are assigned to three subfamilies of *Herpesviridae*. These include the following: herpes simplex virus type 1 [HSV-1, also referred as human herpesvirus 1 (HHV1)], herpes simplex virus type 2 (HSV-2 or HHV2), Varicella-zoster virus (VZV or HHV3), Epstein-Barr virus (EBV or

HHV4), human cytomegalovirus (CMV or HCMV or HHV5), human herpesvirus 6 (HHV6), human herpesvirus 7 (HHV7), and Kaposi sarcoma-associated herpesvirus (KSHV or HHV8).

Herpesviruses replicate in the nucleus. After virus entry the viral genome is translocated to the nucleus of infected cells where expression of viral genes occurs. All herpesviruses have two types of viral life cycle, latent and lytic, with distinctive transcriptional profile. Latent infection is characterized with the expression of a few viral genes (latent transcripts) to maintain viral genome in latently infected cells. Lytic infection is associated with viral genome replication and production of infectious virions and generally leads to destruction of infected cell. In contrast to latent infection almost all of viral lytic genes in viral lytic infection are expressed in 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. At some circumstances the virus in latently infected cells may be reactivated and proceeded to lytic infection. The mechanisms controlling the establishment of latency and reactivation of herpesviruses are not fully understood. Herpesviral genome encodes up to 100 different genes including a variable number of noncoding genes for noncoding RNAs or viral miRNAs [150–152].

Most human herpesviruses are highly prevalent in general population. Initial infection generally occurs in childhood or early adolescence through body contact and follows by establishment of latent infection. Some herpesviruses are sexually transmitted. Blood transfusion, tissue transplantation or congenital transmission represents other ways of acquiring the virus. Primary infection often occurs at epithelia, the point of entry, followed by establishment of latent infection generally in a specialized cell type (neurons or lymphocytes) which serves as a virus reservoir. Recurrence of infection is caused by virus reactivation from latency when virus escapes from host immunological surveillance. Overall symptoms of herpesvirus infections in healthy individuals are generally mild, but may be life threatening in immunocompromised patients. While infections of some herpesviruses such as EBV and KSHV are clearly etiologically linked to the development of several types of cancer, the role of other human herpesviruses in cell transformation remains unknown [153]. Several compounds are used to treat acute herpesvirus infection. No vaccine against herpesviruses is currently available.

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 in multiple time points these techniques cannot 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. DNA contamination problem 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 to DNA based on its size. The number of spliced viral transcripts varies from one herpesvirus to another, with only handful split genes in HSV-1 to about 30 % in KSHV [154]. Both latent and lytic genes could have an intron and sometimes are alternatively spliced.

Herpes Simplex Viruses

HSV-1, a member of *Alphaherpesvirinae* subfamily, is an human neurotropic herpesvirus associated with "herpes labialis," Bell's palsy and vestibular neuritis [155]. After initial infection, HSV-1 establishes the latent infection in sensory neural ganglia from where it periodically reactivates. Viral genome consists of two unique regions (long and short) flanked with inverted repeat regions (internal or terminal) (Fig. 38.16a). HSV-1 encodes at least 84 genes. The majority of gene are named based on their position within a particular part of viral genome like UL1 (unique long region ORF1) or US3 (unique short region ORF3) while others have alternative historical manes, like ICP0 (infected cells protein 0). Only few HSV-1 transcripts are spliced (LAT, ICP0, UL15, US1, US12/ICP47) (see review ref. [156]) including both latent and lytic transcripts.

During latency HSV-1 expressed LAT (*l*atency-*a*ssociated *t*ranscript) RNA from a repeat region of 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 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 removing an intron of 559 or 556-bp depending on the virus strain [160] (Fig. 38.16a). Both LAT RNAs are uncapped without a poly A tail and accumulates in the nucleus of infected cells. HSV-1 LAT RNA is a noncoding regulatory RNA for establishment and maintenance of viral latency by inhibiting expression of viral lytic genes and interfering with cellular apoptosis pathway [161]. Recent studies showed that LAT transcript functions as a precursor for generation of virus-encoded miRNAs [162]. Expression from LAT-DNA was observed also 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 a 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 nonspecific 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. 38.16a). 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 alpha herpesvirus subfamily. Genital infection with HSV-2 causes genital herpes which is considered as 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 a similar genome and gene structures including their LAT and ICP0 regions [168].

a HSV-1
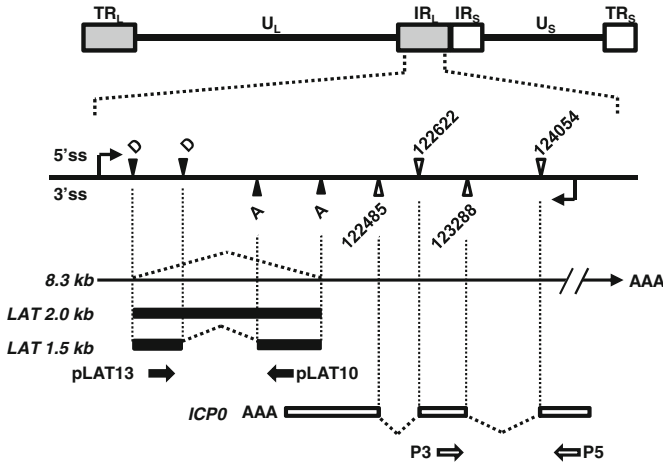
Primer	Position (nt)	Sequence	Amplicons (bp)
pLAT10	120408-120389	5'-ATGGAGCCAGAACCACAGTG-3'	DNA: (947)
pLAT13	119464 -119484	5'-GTAGGTTAGACACCTGCTTCT-3'	RNA: LAT 2.0 kb (947), LAT 1.5 kb (297)
P3	123214-123233	5'-TTCGGTCTCCGCCTGAGAGT-3'	DNA: (922)
P5	124116-124135	5'-GACCCTCCAGCCGCATACGA-3'	RNA: ICP0 unspl. (922), ICP0 spl. 1.5 kb (157)

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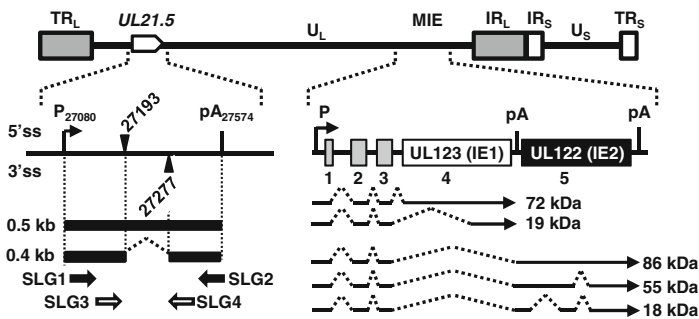
Primer	Position (nt)	Sequence	Amplicons (bp)
SLG1	27120-27139	5'-CTATGGATCTTGAGCTTACT-3'	
SLG2	27410-27429	5'-TCGCTGCCATCTCCGTCTGT-3'	DNA: (258) RNA: unsliced (258) spliced (175)
SLG3	27144-27144	5'-GTGACCTTGACGGTGGCTTT-3'	
SLG4	27382-27401	5'-CGTCATACTCCCCGGAGTAA-3'	

Fig. 38.16 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) consists of two unique regions (long-U_L and short-U_S) flanked with terminal (TR) and internal (IR) repeats.

In general the infection of HSV-1 and -2 is controlled by host immune system. Thus initial or recurrent infections are associated with only mild symptoms. Infection in immunocompromised patients could cause several severe diseases including encephalitis [169]. Genital infections or reactivation of HSV-2 during pregnancy could lead to congenital infection [170]. Detection of viral DNA may not provide sufficient information about virus replication status due to the permanent presence of viral DNA in the infected cells. Detection viral lytic products, such as spliced ICP0 RNA, may be a better predictor of virus reactivation even before occurrence of clinical symptoms for early diagnosis enabling of 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 *Betaherpesvirnae*. High prevalence of CMV infection has been noticed in 50–80 % human population. In majority of healthy individuals, the primary CMV infection occurs asymptomatically, but in some cases could be associated with sore throat, prolong fever or syndrome similar to infectious mononucleosis. After initial infection the virus remains latent in T cells for the rest of the host life without apparent symptoms. In contrast, the CMV infection in immunocompromised individuals, such as newborns, transplant recipients, people with AIDS or cancer patients, could 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 to encode around 200 genes (reviewed in ref. [171]). While the majority of CMV transcripts are intronless, the presence of several split genes has been identified in all kinetic classes of viral genes [172]. A major immediate early region (MIE) located within a unique long (UL) region of CMV genome contains several genes highly expressed at 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

Fig. 38.16 (continued) 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 Accs. 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 Accs. No. X17403.1)

splicing. MIE transcripts IE1 and IE2 are expressed from the same promoter, but alternatively polyadenylated. These transcripts have five major exons and can be alternatively spliced to express additional isoforms of IE1 and IE2 proteins (Fig. 38.16b). Splicing was detected also 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 which encodes viral glycoprotein is expressed from the UL region posited from nt 27,080 to nt 27,574 of CMV genome [173], and has a short intron with 83 nts. Removal of this intron leads to production of a mature mRNA in size of ~0.4-kb (Fig. 38.16b). Both spliced and unspliced UL21.5 RNAs are easily detectable by RT-PCR from infected cells [174].

Allogenic bone marrow transplant recipients are in high risk to develop CMV diseases. Historically, viremia was used as an indicator of CMV disease to guide preemptive treatment. Multiple approaches have been developed to detect CMV viremia in circulating lymphocytes by direct virus isolation with cultivation or by detection of viral antigens or viral DNA [175, 176]. However, the detection of viremia is not sufficient in disease prediction since many viremic patients never develop symptoms. An 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, with no worry by DNA contamination. Detection of spliced immediate early transcripts had been reported in good correlation with detection of viral DNA or viral antigen [178–180]. Detection of late gene UL21.5 has better prediction value and significantly correlates with disease progression [174, 181, 182].

Epstein-Barr Virus

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

for EBV-mediated cell transformation [115]. Several types of EBV latency were defined by variable expression of latent genes in various malignancies [186].

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. 38.17a). EBV encodes at least 80 viral proteins [187] and several noncoding RNAs including viral miRNAs [188, 189]. The ORF names are derived from their positions in *Bam*HI fragment (from A to Z) by orientation (L-left or R-right) and a digital number representing the frame (F) order (for example BZLF1). Other genes retain their historical names based on the gene product function. Number of split genes in EBV is significantly higher than that found in alpha- and beta-herpesviruses. Extensive alternative RNA splicing is prominent especially for almost all 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 in size of ~100 kb. This transcript is originated from one of two alternative promoters, Cp or Wp, which are named by their localization in a particular *Bam*HI fragment of viral genome (Fig. 38.17b). At the early stage of latent infection, the Wp is initially used, but the expressed EBNA1 and EBNA2 from the Wp transactivate the Cp promoter and cause a switch of transcription from Wp to Cp [191]. Usage of Cp promoter is associated with EBV "latency type III." In Burkitt's lymphoma and Burkitt's lymphoma-derived cell lines, EBNA1 expression is initiated from a distal Qp promoter rather than from Cp and Wp and is associated with "latency type I" [192]. EBNA is expressed also in the lytic phase from additional Fp promoter closely localized 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] to encode 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. 38.17c). Thus, the Zp promoter transcript is a monocistronic ZEBRA RNA containing two constitutive introns and splicing of these two introns results in production of a 0.9-kb mRNA to encode ZEBRA protein. Transcription from the Rp promoter leads to produce a 3.8-kb bicistronic transcript, ZEBRA/RTA, which contains two additional introns over 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 a potential to encode ZEBRA and RTA proteins. A third, minor isoform of ZEBRA/RTA transcript is derived from splicing of additional internal intron spanning over BRLF1 ORF to BZLF1 ORF and this splicing produces a RAZ transcript in size of ~0.9-kb to encode 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 of latently infected nondividing B-cells with "latency type 0."
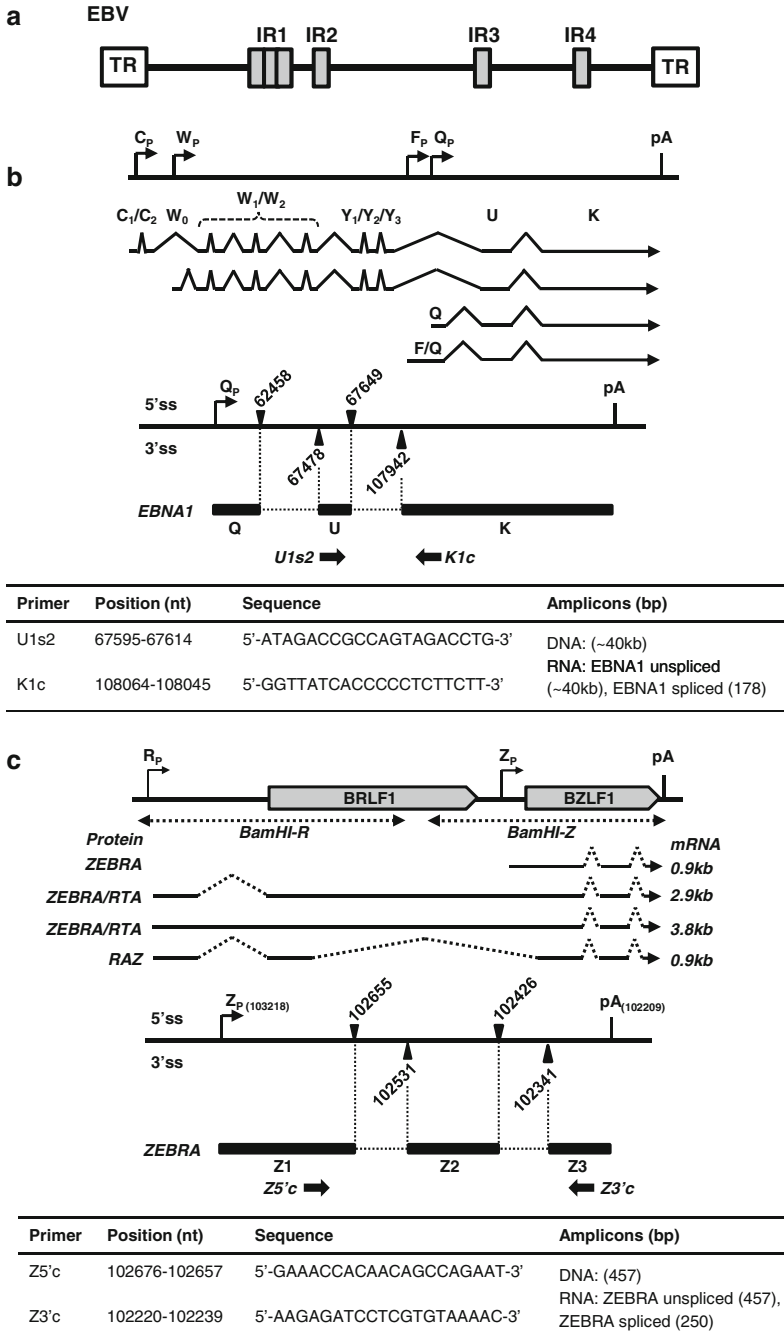
Fig. 38.17 RNA splicing of EBNA1 and ZEBRA transcripts in EBV infections. (**a**) Organization of EBV genome with terminal repeats (TR) and internal repeats (IR1-4). (**b**) Multiple transcripts of EBV latency-associated EBNA1 transcribed from several alternative promoters (C_p-Q_p).

This makes the detection of EBNA-1 expression as a good marker for the presence of EBV in tumors. The expression of ZEBRA during lytic infection could be used to monitor a productive EBV infection and EBV reactivation.

Kaposi Sarcoma-Associated Herpesvirus

Kaposi sarcoma-associated herpesvirus (KSHV) is the latest discovered human herpesvirus [197]. After primary infection, KSHV establishes latent infection in endothelial cells and B cells [198]. In healthy individuals both primary and latent KSHV infections are generally asymptomatic. Suppression of immune system in KSHV-positive individuals, such as 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 ref. [199]). Presence of viral genome and expression of viral-encoded products in all cancer cells support active roles of KSHV in cell transformation.

KSHV belongs to *Gammaherpesvirinae* and has similar genome organization as EBV with long unique region flanked with terminal repeats (Fig. 38.18a). KSHV genome (~165 kb) encodes up to 90 genes named by their position in viral genome from left to right (for example ORF47) [200]. Some genes are designated with 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 % viral genes, including both latent and lytic genes, undergo RNA splicing [154].

During latency, 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 which has an intron containing two alternative 3′ splice site. Alternative RNA splicing and alternative

---

Fig. 38.17 (continued) 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 EBNA1 transcript from exon U to exon K [185] are 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 Accs. No. V01555.2)
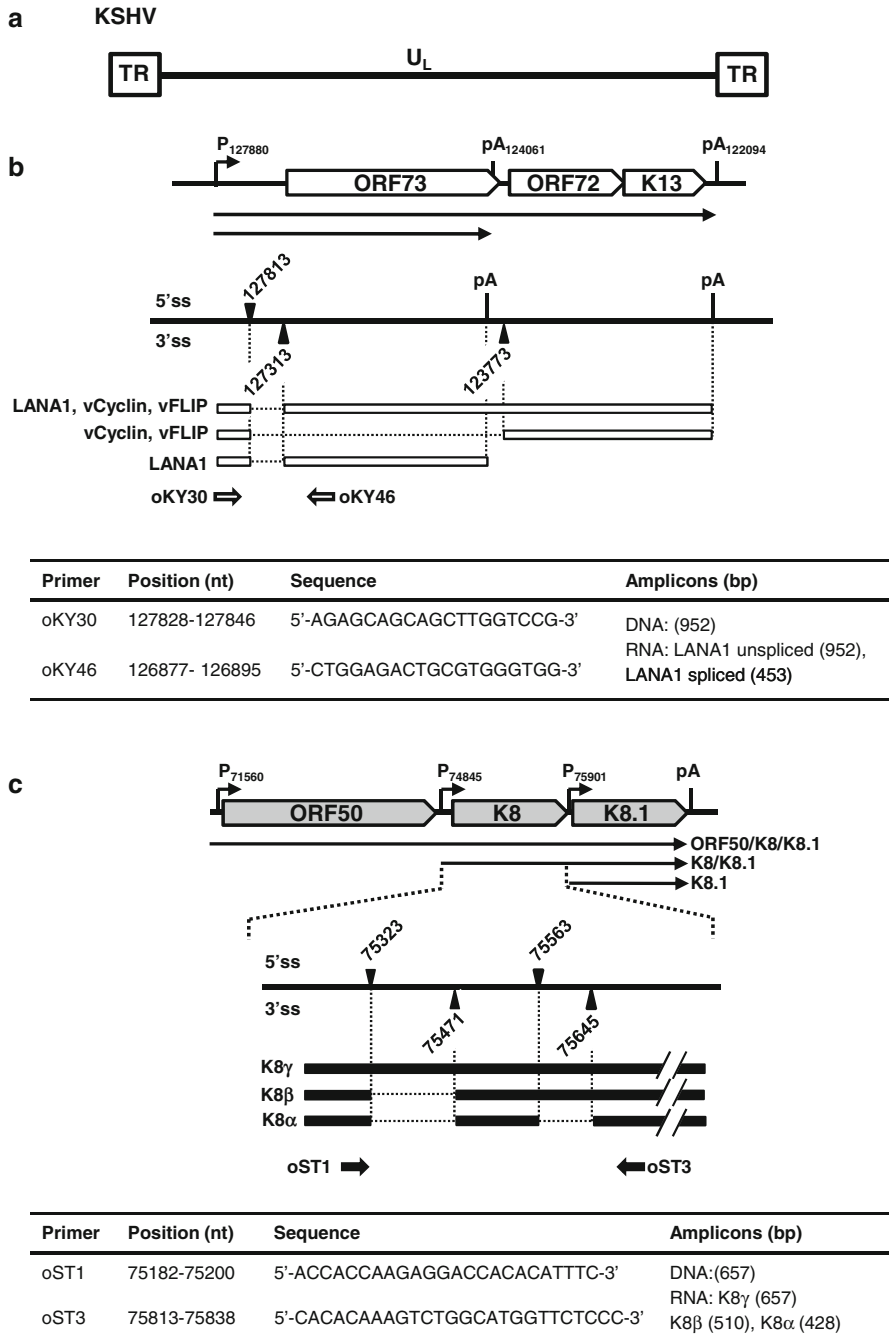
Fig. 38.18 RNA splicing of representative latent transcripts and early transcripts in KSHV infections. (**a**) Genome of KSHV contains a long unique region (U_L) flanked by terminal repeats (TR). (**b**) Gene structure of a latent gene locus containing ORF73/ORF72/K13 genes. Three genes are

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)

RNA polyadenylation of the tricistronic pre-mRNA results in production of three mature mRNAs (5.4-, 3.3-, and 1.7-kb) [202] (Fig. 38.18b). The 5.4-kb transcript most likely for LANA-1 expression is produced by usage of the proximal 3′ splice site, whereas usage of the distal 3′ splice site leads to express the 1.7-kb transcript for vCyclin and vFLICE. Both transcripts are polyadenylated at the same distal polyadenylation site. The minor 3.3-kb transcript use the proximal splice site for RNA splicing, but is polyadenylated at a proximal noncanonical polyadenylation signal (Fig. 38.18b).

KSHV lytic replication is controlled by a major viral transactivator, ORF50 (also referred as Rta) [203, 204]. Similarly to LANA1, ORF50 posits along with K8 and K8.1 in a larger gene locus (ORF50/K8/K8.1 cluster) (Fig. 38.18c) 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 ref. [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 K8 and K8.1 transcripts. The ORF50 transcript is tricistronic, K8 is bicistronic, and K8.1 is monocistronic in nature. The bicistronic K8 full transcript composes of four exons separated by three introns (Fig. 38.18c). 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 spliced K8 region which detects the expression of both ORF50 tricistronic and K8 bicistronic transcripts could be used to monitor viral lytic replication [206].

Fig. 38.18 (continued) 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 Accs. No. U75698.1) are diagrammed further below with coding potentials of each spliced product on the left. *Open boxes* are exons and *dashed 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 structures 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 Accs. No. U75698.1). Names of three common forms of K8 transcripts from alternative RNA splicing are shown on the left. oST1 and oST3 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 our readers with knowledge of viral RNA splicing during viral infection and the detection of spliced viral RNA transcripts as a new approach in diagnostic virology. In the first part, we provided basic information about the mechanisms of RNA splicing and methodological approaches of specific detection of spliced RNA molecules. 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 various viral pathogens often in the same time by multiplexing. The rapid setup of these methods is especially important for quick response to emerging viruses as has recently been proven in the case of SARS and avian flu pandemic, when nucleic acid amplification was rapidly employed to detect and to confirm the infection. The less material requirement and simplicity make these detection methods suitable for applications in low resources setting such laboratories of the first contact and field laboratories. Because amplification of nucleic acid molecules as a routine in many diagnostic laboratories is used to detect the genomic sequences from many viruses, the detection of spliced viral transcripts could be performed simultaneously to already existing methods.

The second part of the chapter summarized our current knowledge about viral RNA splicing events in the majority of known human viruses. We also included some viral agents, such as human circoviruses and adeno-associated viruses, where a direct link between infection and pathological manifestation remains to be determined. In addition, we have provided 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. We believe that systematic study of RNA splicing events during viral infection could lead to better viral diagnosis, better therapy management, and eventually leading to our better understanding of pathogenesis of human viral pathogens.

References

1. Wiedbrauk DL (2009) Nucleic acid amplification and detection methods. In: Specter S, Hodinka RL, Young SA, Wiedbrauk DL (eds) Clinical virology manual, 4th edn. ASM Press, Washington, DC, pp 156–184
2. Berget SM, Moore C, Sharp PA (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A 74(8):3171–3175
3. Chow LT, Gelinas RE, Broker TR, Roberts RJ (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. Cell 12(1):1–8
4. Fong N, Bentley DL (2001) Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev 15(14):1783–1795
5. Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. Nature 416(6880):499–506
6. Tarn WY, Steitz JA (1996) A novel spliceosome containing U11, U12, and U5 snRNPs excises a minor class (AT-AC) intron in vitro. Cell 84(5):801–811

7. Wu Q, Krainer AR (1999) AT-AC pre-mRNA splicing mechanisms and conservation of minor introns in voltage-gated ion channel genes. Mol Cell Biol 19(5):3225–3236

8. Wachtel C, Manley JL (2009) Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Mol Biosyst 5(4):311–316

9. Toor N, Rajashankar K, Keating KS, Pyle AM (2008) Structural basis for exon recognition by a group II intron. Nat Struct Mol Biol 15(11):1221–1222

10. Chasin LA (2007) Searching for splicing motifs. Adv Exp Med Biol 623:85–106

11. Long JC, Caceres JF (2009) The SR protein family of splicing factors: master regulators of gene expression. Biochem J 417(1):15–27

12. Han SP, Tang YH, Smith R (2010) Functional diversity of the hnRNPs: past, present and perspectives. Biochem J 430(3):379–392

13. Kornblihtt AR (2007) Coupling transcription and alternative splicing. Adv Exp Med Biol 623:175–189

14. Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. Cell 144(1):16–26

15. Zheng ZM (2004) Regulation of alternative RNA splicing by exon definition and exon sequences in viral and mammalian gene expression. J Biomed Sci 11(3):278–294

16. Wang ET, Sandberg R, Luo S et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456(7221):470–476

17. Croft L, Schandorff S, Clark F, Burrage K, Arctander P, Mattick JS (2000) ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. Nat Genet 24(4):340–341

18. Livak KJ, Flood SJ, Marmaro J, Giusti W, Deetz K (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl 4(6):357–362

19. Tyagi S, Kramer FR (1996) Molecular beacons: probes that fluoresce upon hybridization. Nat Biotechnol 14(3):303–308

20. Whitcombe D, Theaker J, Guy SP, Brown T, Little S (1999) Detection of PCR products using self-probing amplicons and fluorescence. Nat Biotechnol 17(8):804–807

21. Hughes TR, Hiley SL, Saltzman AL, Babak T, Blencowe BJ (2006) Microarray analysis of RNA processing and modification. Methods Enzymol 410:300–316

22. Cuperlovic-Culf M, Belacel N, Culf AS, Ouellette RJ (2006) Microarray analysis of alternative splicing. OMICS 10(3):344–357

23. Pinzani P, Orlando C, Pazzagli M (2006) Laser-assisted microdissection for real-time PCR sample preparation. Mol Aspects Med 27(2–3):140–159

24. Esposito G (2007) Complementary techniques: laser capture microdissection—increasing specificity of gene expression profiling of cancer specimens. Adv Exp Med Biol 593:54–65

25. Niedobitek G, Herbst H (2006) In situ detection of Epstein-Barr virus and phenotype determination of EBV-infected cells. Methods Mol Biol 326:115–137

26. Knoll JH, Lichter P, Bakdounes K, Eltoum IE (2007) In situ hybridization and detection using nonisotopic probes. Curr Protoc Mol Biol Chapter 14: Unit 14.7

27. Speel EJ, Hopman AH, Komminoth P (2006) Tyramide signal amplification for DNA and mRNA in situ hybridization. Methods Mol Biol 326:33–60

28. Yamamichi N, Shimomura R, Inada K et al (2009) Locked nucleic acid in situ hybridization analysis of miR-21 expression during colorectal cancer development. Clin Cancer Res 15(12):4009–4016

29. Bonvicini F, Mirasoli M, Gallinella G, Zerbini M, Musiani M, Roda A (2007) PNA-based probe for quantitative chemiluminescent in situ hybridisation imaging of cellular parvovirus B19 replication kinetics. Analyst 132(6):519–523

30. Kang JG, Majerciak V, Uldrick TS et al (2011) 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 225(3):378–389

31. Blanco AM, Artero R (2010) A practical approach to FRET-based PNA fluorescence in situ hybridization. Methods 52(4):343–351

32. de Planell-Saguer M, Rodicio MC, Mourelatos Z (2010) Rapid in situ codetection of noncoding RNAs and proteins in cells and formalin-fixed paraffin-embedded tissue sections without protease treatment. Nat Protoc 5(6):1061–1073

33. Su Z, Ning B, Fang H et al (2011) Next-generation sequencing and its applications in molecular diagnostics. Exp Rev Mol Diagn 11(3):333–343

34. Nakamura S, Yang CS, Sakon N et al (2009) Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PLoS One 4(1):e4219

35. Davis CA, Grate L, Spingola M, Ares M Jr (2000) Test of intron predictions reveals novel splice sites, alternatively spliced mRNAs and new introns in meiotically regulated genes of yeast. Nucleic Acids Res 28(8):1700–1706

36. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921

37. Bouvier NM, Palese P (2008) The biology of influenza viruses. Vaccine 26(Suppl 4): D49–D53

38. Lamb RA, Horvath CM (1991) Diversity of coding strategies in influenza viruses. Trends Genet 7(8):261–266

39. Lamb RA, Choppin PW (1979) Segment 8 of the influenza virus genome is unique in coding for two polypeptides. Proc Natl Acad Sci U S A 76(10):4908–4912

40. Lamb RA, Lai CJ (1980) Sequence of interrupted and uninterrupted mRNAs and cloned DNA coding for the two overlapping nonstructural proteins of influenza virus. Cell 21(2):475–485

41. Briedis DJ, Lamb RA (1982) Influenza B virus genome: sequences and structural organization of RNA segment 8 and the mRNAs coding for the NS1 and NS2 proteins. J Virol 42(1):186–193

42. Nakada S, Graves PN, Palese P (1986) The influenza C virus NS gene: evidence for a spliced mRNA and a second NS gene product (NS2 protein). Virus Res 4(3):263–273

43. Lamb RA, Lai CJ, Choppin PW (1981) 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 78(7):4170–4174

44. Shih SR, Suen PC, Chen YS, Chang SC (1998) A novel spliced transcript of influenza A/WSN/33 virus. Virus Genes 17(2):179–183

45. Chiang C, Chen GW, Shih SR (2008) Mutations at alternative 5′ splice sites of M1 mRNA negatively affect influenza A virus viability and growth rate. J Virol 82(21):10873–10886

46. Yamashita M, Krystal M, Palese P (1988) Evidence that the matrix protein of influenza C virus is coded for by a spliced mRNA. J Virol 62(9):3348–3355

47. Pekosz A, Lamb RA (1998) 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 95(22):13233–13238

48. Hongo S, Sugawara K, Muraki Y et al (1999) Influenza C virus CM2 protein is produced from a 374-amino-acid protein (P42) by signal peptidase cleavage. J Virol 73(1):46–50

49. Stoltzfus CM (2009) Chapter 1. Regulation of HIV-1 alternative RNA splicing and its role in virus replication. Adv Virus Res 74:1–40

50. Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN (1990) Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol 64(6):2519–2529

51. Tazi J, Bakkour N, Marchand V, Ayadi L, Aboufirassi A, Branlant C (2010) Alternative splicing: regulation of HIV-1 multiplication as a target for therapeutic action. FEBS J 277(4): 867–876

52. Landry S, Halin M, Lefort S et al (2007) Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology 4:71

53. Fischer U, Pollard VW, Luhrmann R et al (1999) Rev-mediated nuclear export of RNA is dominant over nuclear retention and is coupled to the Ran-GTPase cycle. Nucleic Acids Res 27(21):4128–4134

54. Lutzelberger M, Reinert LS, Das AT, Berkhout B, Kjems J (2006) A novel splice donor site in the gag-pol gene is required for HIV-1 RNA stability. J Biol Chem 281(27):18644–18651

55. Benko DM, Schwartz S, Pavlakis GN, Felber BK (1990) A novel human immunodeficiency virus type 1 protein, tev, shares sequences with tat, env, and rev proteins. J Virol 64(6):2505–2518

56. Salfeld J, Gottlinger HG, Sia RA, Park RE, Sodroski JG, Haseltine WA (1990) A tripartite HIV-1 tat-env-rev fusion protein. EMBO J 9(3):965–970

57. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC (1980) 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 77(12):7415–7419

58. Yoshida M, Miyoshi I, Hinuma Y (1982) 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 79(6):2031–2035

59. Gessain A, Barin F, Vernant JC et al (1985) Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. Lancet 2(8452):407–410

60. Osame M, Usuku K, Izumo S et al (1986) HTLV-I associated myelopathy, a new clinical entity. Lancet 1(8488):1031–1032

61. Gessain A, de TG (1996) Geographic and molecular epidemiology of primate T lymphotropic retroviruses: HTLV-I, HTLV-II, STLV-I, STLV-PP, and PTLV-L. Adv Virus Res 47:377–426

62. Bindhu M, Nair A, Lairmore MD (2004) Role of accessory proteins of HTLV-1 in viral replication, T cell activation, and cellular gene expression. Front Biosci 9:2556–2576

63. Cavanagh MH, Landry S, Audet B et al (2006) HTLV-I antisense transcripts initiating in the 3′LTR are alternatively spliced and polyadenylated. Retrovirology 3:15

64. Halin M, Douceron E, Clerc I et al (2009) 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 114(12):2427–2438

65. Nishizawa T, Okamoto H, Konishi K, Yoshizawa H, Miyakawa Y, Mayumi M (1997) A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. Biochem Biophys Res Commun 241(1):92–97

66. Springfeld C, Bugert JJ, Schnitzler P, Tobiasch E, Kehm R, Darai G (2000) TT virus as a human pathogen: significance and problems. Virus Genes 20(1):35–45

67. Miyata H, Tsunoda H, Kazi A et al (1999) 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 73(5):3582–3586

68. Mushahwar IK, Erker JC, Muerhoff AS et al (1999) Molecular and biophysical characterization of TT virus: evidence for a new virus family infecting humans. Proc Natl Acad Sci U S A 96(6):3177–3182

69. Okamoto H, Nishizawa T, Tawara A et al (2000) TT virus mRNAs detected in the bone marrow cells from an infected individual. Biochem Biophys Res Commun 279(2):700–707

70. Kamahora T, Hino S, Miyata H (2000) Three spliced mRNAs of TT virus transcribed from a plasmid containing the entire genome in COS1 cells. J Virol 74(21):9980–9986

71. Qiu J, Kakkola L, Cheng F et al (2005) Human circovirus TT virus genotype 6 expresses six proteins following transfection of a full-length clone. J Virol 79(10):6505–6510

72. Takahashi K, Iwasa Y, Hijikata M, Mishiro S (2000) Identification of a new human DNA virus (TTV-like mini virus, TLMV) intermediately related to TT virus and chicken anemia virus. Arch Virol 145(5):979–993

73. Yang JD, Roberts LR (2010) Hepatocellular carcinoma: a global view. Nat Rev Gastroenterol Hepatol 7(8):448–458

74. Studach LL, Rakotomalala L, Wang WH et al (2009) Polo-like kinase 1 inhibition suppresses hepatitis B virus X protein-induced transformation in an in vitro model of liver cancer progression. Hepatology 50(2):414–423

75. Seeger C, Mason WS (2000) Hepatitis B virus biology. Microbiol Mol Biol Rev 64(1):51–68

76. Fang ZL, Sabin CA, Dong BQ et al (2009) 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 50(2):273–280

77. Su TS, Lai CJ, Huang JL et al (1989) Hepatitis B virus transcript produced by RNA splicing. J Virol 63(9):4011–4018

78. Suzuki T, Masui N, Kajino K, Saito I, Miyamura T (1989) 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 86(21):8422–8426

79. Chen PJ, Chen CR, Sung JL, Chen DS (1989) Identification of a doubly spliced viral transcript joining the separated domains for putative protease and reverse transcriptase of hepatitis B virus. J Virol 63(10):4165–4171

80. Suzuki T, Kajino K, Masui N, Saito I, Miyamura T (1990) Alternative splicing of hepatitis B virus RNAs in HepG2 cells transfected with the viral DNA. Virology 179(2):881–885

81. Sommer G, Heise T (2008) Posttranscriptional control of HBV gene expression. Front Biosci 13:5533–5547

82. Abraham TM, Lewellyn EB, Haines KM, Loeb DD (2008) Characterization of the contribution of spliced RNAs of hepatitis B virus to DNA synthesis in transfected cultures of Huh7 and HepG2 cells. Virology 379(1):30–37

83. Rosmorduc O, Petit MA, Pol S et al (1995) In vivo and in vitro expression of defective hepatitis B virus particles generated by spliced hepatitis B virus RNA. Hepatology 22(1):10–19

84. Lin X, Ma ZM, Yao X, Zhang YP, Wen YM (2002) Replication efficiency and sequence analysis of full-length hepatitis B virus isolates from hepatocellular carcinoma tissues. Int J Cancer 102(5):487–491

85. Lee GH, Wasser S, Lim SG (2008) Hepatitis B pregenomic RNA splicing—the products, the regulatory mechanisms and its biological significance. Virus Res 136(1–2):1–7

86. Huang HL, Jeng KS, Hu CP, Tsai CH, Lo SJ, Chang C (2000) Identification and characterization of a structural protein of hepatitis B virus: a polymerase and surface fusion protein encoded by a spliced RNA. Virology 275(2):398–410

87. Soussan P, Garreau F, Zylberberg H, Ferray C, Brechot C, Kremsdorf D (2000) In vivo expression of a new hepatitis B virus protein encoded by a spliced RNA. J Clin Invest 105(1):55–60

88. Assrir N, Soussan P, Kremsdorf D, Rossignol JM (2010) Role of the hepatitis B virus proteins in pro- and anti-apoptotic processes. Front Biosci 15:12–24

89. Berns KI (1990) Parvovirus replication. Microbiol Rev 54(3):316–329

90. Myers MW, Laughlin CA, Jay FT, Carter BJ (1980) Adenovirus helper function for growth of adeno-associated virus: effect of temperature-sensitive mutations in adenovirus early gene region 2. J Virol 35(1):65–75

91. Handa H, Carter BJ (1979) Adeno-associated virus DNA replication complexes in herpes simplex virus or adenovirus-infected cells. J Biol Chem 254(14):6603–6610

92. Geoffroy MC, Salvetti A (2005) Helper functions required for wild type and recombinant adeno-associated virus growth. Curr Gene Ther 5(3):265–271

93. Brown KE (2010) The expanding range of parvoviruses which infect humans. Rev Med Virol 20(4):231–244

94. Gao G, Vandenberghe LH, Alvira MR et al (2004) Clades of adeno-associated viruses are widely disseminated in human tissues. J Virol 78(12):6381–6388

95. Erles K, Rohde V, Thaele M, Roth S, Edler L, Schlehofer JR (2001) DNA of adeno-associated virus (AAV) in testicular tissue and in abnormal semen samples. Hum Reprod 16(11):2333–2337

96. Tobiasch E, Rabreau M, Geletneky K et al (1994) Detection of adeno-associated virus DNA in human genital tissue and in material from spontaneous abortion. J Med Virol 44(2):215–222

97. Blechacz B, Russell SJ (2004) Parvovirus vectors: use and optimisation in cancer gene therapy. Exp Rev Mol Med 6(16):1–24

98. Qiu J, Pintel DJ (2002) The adeno-associated virus type 2 Rep protein regulates RNA processing via interaction with the transcription template. Mol Cell Biol 22(11):3639–3652
 99. Weindler FW, Heilbronn R (1991) A subset of herpes simplex virus replication genes provides helper functions for productive adeno-associated virus replication. J Virol 65(5): 2476–2483
100. Qiu J, Nayak R, Tullis GE, Pintel DJ (2002) 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 76(24):12435–12447
101. Cossart YE, Field AM, Cant B, Widdows D (1975) Parvovirus-like particles in human sera. Lancet 1(7898):72–73
102. Candotti D, Etiz N, Parsyan A, Allain JP (2004) Identification and characterization of persistent human erythrovirus infection in blood donor samples. J Virol 78(22):12169–12178
103. Anderson MJ, Lewis E, Kidd IM, Hall SM, Cohen BJ (1984) An outbreak of erythema infectiosum associated with human parvovirus infection. J Hyg (Lond) 93(1):85–93
104. Woolf AD, Campion GV, Chishick A et al (1989) Clinical manifestations of human parvovirus B19 in adults. Arch Intern Med 149(5):1153–1156
105. Potter CG, Potter AC, Hatton CS et al (1987) Variation of erythroid and myeloid precursors in the marrow and peripheral blood of volunteer subjects infected with human parvovirus (B19). J Clin Invest 79(5):1486–1492
106. Jordan JA (1996) Identification of human parvovirus B19 infection in idiopathic nonimmune hydrops fetalis. Am J Obstet Gynecol 174(1 Pt 1):37–42
107. Ozawa K, Ayub J, Hao YS, Kurtzman G, Shimada T, Young N (1987) Novel transcription map for the B19 (human) pathogenic parvovirus. J Virol 61(8):2395–2406
108. Lynch JP III, Fishbein M, Echavarria M (2011) Adenovirus. Semin Respir Crit Care Med 32(4):494–511
109. Berk AJ (2005) Recent lessons in gene expression, cell cycle control, and cell biology from adenovirus. Oncogene 24(52):7673–7685
110. Akusjarvi G (2008) Temporal regulation of adenovirus major late alternative RNA splicing. Front Biosci 13:5006–5015
111. Stephens C, Harlow E (1987) Differential splicing yields novel adenovirus 5 E1A mRNAs that encode 30 kd and 35 kd proteins. EMBO J 6(7):2027–2035
112. Cuesta R, Xi Q, Schneider RJ (2000) Adenovirus-specific translation by displacement of kinase Mnk1 from cap-initiation complex eIF4F. EMBO J 19(13):3465–3474
113. Akusjarvi G, Stevenin J (2003) Remodelling of the host cell RNA splicing machinery during an adenovirus infection. Curr Top Microbiol Immunol 272:253–286
114. Akusjarvi G, Persson H (1981) Controls of RNA splicing and termination in the major late adenovirus transcription unit. Nature 292(5822):420–426
115. Zheng ZM (2010) Viral oncogenes, noncoding RNAs, and RNA splicing in human tumor viruses. Int J Biol Sci 6(7):730–755
116. Gardner SD, Field AM, Coleman DV, Hulme B (1971) New human papovavirus (B.K.) isolated from urine after renal transplantation. Lancet 1(7712):1253–1257
117. Padgett BL, Walker DL, ZuRhein GM, Eckroade RJ, Dessel BH (1971) Cultivation of papova-like virus from human brain with progressive multifocal leucoencephalopathy. Lancet 1(7712):1257–1260
118. Allander T, Andreasson K, Gupta S et al (2007) Identification of a third human polyomavirus. J Virol 81(8):4130–4136
119. Gaynor AM, Nissen MD, Whiley DM et al (2007) Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog 3(5):e64
120. Feng H, Shuda M, Chang Y, Moore PS (2008) Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319(5866):1096–1100
121. Schowalter RM, Pastrana DV, Pumphrey KA, Moyer AL, Buck CB (2010) Merkel cell polyomavirus and two previously unknown polyomaviruses are chronically shed from human skin. Cell Host Microbe 7(6):509–515

122. van der ME, Janssens RW, Lauber C, Bouwes Bavinck JN, Gorbalenya AE, Feltkamp MC (2010) Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. PLoS Pathog 6(7):e1001024

123. Scuda N, Hofmann J, Calvignac-Spencer S et al (2011) A novel human polyomavirus closely related to the African green monkey-derived lymphotropic polyomavirus. J Virol 85(9): 4586–4590

124. Sweet BH, Hilleman MR (1960) The vacuolating virus, S.V. 40. Proc Soc Exp Biol Med 105:420–427

125. Knowles WA, Pipkin P, Andrews N et al (2003) Population-based study of antibody to the human polyomaviruses BKV and JCV and the simian polyomavirus SV40. J Med Virol 71(1):115–123

126. Krumbholz A, Bininda-Emonds OR, Wutzler P, Zell R (2009) Phylogenetics, evolution, and medical importance of polyomaviruses. Infect Genet Evol 9(5):784–799

127. Shuda M, Feng H, Kwun HJ et al (2008) T antigen mutations are a human tumor-specific signature for Merkel cell polyomavirus. Proc Natl Acad Sci U S A 105(42):16272–16277

128. Shuda M, Kwun HJ, Feng H, Chang Y, Moore PS (2011) Human Merkel cell polyomavirus small T antigen is an oncoprotein targeting the 4E-BP1 translation regulator. J Clin Invest 121(9):3623–3634

129. Houben R, Adam C, Baeurle A et al (2012) 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 130(4):847–856

130. Zerrahn J, Knippschild U, Winkler T, Deppert W (1993) Independent expression of the transforming amino-terminal domain of SV40 large I antigen from an alternatively spliced third SV40 early mRNA. EMBO J 12(12):4739–4746

131. Abend JR, Joseph AE, Das D, Campbell-Cecen DB, Imperiale MJ (2009) A truncated T antigen expressed from an alternatively spliced BK virus early mRNA. J Gen Virol 90(Pt 5): 1238–1245

132. Trowbridge PW, Frisque RJ (1995) Identification of three new JC virus proteins generated by alternative splicing of the early viral mRNA. J Neurovirol 1(2):195–206

133. Zheng ZM, Baker CC (2006) Papillomavirus genome structure, expression, and post-transcriptional regulation. Front Biosci 11:2286–2302

134. Bernard HU, Burk RD, Chen Z, van DK, Hausen H, de Villiers EM (2010) Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. Virology 401(1):70–79

135. Munoz N, Bosch FX, de SS et al (2003) Epidemiologic classification of human papillomavirus types associated with cervical cancer. N Engl J Med 348(6):518–527

136. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S (2007) Human papillomavirus and cervical cancer. Lancet 370(9590):890–907

137. Chen HC, Schiffman M, Lin CY et al (2011) Persistence of type-specific human papillomavirus infection and increased long-term risk of cervical cancer. J Natl Cancer Inst 103(18):1387–1396

138. Scheffner M, Werness BA, Huibregtse JM, Levine AJ, Howley PM (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. Cell 63(6):1129–1136

139. Dyson N, Howley PM, Munger K, Harlow E (1989) The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. Science 243(4893):934–937

140. Tang S, Tao M, McCoy JP Jr, Zheng ZM (2006) 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 80(9):4249–4263

141. Wang X, Meyers C, Wang HK, Chow LT, Zheng ZM (2011) Construction of a full transcription map of human papillomavirus type 18 during productive viral infection. J Virol 85(16):8080–8092

142. Khan MJ, Castle PE, Lorincz AT et al (2005) 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 97(14):1072–1079

143. Stoler MH, Wright TC Jr, Sharma A, Apple R, Gutekunst K, Wright TL (2011) High-risk human papillomavirus testing in women with ASC-US cytology: results from the ATHENA HPV study. Am J Clin Pathol 135(3):468–475

144. Getman D, Aiyer A, Dockter J, Giachetti C, Zhang F, Ginocchio CC (2009) Efficiency of the APTIMA HPV Assay for detection of HPV RNA and DNA targets. J Clin Virol 45(Suppl 1):S49–S54

145. Ratnam S, Coutlee F, Fontaine D et al (2011) 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 49(2):557–564

146. Clad A, Reuschenbach M, Weinschenk J, Grote R, Rahmsdorf J, Freudenberg N (2011) 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 49(3):1071–1076

147. Molden T, Kraus I, Skomedal H, Nordstrom T, Karlsen F (2007) PreTect HPV-Proofer: real-time detection and typing of E6/E7 mRNA from carcinogenic human papillomaviruses. J Virol Methods 142(1–2):204–212

148. Benevolo M, Vocaturo A, Caraceni D et al (2011) 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 49(7):2643–2650

149. Ratnam S, Coutlee F, Fontaine D et al (2010) 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 48(8):2779–2785

150. Steitz J, Borah S, Cazalla D et al (2011) Noncoding RNPs of viral origin. Cold Spring Harb Perspect Biol 3(3):a005165

151. Cullen BR (2011) Herpesvirus microRNAs: phenotypes and functions. Curr Opin Virol 1(3):211–215

152. Plaisance-Bonstaff K, Renne R (2011) Viral miRNAs. Methods Mol Biol 721:43–66

153. Taylor GS, Blackbourn DJ (2011) Infectious agents in human cancers: lessons in immunity and immunomodulation from gammaherpesviruses EBV and KSHV. Cancer Lett 305(2):263–278

154. Zheng ZM (2003) Split genes and their expression in Kaposi's sarcoma-associated herpesvirus. Rev Med Virol 13(3):173–184

155. Arduino PG, Porter SR (2008) Herpes Simplex Virus Type 1 infection: overview on relevant clinico-pathological features. J Oral Pathol Med 37(2):107–121

156. Rajcani J, Andrea V, Ingeborg R (2004) Peculiarities of herpes simplex virus (HSV) transcription: an overview. Virus Genes 28(3):293–310

157. Javier RT, Stevens JG, Dissette VB, Wagner EK (1988) A herpes simplex virus transcript abundant in latently infected neurons is dispensable for establishment of the latent state. Virology 166(1):254–257

158. Croen KD, Ostrove JM, Dragovic LJ, Smialek JE, Straus SE (1987) 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 317(23):1427–1432

159. Farrell MJ, Dobson AT, Feldman LT (1991) Herpes simplex virus latency-associated transcript is a stable intron. Proc Natl Acad Sci U S A 88(3):790–794

160. Spivack JG, Woods GM, Fraser NW (1991) 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 65(12):6800–6810

161. Henderson G, Peng W, Jin L et al (2002) Regulation of caspase 8- and caspase 9-induced apoptosis by the herpes simplex virus type 1 latency-associated transcript. J Neurovirol 8(Suppl 2):103–111

162. Umbach JL, Kramer MF, Jurak I, Karnowski HW, Coen DM, Cullen BR (2008) MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. Nature 454(7205):780–783

163. vi-Rao GB, Goodart SA, Hecht LM, Rochford R, Rice MK, Wagner EK (1991) Relationship between polyadenylated and nonpolyadenylated herpes simplex virus type 1 latency-associated transcripts. J Virol 65(5):2179–2190

164. Chen J, Silverstein S (1992) Herpes simplex viruses with mutations in the gene encoding ICP0 are defective in gene expression. J Virol 66(5):2916–2927
165. Perry LJ, Rixon FJ, Everett RD, Frame MC, McGeoch DJ (1986) Characterization of the IE110 gene of herpes simplex virus type 1. J Gen Virol 67(Pt 11):2365–2380
166. Everett RD, Cross A, Orr A (1993) A truncated form of herpes simplex virus type 1 immediate-early protein Vmw110 is expressed in a cell type dependent manner. Virology 197(2):751–756
167. Spatz SJ, Nordby EC, Weber PC (1996) Mutational analysis of ICP0R, a transrepressor protein created by alternative splicing of the ICP0 gene of herpes simplex virus type 1. J Virol 70(11):7360–7370
168. Tang S, Bertke AS, Patel A, Wang K, Cohen JI, Krause PR (2008) 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 105(31):10931–10936
169. Berger JR, Houff S (2008) Neurological complications of herpes simplex virus type 2 infection. Arch Neurol 65(5):596–600
170. Desselberger U (1998) Herpes simplex virus infection in pregnancy: diagnosis and significance. Intervirology 41(4–5):185–190
171. Bankier AT, Beck S, Bohni R et al (1991) The DNA sequence of the human cytomegalovirus genome. DNA Seq 2(1):1–12
172. Rawlinson WD, Barrell BG (1993) Spliced transcripts of human cytomegalovirus. J Virol 67(9):5502–5513
173. Bresnahan WA, Shenk T (2000) A subset of viral transcripts packaged within human cytomegalovirus particles. Science 288(5475):2373–2376
174. Nelson PN, Rawal BK, Boriskin YS et al (1996) A polymerase chain reaction to detect a spliced late transcript of human cytomegalovirus in the blood of bone marrow transplant recipients. J Virol Methods 56(2):139–148
175. Gerna G, Revello MG, Percivalle E, Morini F (1992) Comparison of different immunostaining techniques and monoclonal antibodies to the lower matrix phosphoprotein (pp 65) for optimal quantitation of human cytomegalovirus antigenemia. J Clin Microbiol 30(5):1232–1237
176. Fernando S, Booth J, Boriskin Y et al (1994) Association of cytomegalovirus infection with post-transplantation cardiac rejection as studied using the polymerase chain reaction. J Med Virol 42(4):396–404
177. Greijer AE, Adriaanse HM, Kahl M et al (2001) Quantitative competitive NASBA for measuring mRNA expression levels of the immediate early 1, late pp 67, and immune evasion genes US3, US6 and US11 in cells infected with human cytomegalovirus. J Virol Methods 96(2):133–147
178. Randhawa PS, Manez R, Frye B, Ehrlich GD (1994) Circulating immediate-early mRNA in patients with cytomegalovirus infections after solid organ transplantation. J Infect Dis 170(5):1264–1267
179. Velzing J, Rothbarth PH, Kroes AC, Quint WG (1994) Detection of cytomegalovirus mRNA and DNA encoding the immediate early gene in peripheral blood leukocytes from immunocompromised patients. J Med Virol 42(2):164–169
180. Meyer T, Reischl U, Wolf H, Schuller C, Arndt R (1994) Identification of active cytomegalovirus infection by analysis of immediate-early, early and late transcripts in peripheral blood cells of immunodeficient patients. Mol Cell Probes 8(4):261–271
181. Boriskin YS, Fuller K, Powles RL et al (2002) 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 24(1–2):13–23
182. Andre E, Imbert-Marcille BM, Cantarovich D, Besse B, Ferre-Aubineau V, Billaudel S (1999) 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 34(4):287–291

183. Yao QY, Ogan P, Rowe M, Wood M, Rickinson AB (1989) Epstein-Barr virus-infected B cells persist in the circulation of acyclovir-treated virus carriers. Int J Cancer 43(1):67–71
184. Kutok JL, Wang F (2006) Spectrum of Epstein-Barr virus-associated diseases. Annu Rev Pathol 1:375–404
185. Qu L, Green M, Webber S, Reyes J, Ellis D, Rowe D (2000) Epstein-Barr virus gene expression in the peripheral blood of transplant recipients with persistent circulating virus loads. J Infect Dis 182(4):1013–1021
186. Young LS, Arrand JR, Murray PG (2007) EBV gene expression and regulation. In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (eds) Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, Cambridge. pp 461–489
187. Baer R, Bankier AT, Biggin MD et al (1984) DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310(5974):207–211
188. Swaminathan S (2008) Noncoding RNAs produced by oncogenic human herpesviruses. J Cell Physiol 216(2):321–326
189. Barth S, Meister G, Grasser FA (2011) EBV-encoded miRNAs. Biochim Biophys Acta 1809(11–12):631–640
190. Kennedy G, Sugden B (2003) EBNA-1, a bifunctional transcriptional activator. Mol Cell Biol 23(19):6901–6908
191. Woisetschlaeger M, Yandava CN, Furmanski LA, Strominger JL, Speck SH (1990) Promoter switching in Epstein-Barr virus during the initial stages of infection of B lymphocytes. Proc Natl Acad Sci U S A 87(5):1725–1729
192. Nonkwelo C, Ruf IK, Sample J (1997) The Epstein-Barr virus EBNA-1 promoter Qp requires an initiator-like element. J Virol 71(1):354–361
193. Nonkwelo C, Henson EB, Sample J (1995) Characterization of the Epstein-Barr virus Fp promoter. Virology 206(1):183–195
194. Miller G, El-Guindy A, Countryman J, Ye J, Gradoville L (2007) Lytic cycle switches of oncogenic human gammaherpesviruses. Adv Cancer Res 97:81–109
195. Manet E, Gruffat H, Trescol-Biemont MC et al (1989) Epstein-Barr virus bicistronic mRNAs generated by facultative splicing code for two transcriptional trans-activators. EMBO J 8(6):1819–1826
196. Furnari FB, Zacny V, Quinlivan EB, Kenney S, Pagano JS (1994) RAZ, an Epstein-Barr virus transdominant repressor that modulates the viral reactivation mechanism. J Virol 68(3): 1827–1836
197. Chang Y, Cesarman E, Pessin MS et al (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science 266(5192):1865–1869
198. Ganem D (2006) KSHV infection and the pathogenesis of Kaposi's sarcoma. Annu Rev Pathol 1:273–296
199. Ablashi DV, Chatlynne LG, Whitman JE Jr, Cesarman E (2002) Spectrum of Kaposi's sarcoma-associated herpesvirus, or human herpesvirus 8, diseases. Clin Microbiol Rev 15(3): 439–464
200. Russo JJ, Bohenzky RA, Chien MC et al (1996) Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus (HHV8). Proc Natl Acad Sci U S A 93(25):14862–14867
201. Verma SC, Lan K, Robertson E (2007) Structure and function of latency-associated nuclear antigen. Curr Top Microbiol Immunol 312:101–136
202. Dittmer D, Lagunoff M, Renne R, Staskus K, Haase A, Ganem D (1998) A cluster of latently expressed genes in Kaposi's sarcoma-associated herpesvirus. J Virol 72(10):8309–8315
203. Sun R, Lin SF, Gradoville L, Yuan Y, Zhu F, Miller G (1998) A viral gene that activates lytic cycle expression of Kaposi's sarcoma-associated herpesvirus. Proc Natl Acad Sci U S A 95(18):10866–10871
204. Lukac DM, Kirshner JR, Ganem D (1999) 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 73(11):9348–9361

205. Yamanegi K, Tang S, Zheng ZM (2005) 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 79(22):14207–14221

206. Majerciak V, Pripuzova N, McCoy JP, Gao SJ, Zheng ZM (2007) 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 81(3):1062–1071

207. Cheung TK, Guan Y, Ng SS et al (2005) 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 86(Pt 5):1447–1454

208. Purcell DF, Martin MA (1993) Alternative splicing of human immunodeficiency virus type 1 mRNA modulates viral protein expression, replication, and infectivity. J Virol 67(11):6365–6378

209. Yamano Y, Nagai M, Brennan M et al (2002) 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 99(1):88–94

210. Preiss S, Littlejohn M, Angus P et al (2008) Defective hepatitis B virus DNA is not associated with disease status but is reduced by polymerase mutations associated with drug resistance. Hepatology 48(3):741–749

211. Nguyen QT, Wong S, Heegaard ED, Brown KE (2002) Identification and characterization of a second novel human erythrovirus variant, A6. Virology 301(2):374–380

212. Gama-Carvalho M, Condado I, Carmo-Fonseca M (2003) Regulation of adenovirus alternative RNA splicing correlates with a reorganization of splicing factors in the nucleus. Exp Cell Res 289(1):77–85

213. White MK, Safak M, Khalili K (2009) Regulation of gene expression in primate polyomaviruses. J Virol 83(21):10846–10856

214. Tanaka S, Minagawa H, Toh Y, Liu Y, Mori R (1994) Analysis by RNA-PCR of latency and reactivation of herpes simplex virus in multiple neuronal tissues. J Gen Virol 75(Pt 10):2691–2698

215. Majerciak V, Yamanegi K, Allemand E, Kruhlak M, Krainer AR, Zheng ZM (2008) 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 82(6):2792–2801

216. Tang S, Zheng ZM (2002) Kaposi's sarcoma-associated herpesvirus K8 exon 3 contains three 5′-splice sites and harbors a K8.1 transcription start site. J Biol Chem 277(17): 14547–14556