Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Chapter 24

EMERGING VIRUSES: Their Diseases and Identification

Paul Kellam

INTRODUCTION

Diseases caused by microbial infections have been present throughout human evolution. A large proportion are the result of virus infections. Since the 1970s, new human viruses have been identified which cause widely varying diseases. These viruses fall into the well-documented field of emerging infectious diseases. In fact, three distinct situations are included in this category: (a) resurgent or recurrent old diseases (usually caused by 'new' or mutated previously known agents); (b) well-known human diseases with epidemiological evidence of transmission, but with the discovery of a newly identified agent; or (c) diseases truly new to humans but caused by pre-existing zoonotic agents.

EMERGING VIRUS DISEASES

A commonly cited example of resurgent or recurrent disease is the yearly appearance of new antigenically different influenza viruses. These new variants are able to evoke disease in their host while causing the centuries-old symptoms of influenza. The variation that occurs in influenza viruses is of two kinds. The first, called antigenic drift, occurs as a result of accumulation of point mutations in the virus surface proteins, haemagglutinin (H) and neuraminidase (N). This can lead to viral escape from the immune response and the emergence of new influenza epidemics. The second kind of change, called antigenic shift, occurs when complete genome segments encoding the H or N genes of one virus variant are replaced by the corresponding segments from another variant. This process is known as reassortment. Both wildfowl and swine are endemically infected with influenza virus, and, periodically, avian and human influenza viruses co-infect pigs. It is therefore thought that pigs act as a mixing pot in the formation of pandemic influenza variants by allowing reassortment between avian, swine and human influenza viruses to occur. However, the 1918 'Spanish' pandemic strain of influenza was most likely derived solely from strains that infected humans and swine, not the avian subgroup (Taubenberger et al, 1997). This suggests that close observation of new pig influenza viruses could highlight the presence of potential new human pandemic variants.

HIV and the New Viruses Second Edition
ISBN 0-12-200741-7

New infectious diseases falling into the latter two categories – new disease or new agent – continue to be identified. Table 24.1 lists the major human viral diseases identified since 1977. This table also includes new human viruses for which a link with a specific human disease has not been conclusively established, for example hepatitis G virus (Simons et al, 1995a,b; Linnen et al, 1996), Borna disease virus (Lipkin et al, 1990; VandeWoude et al, 1990) and human herpesvirus 7 (Frenkel et al, 1990). Also included are subgenomic viral sequences identified in human tissue, namely human retrovirus 5 (HRV-5) (Griffiths et al, 1997) and multiple sclerosis-associated retrovirus (MSRV) (Perron et al, 1997), which have not yet yielded complete virus genomes.

Contributors to this book have described many viral diseases truly new in humans, but most probably caused by a zoonosis. Human immunodeficiency virus types 1 and 2 (HIV-1, HIV-2) have homology to Old World monkey virus, simian immunodeficiency virus (SIV), suggesting that a recent zoonosis resulted in the appearance of the new human disease of acquired immune deficiency syndrome (AIDS) (Allan et al, 1991; Myers et al, 1992). Another example of a new zoonotic infection, hantavirus pulmonary

Table 24.1 ***Viruses associated with human disease identified since 1977***

Year	*Agent*	*Disease*	*References*
1977	Ebola virus	Ebola haemorrhagic fever	Johnson et al (1977)
1977	Hantaan virus	Haemorrhagic fever with renal syndrome	Lee et al (1978)
1980	Human T cell lymphotropic virus type 1 (HTLV-1)	T cell lymphoma/leukaemia	Poiesz et al (1980)
1982	Human T cell lymphotropic virus type 2 (HTLV-2)	Hairy cell leukaemia	Kalyanaraman et al (1981)
1983	Human papillomavirus types 16 and 18	Cervical cancer	Durst et al (1983)
1983	Human immunodeficiency virus type 1 (HIV-1)	AIDS	Barre Sinoussi et al (1983)
1986	Human immunodeficiency virus type 2 (HIV-2)	AIDS	Clavel et al (1986)
1988	Human herpesvirus 6 (HHV-6)	Roseola subitum	Yamanishi et al (1988)
1989	Hepatitis C virus (HCV)	Non-A, non-B hepatitis	Choo et al (1989)
1990	Borna disease virus	NAD	VandeWoude et al (1990) Lipkin et al (1990)
1990	Hepatitis E virus (HEV)	Acute faecal–oral hepatitis	Reyes et al (1990)
1990	Human herpesvirus 7 (HHV-7)	NAD	Frenkel et al (1990)
1991	Guantanto virus	Venezuelan haemorrhagic fever	Salas et al (1991)
1993	Hantavirus (SNV)	Hantavirus pulmonary syndrome (HPS)	Nichol et al (1993)
1994	Sabia virus	Brazilian haemorrhagic fever	Lisieux et al (1994)
1994	Human herpesvirus 8 (HHV-8)	Kaposi's sarcoma	Chang et al (1994)
1995	Hepatitis G virus (HGV, GBV-C)	NAD	Linnen et al (1996) Simons et al (1995a,b)
1997	Multiple sclerosis-associated retrovirus	Multiple sclerosis	Perron et al (1997)
1997	Human retrovirus 5 (HRV-5)	NAD	Griffiths et al (1997)

*NAD, no associated human disease.

syndrome (HPS), demonstrates how modern molecular virology can be instrumental in identifying new human pathogens. In May 1993 an outbreak of respiratory illness with a mortality rate in excess of 75% was reported in a border region of New Mexico, Arizona, Utah and Colorado in southwestern USA. Serology surveys initially failed to identify known agents associated with respiratory disease but did detect cross-reactive antibodies to a previously characterized hantavirus. This was thought unusual, as hantaviruses had not previously been associated with human disease in North America, nor had they been associated with a severe predominantly respiratory illness. Using degenerate PCR primers to conserved regions of the hantavirus genome led to the identification of a new hantavirus that differed from known hantaviruses by 30% at the nucleotide level (Nichol et al, 1993). The usual hosts for hantaviruses are rodents, and trapping of rodents in the Four Corners area revealed 33% of deer mice were seropositive for the new hantavirus. The outbreak of HPS was subsequently shown to be associated with a 10-fold increase in the deer mouse population in the Four Corners area. Four months after the HPS outbreak a culture system was established and the hantavirus responsible for HPS was finally designated Sin Nombre virus (SNV) (Elliott et al, 1994).

An example of a newly identified virus causing disease in humans with no currently known animal or insect host is provided by hepatitis C virus (HCV). Epidemiological evidence suggested that the causes of non-A, non-B hepatitis (NANBH) were of an infectious origin. However, conventional virological techniques had failed to identify the agent, even though evidence suggested NANBH was caused by a blood-borne, small enveloped virus, readily transmissible to chimpanzees. The hepatitis C virus was discovered by screening a cDNA expression library prepared from virus particles pelleted from a chimpanzee with a high NANBH virus titre. When the library was screened with serum from a patient with chronic NANBH one clone was identified from a library of one million clones (Choo et al, 1989). Identification of this clone led to the rapid characterization of the entire virus genome, and HCV was formally assigned to a new virus subfamily within the family *Flaviviridae*. Serology assays have now demonstrated that HCV is the major cause of NANBH throughout the world (Kuo et al, 1989; Linnen et al, 1996).

These examples show clearly how modern molecular biology techniques can be used to identify completely new viruses. These viruses can be associated with a new disease, or associated with a well-characterized disease present in humans for many years. In identifying an emerging virus, one is often presented with epidemiological data and clinical specimens that have no reactivity with diagnostic reagents available for known pathogens. The primary aim therefore is to identify any new infectious agent and build a body of data to support the existence of a causal link between organism and disease.

IDENTIFICATION OF EMERGING VIRUSES

Virus isolation and culture in vitro have long been the definitive method of new virus characterization and this is often the first method employed. However, in many viral diseases this approach may have several limitations. The main limitation is the need for a permissive host cell to propagate the new virus. Equally important is an effective detection method to show infection of the permissive cells. Classically, this was achieved

through the observation of cytopathic effect or the detection of specific viral gene products. Isolation in cell culture will not work if the virus does not grow in vitro, or if no detectable cytopathic effect or viral gene product is produced. Indeed, in the case of HPS, discovery of a new hantavirus preceded the establishment of a tissue culture system for virus propagation (Elliott et al, 1994). For HCV, good culture systems still do not exist 10 years after the discovery of the virus, and keratinocytes permissive for human papillomavirus type16 replication are still not suitable for routine virus isolation 16 years after its discovery by molecular hybridization (Durst et al, 1983).

Electron microscopy (EM) of disease tissue and in vitro infected cell cultures is another method employed. It may reveal virus particles and common morphological features of a particular virus family, leading to clues for the further characterization of the virus. However, EM can suffer from artifactual structures present in the prepared material. In the late 1960s and throughout the 1970s, a number of groups reported the detection by EM of virus-like particles in human tumours (Dmochowski et al, 1968; Feller and Chopra, 1971; Seman et al, 1971; Denner and Dorfman, 1977). In contrast, Dalton (1975) demonstrated that many of these virus-like particles may simply be the result of breakdown products from normal cells. Also, the assumption that virus architecture is constant in a family of viruses is not always true, with viruses classified by EM requiring reclassification following molecular analysis of genome structure. One example of this is the *Arterivirus* genus of the coronavirus-like (CVL) superfamily. Equine anaemia virus (EAV – the prototype of the genus), lactate dehydrogenase-elevating virus (LDV), simian haemorrhagic fever virus (SHFV) and porcine reproductive and respiratory syndrome virus (PRRSV) are morphologically similar to viruses of the *Togaviridae*, and together with their genome size this resulted in their initial classification in the togavirus family. However, the arterivirus genome organization and replication strategy, along with the homology of proteins such as the RNA replicase, have since resulted in the reclassification of arteriviruses within the CVL superfamily. The final taxonomic fate of this superfamily is still to be decided (Snijder and Horzinek, 1993; Snijder et al, 1993). Electron microscopy can be successful when a completely new structure is observed, as with the observation of the worm-like virions of Marburg virus. This virus was later classified as the type species of the *Filovirus* family (Kiley et al, 1982).

Other standard techniques based on immunofluorescence assays use patient sera to detect viral antigens in infected pathological specimens. This was used successfully to detect SNV in hantavirus pulmonary syndrome. However, if the patient sera or clinical specimens do not react with known virus groups or virus-specific antibodies, this method provides little useful information. The tendency to use increasing specificity in routine diagnostic tests works against identification of related yet distinct virus strains.

A combination of immunological and molecular biological techniques has been used successfully to identify new human viruses. A general scheme for immunoscreening is outlined in Figure 24.1A. This relies on the production of high-quality cDNA expression libraries derived from infected tissue or, in the case of positive-sense RNA viruses, directly from virus present in plasma. Three hepatitis viruses have been identified using this technique. Hepatitis C virus was identified by Choo et al (1989) following the isolation of a virus-specific cDNA clone from NANB-infected chimpanzee plasma as described earlier. The same approach was successful in identifying hepatitis G virus (Linnen et al, 1996); this virus was found to be almost identical to the virus GBV-C, another human hepatitis-associated virus (Simons et al, 1995b), identified using degen-

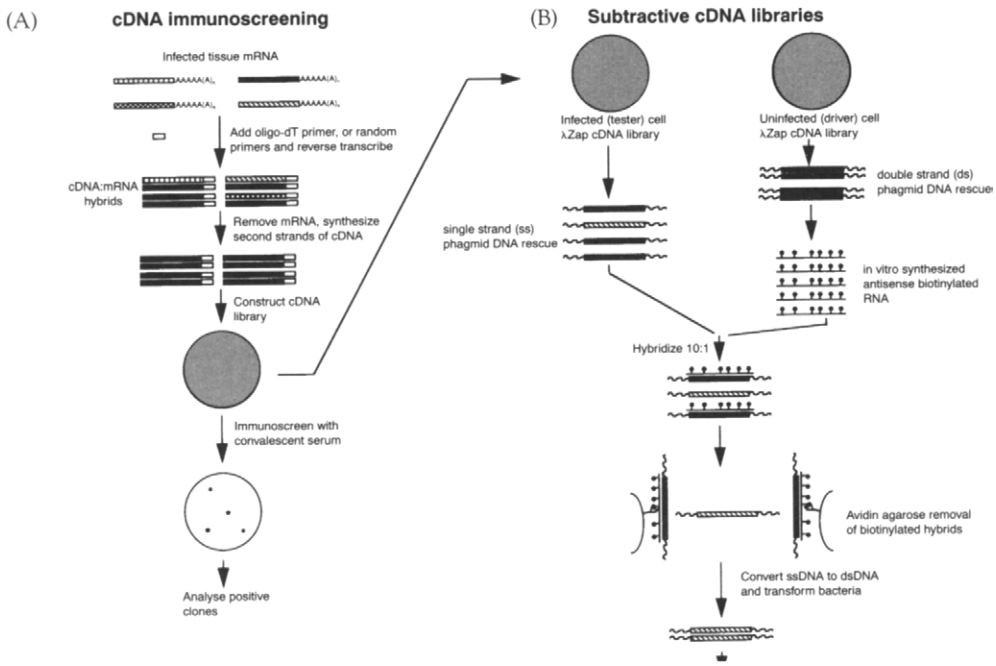

Figure 24.1 ***The cDNA library approach to identifying new viruses***
Different cDNA library approaches to the identification of unculturable infectious agents. (A) Using cDNA immunoscreening. Modified from Gao and Moore (1996) with permission. (B) Using subtractive cDNA libraries, based on the method of Schweinfest et al (1990). Filled long boxes represent cDNA.

erate PCR primers. In addition hepatitis E virus (HEV)-specific cDNA clones were identified from a cDNA library derived from total RNA isolated from the bile juice of two experimentally infected rhesus monkeys (Uchida et al, 1992). An absolute requirement for cDNA immunoscreening is the availability of high antibody titre immune serum from patients or experimentally infected animals. However, these sera may also contain many antibodies that cross-react with human antigens expressed in the library, leading to false-positive clones being identified. This can be a particular problem for diseases in which autoantibodies are common. The requirement for a large representative cDNA library made in the absence of knowledge of the virus genome entails the use of randomly primed cDNA rather than conventional oligo-dT primed libraries. This was particularly relevant in the case of HCV where the viral positive-sense RNA genome contains a polyuridine rather than a polyadenine 3′ tail (Kolykhalov et al, 1996).

SEQUENCE-BASED METHODS OF IDENTIFYING NEW VIRUSES

The use of powerful molecular-based methods to analyse well-characterized biological specimens has created a new era of molecular identification of emerging viruses. The major advantage of molecular methodologies is the ability to look rapidly for new viruses, known viruses, or related but previously undetected members of established

virus families. This was resoundingly successful for the discovery of genital papillomaviruses (Durst et al, 1983).

With the widespread use of the polymerase chain reaction (PCR) and the availability of extensive sequence databases of virus genomes it is often possible to design PCR primers to conserved regions of virus genomes. It is then possible to survey samples with these primers to look for the presence of a given virus in new pathological specimens. These are very powerful identification tools but as such must be applied with care. The most common problems encountered are the nature of the assay conditions, where in addition to the general problems of PCR contamination, small modifications can dramatically alter the sensitivity of the PCR signal produced. Without appropriate controls, conditions can be accidentally contrived to amplify and detect many irrelevant DNA sequences from a disease. In addition, detection of virus genomes in disease tissue does not automatically produce a link between virus and disease (Fredericks and Relman, 1996). This has particularly relevance to PCR screening for known or new viruses. For example, a number of known infectious agents have been implicated in the pathogenesis of multiple sclerosis (MS) (Allen and Brankin, 1993; Kurtzke, 1993; Challoner et al, 1995; Perron et al, 1997), but none of these associations has been conclusively proved to cause MS (Rice, 1992).

To look for new members of virus families, degenerate PCR primers can be designed to conserved regions of virus genomes at the nucleic acid level or, preferably, can be designed to conserved amino acid regions of virus proteins based on codon degeneracy for each amino acid. The latter cover a much wider variety of nucleic acid sequences and are therefore more divergent when used to look for new virus genomes. Many groups have published degenerate PCR primers to diverse virus families (Table 24.2). These have been used successfully to identify many new viruses ranging from human viruses such as GBV-C (hepatitis G virus) (Simons et al, 1995a), SNV (Nichol et al, 1993) and HRV-5 (Griffiths et al, 1997), to many animal viruses, pig endogenous retrovirus (PERV) (Patience et al, 1997), walleye dermal sarcoma virus (Zhang and Martineau, 1996) and a macaque gammaherpesvirus, retroperitoneal fibromatosis herpesvirus (RFHV Mn) (Rose et al, 1997). However, the lack of specific controls for all degenerate primers means standardization can only be achieved on existing members of virus families, thereby not guaranteeing detection of an unknown virus. Again, the problems of inappropriate amplification conditions can lead to the production of many false positives or, in some cases, no amplification. It is therefore necessary to optimize carefully degenerate PCR conditions on relevant controls. This usually requires optimization of reaction buffers and primer annealing temperatures, and often the use of nested or seminested PCR strategies. In addition, special amplification techniques such as 'Touchdown' PCR can often improve PCR amplification when using degenerate primers (Don et al, 1991; Zhang and Martineau, 1996).

The most recent array of molecular techniques to be adapted for use in new virus discovery are purely nucleic acid-based and make no assumptions about the nature of the viral agent present. These methods have evolved from one aim of molecular medicine, which is to identify differences at the nucleic acid level between disease-associated tissue and normal tissue. These methodologies have in common that one nucleic acid population ('uninfected' or 'driver') is hybridized in excess with a second population ('infected' or 'tested') to remove common sequences, thereby enriching target sequences unique to the tester. The methods can be broadly divided into physical subtraction tech-

niques and PCR-based kinetic enrichment techniques. Each has its own relative merits which are ultimately dependent on the nature of the nucleic acid sample to be analysed.

Physical subtraction techniques

Physical subtraction techniques are applicable only to detecting differences in mRNA expression between one cell type and another. Early uses of these techniques simply involved the solution hybridization of an excess of driver mRNA with cDNA made from tester mRNA. Common sequences present in both samples form cDNA:mRNA hybrids, leaving the unique sequences in the tester cDNA unhybridized. The double-stranded hybrids are removed by hydroxyapatite chromatography, exploiting the higher affinity of hydroxyapatite for double-stranded nucleic acids. The remaining subtracted cDNA is used to construct a subtracted cDNA library or, if labelled, as an enriched cDNA probe for library screening. This method was used in combination with immunoscreening to identify the virus associated with Borna disease (Lipkin et al, 1990; VandeWoude et al, 1990). Borna disease is an infectious neurological disease that occurs sporadically in horses and sheep in central Europe, and may also be associated with certain human neuropsychiatric disorders (Bode et al, 1995).

Modern versions of this technique have been developed that use cDNA libraries constructed in the cloning vector λZap (Stratagene, La Jolla, USA). The libraries are made from driver and tester cell lines or tissue samples (Figure 24.1B). This method has been used to isolate rare cDNAs (less than 0.01% abundance) from colon and hepatic carcinoma tissue (Schweinfest et al, 1990). The λZap vectors allow the rescue, directly from the library, of single-stranded DNA phagemids and DNA (ssDNA) or double-stranded DNA plasmids (dsDNA), all containing representative cDNAs. Initial methods made use of non-directionally cloned cDNA libraries (Schweinfest et al, 1990). However, newer vectors allowing the production of directionally cloned cDNAs can be used in a modified version of this technique (Figure 24.1B).

Both primary cDNA libraries are amplified and approximately 2 million plaque-forming units of the tester library are used to produce single-strand phage DNA. The same amount of driver library is used to produce double-stranded plasmid DNA. In vitro transcribed RNA incorporating dUTP-biotin is made from the driver dsDNA using a T3 promoter located in the vector. These biotinylated RNAs are complementary to the tester ssDNA and are subsequently hybridized to the tester ssDNA. The complementary hybrids are removed using streptavidin beads. Multiple rounds of hybridization and subtraction can be performed, enriching for rare cDNAs in the tester population. Following subtraction the tester ssDNA is made double-stranded and transformed into *Escherichia coli*. This represents the enriched subtracted library which can then be analysed further.

PCR-based techniques

Representational difference analysis

Representational difference analysis (RDA) represented the first global approach to the analysis of differences between cellular genomes. Although originally developed to look for differences between tumour cell genomic DNA and normal cells (Lisitsyn et al,

Table 24.2 ***Degenerate PCR primers***

Virus family	*Gene*	*Primer name*	*Primary PCR primers (5′ to 3′)**	*Nested/seminested PCR primers (5′ to 3′)*	*Reference*
Retrovirus	RT	5′MOP-1	TGGAAAGTGYTRCCMCARGG		Shih et al (1989)
		3′MOP-1	GGMGGCCAGCAGSAKGTCATCCAYGTA		
Retrovirus	RT	5′MOP-2	CCWTGGAATACTCCYRTWTT		Li et al (1996)
		3′MOP-2	GTCKGAACCAATTWATATYYCC		
		5′MOP-1		TGGAAAGTGYTRCCMCARGG	
		3′MOP-2		GGMGGCCAGCAGSAKGTCATC CAYGTA	
Retrovirus (A, B, D type)	RTing and protease	ABDPOL	TCCCCTTGGAATACTCCTGTTTTYGT		Medstrand et al (1992)
		ABDPRO	CATTCCTTGTGGTAAAACTTTCCAYTG		
Retrovirus (BLV, HTLV)	Integrase	110(+)	CCCTACAATCCCACMAGCYTCRG		Dube et al (1997)
		111(−)	RTGGTKATTTSCCATCKGGTYTT		
Retrovirus (lentiviruses)	RT	LV1	CCGGATCCDCAYCCNGSAGGAYTAMAA		Gelman et al (1992)
		LV2	GGTCTAGAYRYARTTCATAACCCAKCCA		
		LV3		CCGGATCCGAYRTRGGKGAYGCMTA	
		DDMY		CCGGATCCRTCRTCCATRTA	
Herpesvirus	Pol[a]	DFA	GAYTTYGCNAGYYTNTAYCC		VanDevanter et al (1996)
		ILK	TCCTGGACAAGCARNYSGCNMTNAA		
		KG1	GTCTTGCTCACCAGNTCNACNCCYTT		
		TGV		TGTAACTCGGTGTAYGGNTTYACNG GNGT	
		IYG		CACAGAGTCCGTRTCNCCRTADAT	
Herpesvirus	Pol[b]	DFASA	GTGTTCGACTTYGCNAGYYTNTAYCC		Rose et al (1997)
		GDTD1B	CGGCATGCGACAAACACGGAGTCNG TRTCNCCRTA		
		VYGA		ACGTGCAACGCGGTGTAYGGNKTNA CNGG	
		PCLNA		GTCGCCTCTGGCATCCTNCCNTGYC TNAA	
		KMLEA		CAGGGCCGGAAGATGCTGGARACRT CNCARGC	
Papillomavirus	L1[c]	GP17	CGGGATCCGGNMGNGGNCARCCNY TNGG		Snijders et al (1991)
		GP18	CGGGATCCAYNCCRTTRTTRTGNCCYTG		
		GPR22		ARGAYGGNGAYATGRYNGAYAYNGG NTWYGG	
Coronavirus	S gene	55	GGAKAAGGTKAATGARTGYGT		Tobler and Ackermann (1996)
		56	CCAKACVTACCAWGGCCAYTT		

Hantavirus	G2[d]	+2548	GATATGAATGATTGYTTTGT		Nichol et al (1993)
		−2859	CCATCAGGGTCTYTCCA		
		+2590		TGTATAATTGGGACWGTATCTAA	
		−2751		GCAAAGTTACATTTYTTCCT	
		+2671	TTTAAGCAATGGTGYACTACWAC		
		−3108	CCATAACACATWGCAGC		
		+2770		AGAAAGAAATGTGCATTTGC	
		−3012		CCTGAACCCCATGCHCCATC	
Morbillivirus	P gene[c]	UPPF	ATGTTTATGATCACAGCGGT		Barrett et al (1993)
		UPPR	ATTGGGTTGCACCACTTGTC		Shimizu et al (1994)
		MBV1	TATGCTGGGTGAAAGTAAGATCT		
		MBV3	GGATTGCTGAAATGATTTGTGAT		
		MBV2		TTTGGCATTGAAACTATGTATCC	
Caliciviruses	orf1[f]	NVp110	ACDATYTCATCATCACCATA		Le Guyader et al (1996)
		NVp35	CTTGTTGGTTTGAGGCCATAT		
		NVp36	ATAAAAGTTGGCATGAACA		
		NVp69	GGCCTGCCATCTGGATTGCC		
		NI	GAATTCCATCGCCCACTGGCT		
Hepatitis C viruses[g]	NS3 helicase	5′	TYGCYACKGCKACCCCHCCKG		Simons et al (1995a)
		3′	TGCCMGCTYTCCCMCKGCC		
		3′		CRATRGTRAWRGTMGGGTCMAGG	
Flavivirus	NS5	FG1	TCAAGGAACTCCACACATGAGATG TACT		Fulop et al (1993)
		FG2	TGTATGCTGACACAGCAGGATGGG ACAC		

*R = A or G; Y = C or T; M = A or C; S = C or G; W = A or T; K = G or T; D = G, T or A; H = A, C or T; V = A, C or G; N = all four nucleotides.

[a]5′ primers DFA and ILK are used in combination with 3′ primer KG1.

[b] For pan-herpesvirus degenerate PCR, primary PCR is performed with primers DFSA and GDTD1B followed by seminested PCR with primers VYGA and GDTD1B. For gammaherpesvirus degenerate PCR, primary PCR is performed with primers VYGA and GDTD1B followed by two seminested PCRs with PCLNA and GDTD1B followed by primers KMLEA and GDTD1B.

[c]Oligonucleotide GPR22 is used as a radioactive probe to detect PCR products.

[d]Nested primer set +2548, −2859, +2590, −2751 detects Hantaan and Seoul serotype viruses and primer set +2671, −3108, +2770, −3012 detects Hantaan and Puumala

1993; Lisitsyn and Wigler, 1995), it has been adapted to identify differences in mRNA expression (Hubank and Schatz, 1994). The technique has been used successfully to identify new human viruses associated with Kaposi's sarcoma, namely human herpesvirus 8 (Chang et al, 1994) and the viruses GBV-A and B (Simons et al, 1995b), which are related to GBV-C/HGV (Simons et al, 1995a; Linnen et al, 1996), associated with viral hepatitis. More controversially, RDA has been used to identify human herpesvirus 6 sequences in plaques from patients with multiple sclerosis (Challoner et al, 1995) and cDNA clones from infectious Creutzfeldt–Jakob brain fractions that have no homology to any known database sequences (Dron and Manuelidis, 1996).

Representational difference analysis combines three elements: representation, subtractive enrichment and kinetic enrichment. The procedure is carried out in two stages. The first comprises the preparation of representations for driver and tester DNAs or cDNAs. Small restriction fragments derived from the starting nucleic acid are ligated to oligonucleotide adaptors and amplified by PCR. The second stage comprises the reiterative hybridization/selection steps (Figure 24.2A). Prior to hybridization, the oligonucleotide adaptors used for the initial representation PCR step are cleaved from both driver and tester amplicons and a new set of defined but different sequence adaptors is ligated onto the 5′ ends of only the tester amplicons. After hybridization of tester and driver amplicons the mixture of molecules is treated with DNA polymerase. This adds a copy of the defined oligonucleotide to both 3′ ends of only the self-annealed tester DNA fragments. The defined oligonucleotide adaptor/primer is then used during PCR of the mixture such that only the tester-annealed DNA fragments can participate in exponential amplification to yield a difference product. The cycle of cleavage of old adaptors and ligating new adaptors to the difference products, hybridizing with excess driver and PCR amplification is repeated two or three times and the final PCR products are cloned and analysed further.

Suppression subtraction hybridization

Suppression subtraction hybridization (SSH), based on similar principles to RDA, was described in June 1996 (Diatchenko et al, 1996). This technique is designed selectively to amplify differentially expressed cDNA fragments and simultaneously to suppress non-targeted DNA amplification. Like RDA, it can be used on a variety of nucleic acid targets (Figure 24.2B), and in both methods representations of both tester and driver DNAs are prepared by restriction enzyme digest. Tester DNA is then subdivided into two portions and each is ligated with a different oligonucleotide adaptor to the 5′ end. However, in contrast to RDA, SSH involves two hybridization steps. In the first, an excess of denatured driver cDNA is added to each population of denatured tester cDNA. Owing to second-order reaction kinetics of hybridization, single-stranded molecules corresponding to high- and low-abundance sequences become normalized. Normalization occurs because the annealing process is faster for more abundant molecules and results in a proportion of the low-abundance tester cDNA remaining single-stranded. During the second hybridization the two primary hybrid samples are mixed together. Since the samples are not heat-denatured, only the remaining normalized and subtracted single-strand tester cDNAs are able to associate to form new hybrids. These hybrids have different oligonucleotide adaptor sequences at their 5′ ends. Fresh denatured driver DNA is then added to enrich further for the differentially expressed sequences. Following hybridization and DNA polymerase end filling, the entire population is PCR amplified

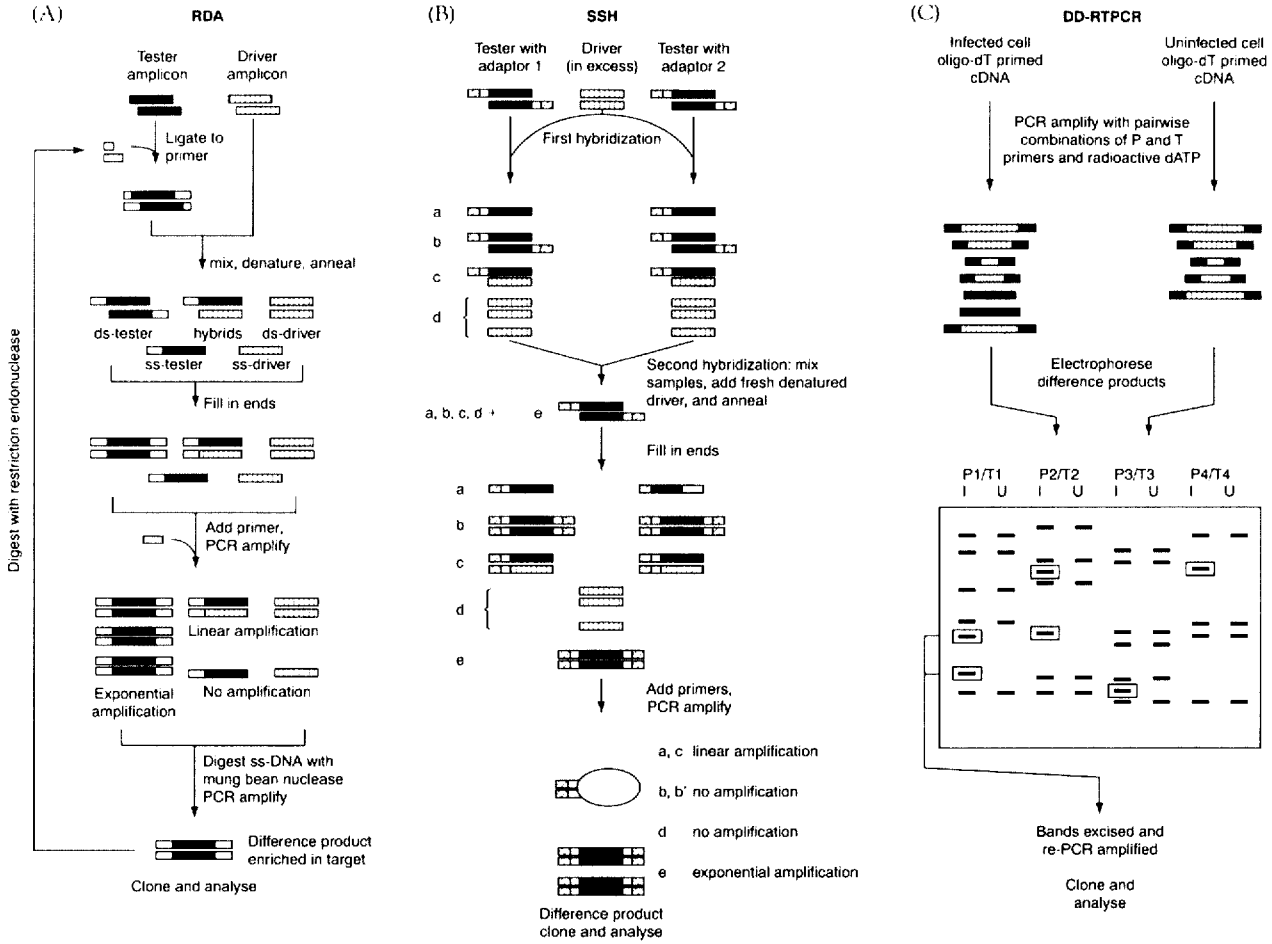

Figure 24.2 ***The PCR approach to identifying new viruses***
Different sequence-based PCR methods used to identify unculturable infectious agents. (A) Representation difference analysis (RDA). Modified from Gao and Moore (1996) with permission. (B) Suppression subtraction hybridization (SSH). Modified from Diatchenko et al (1996) with permission. (C) Differential display RT PCR (DD-RTPCR). Long boxes represent DNA or cDNA, small boxes represent oligonucleotide primers. Note that in (B) two types of oligonucleotide adaptors are used, represented by different shading. Each set of adaptors has a common sequence shown by an open box corresponding to nested PCR primer sites. In (C), 'T primers' refer to composite oligo-dT primers and 'P primers' refer to composite 5' primers.

with oligonucleotides for both adaptors. Hybrids with the same adaptor at either end are suppressed from PCR amplification owing to the formation of panhandle structures between the terminally repeated primer sites (Diatchenko et al, 1996). Only hybrids with different adaptors at either end are exponentially amplified. A final nested PCR with internal common primers allows the cloning of the difference products and analysis. This method has not yet been used for virus identification, but like RDA it has enormous potential.

Differential display

Differential display reverse transcription PCR (DD-RTPCR), described by Liang and Pardee (1992), is only applicable to differences in mRNA expressed between cells and is essentially semirandomly primed PCR. This method is based on the assumption that every individual mRNA molecule can be reverse transcribed and amplified by PCR. Using the original principles described, DD-RTPCR has now been further developed and refined (Ayala et al, 1995) (Figure 24.2C).

Four composite 3′ primers containing a 10 base common sequence at the 5′ end followed by polyT and ending with a mixed base (A, C or G) plus a 3′ fixed base were designed. One such improved primer set is illustrated in Figure 24.3 (Ayala et al, 1995).

These primers are used to prime first strand cDNA synthesis of the driver and tester mRNA or total RNA. In addition, composite arbitrary primers containing the sequence CGTGAATTCG added to the 5′ end of different random 10-mer sequences were also designed. These are then used with the 3′ primers in a radioactively labelled PCR reaction involving initial low-stringency cycles followed by high-stringency cycles. The PCR products are resolved on either native or denaturing polyacrylamide gels. Differentially expressed or novel mRNAs are represented by specific bands in the tester tracks of the gels compared with the driver. The greater the number of different 5′ arbitrary primers used, the higher the chance of observing differences. However, as each new 5′ primer produces eight PCR reactions (four for tester and four for driver), numbers can become limiting. Any differentially displayed band is excised from the original gel, reamplified, cloned and studied further. Like SSH, DD-RTPCR has yet to result in the identification of a new virus.

NEW VIRUS DISCOVERY

The right combination of difference methods needs to be used for the variety of samples that are likely to arise in new virus discovery. If the biological samples available are cell-free, the choice of methods is limited. With no knowledge of the nature of the viral genome (i.e. DNA, positive-sense RNA or negative-sense RNA), only degenerate PCR, RDA or SSH methods are applicable. For infected cell lines or tissue samples, all methods are applicable. Again, if the nature of a new viral genome is not known, the most

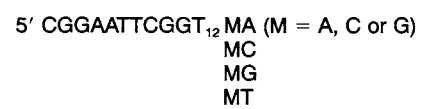

Figure 24.3 ***Improved primer set for DD-RTPCR***

global method of analysis would be to study new mRNA species. The use of physical subtraction methods requires a relatively abundant supply of polyA+ mRNA which may limit the technique. However, since this method avoids PCR and thus PCR contamination, a problem in all of the PCR-based methods, it is still very useful. Another disadvantage of the physical subtraction method is the requirement for cDNA libraries, which are costly and time-consuming to make.

Both RDA and SSH have the disadvantages of requiring complex restriction digests, primer ligations, hybridization and PCR. Contamination with common laboratory DNA at the initial stages can lead to false positive clones. In addition, failure at any one stage can lead to no difference PCR products or false-positives which are impossible to control for when working with authentic samples. The RDA method is time-consuming compared with SSH, and SSH also requires less initial tester and driver nucleic acid than RDA. However, like DD-RTPCR, SSH has yet to be proved in the field of new virus discovery.

Differential display is very sensitive and prone to false positives, although no more so than other subtractive methods (Wan et al, 1996). The method also has the disadvantage of primer dependency for PCR amplification. Owing to the use of polyT primers DD-RTPCR does not allow detection of positive-sense virus genomes with alternative tails to polyA. However, it is possible to design and use alternative polynucleotide primers to cover such diverse virus groups. Overall, DD-RTPCR is quick and provides a qualitative result of the differences between two samples.

Following the identification of a candidate clone, further analysis to characterize whether the candidate is truly exogenous in origin is required. Northern and Southern blot analysis and/or PCR should reveal whether a clone is exogenous. Analysis of open reading frames should show whether the clone encodes for any known homologous protein and whether these are related to known viruses. However, as with many major advances in technology, sequenced-based approaches for new virus identification can lead to misleading conclusions. In particular, proving the link between a disease and an unculturable virus identified from nucleic acid sequences has stimulated much debate. This has led to the formation of revised guidelines for defining the causal relationship between a virus and a disease (Fredericks and Relman, 1996). The presence of viral sequences within disease tissue in the absence of other corroborating data is insufficient to prove that a new virus causes a particular disease. Detection of viruses may reflect either the presence of virus in surrounding cells and tissues, or the ability of a virus to replicate within the microenvironment of the disease tissue, rather than the virus causing the observed disease. In the example of hantavirus pulmonary syndrome, layers of evidence are compiled to reach the conclusion of disease causation. The opposite has been true for HGV. Following the virus discovery, the accumulation of more data now indicates that the virus is not the cause of hepatitis in humans but rather is a virus from a subgroup of individuals in whom it causes no apparent disease (Alter et al, 1997a,b).

The list of diseases with possible microbial aetiology is still large and new pathogens are likely to be identified for some of these by the molecular biological methods described. Specific diseases such as Kawasaki syndrome (Kawasaki et al, 1974), sarcoidosis (Newman et al, 1997) and multiple sclerosis (Allen and Brankin, 1993), as well as more general disease groups, such as neurodegenerative disorders, forms of arthritis, inflammatory bowel disease, autoimmune disease and certain cancers, may have an infectious origin. What may make new aetiological agents hard to define will be those

agents that are part of a multifactorial origin of disease. Disease may result from multiple viral exposure, as is the case for dengue fever. Alternatively, viruses may act as triggers for a disease, or cause disease following long-term chronic asymptomatic infections. Clearly, molecular biological approaches should help unravel the causes of many pre-existing diseases, while also enabling the rapid identification of any new, virulent, emerging infectious disease.

REFERENCES

Allan JS, Short M, Taylor ME et al (1991) Species-specific diversity among simian immunodeficiency viruses from African green monkeys. *J Virol* **65:** 2816–2828.

Allen I & Brankin B (1993) Pathogenesis of multiple sclerosis – the immune diathesis and the role of viruses. *J Neuropathol Exp Neurol* **52:** 95–105.

Alter HJ, Nakatsuji Y, Melpolder J et al (1997a) The incidence of transfusion-associated hepatitis G virus infection and its relation to liver disease. *N Engl J Med* **336:** 747–754.

Alter MJ, Gallagher M, Morris TT et al (1997b) Acute non-A-E hepatitis in the United States and the role of hepatitis G virus infection. Sentinel Counties Viral Hepatitis Study Team. *N Engl J Med* **336:** 741–746.

Ayala M, Balint RF, Fernandez-de-Cossio ME, Canaan-Haden L, Larrick JW & Gavilondo JV (1995) New primer strategy improves precision of differential display. *BioTechniques* **18:** 842–850.

Barre Sinoussi F, Chermann JC, Rey F et al (1983) Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220:** 868–871.

Barrett T, Visser IK, Mamaev L, Goatley L, van Bressem MF & Osterhaust AD (1993) Dolphin and porpoise morbilliviruses are genetically distinct from phocine distemper virus. *Virology* **193:** 1010–1012.

Bode L, Zimmermann W, Ferszt R, Steinbach F & Ludwig H (1995) Borna disease virus genome transcribed and expressed in psychiatric patients. *Nature Med* **1:** 232–236.

Challoner PB, Smith KT, Parker JD et al (1995) Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc Natl Acad Sci USA* **92:** 7440–7444.

Chang Y, Cesarman E, Pessin MS et al (1994) Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266:** 1865–1869.

Choo QL, Kuo G, Weiner AJ, Overby LR, Bradley DW & Houghton M (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244:** 359–362.

Clavel F, Guetard D, Brun Vezinet F et al (1986) Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233:** 343–346.

Dalton AJ (1975) Microvesicles and vesicles of multivesicular bodies versus 'virus-like' particles. *J Natl Cancer Inst* **54:** 1137–1148.

Denner J & Dorfman NA (1977) Small virus-like particles in leukosis-like syndrome induced by certain antigens and immunostimulators. *Acta Biol Med Ger* **36:** 1451–1458.

Diatchenko L, Lau YF, Campbell AP et al (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93:** 6025–6030.

Dmochowski L, Langford PL, Williams WC, Liebelt AG & Liebelt RA (1968) Electron microscopic and bioassay studies of milk from mice of high and low mammary-cancer and high and low leukemia strains. *J Natl Cancer Inst* **40:** 1339–1358.

Don RH, Cox PT, Wainwright BJ, Baker K & Mattick JS (1991) 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucl Acids Res* **19:** 4008.

Dron M & Manuelidis L (1996) Visualization of viral candidate cDNAs in infectious brain fractions from Creutzfeldt–Jakob disease by representational difference analysis. *J Neurovirol* **2:** 240–248.

Dube S, Bachman S, Spicer T et al (1997) Degenerate and specific PCR assays for the detection of bovine leukaemia virus and primate T cell leukaemia/lymphoma virus *pol* DNA and RNA: phylogenetic comparisons of amplified sequences from cattle and primates from around the world. *J Gen Virol* **78:** 1389–1398.

Durst M, Gissmann L, Ikenberg H & zur Hausen H (1983) A papillomavirus DNA from a cervical carcinoma and its prevalence in cancer biopsy samples from different geographic regions. *Proc Natl Acad Sci USA* **80:** 3812–3815.

Elliott LH, Ksiazek TG, Rollin PE et al (1994) Isolation of the causative agent of hantavirus pulmonary syndrome. *Am J Trop Med Hyg* **51:** 102–108.

Feller WF & Chopra HC (1971) Virus-like particles in human milk. *Cancer* **28:** 1425–1430.

Fredericks DN & Relman DA (1996) Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin Microbiol Rev* **9:** 18–33.

Frenkel N, Schirmer EC, Wyatt LS et al (1990) Isolation of a new herpesvirus from human CD4+ T cells [published erratum appears in *Proc Natl Acad Sci USA* (1990) **87(19):** 7797]. *Proc Natl Acad Sci USA* **87:** 748–752.

Fulop L, Barrett AD, Phillpotts R, Martin K, Leslie D & Titball RW (1993) Rapid identification of flaviviruses based on conserved NS5 gene sequences. *J Virol Methods* **44:** 179–188.

Gao SJ & Moore PS (1996) Molecular approaches to the identification of unculturable infectious agents. *Emerg Infect Dis* **2:** 159–167.

Gelman IH, Zhang J, Hailman E, Hanafusa H & Morse SS (1992) Identification and evaluation of new primer sets for the detection of lentivirus proviral DNA. *AIDS Res Hum Retroviruses* **8:** 1981–1989.

Griffiths DJ, Venables PJW, Weiss RA & Boyd MT (1997) A novel exogenous retrovirus sequence identified in humans. *J Virol* **71:** 2866–2872.

Hubank M & Schatz DG (1994) Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucl Acids Res* **22:** 5640–5648.

Johnson KM, Lange JV, Webb PA & Murphy FA (1977) Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire. *Lancet* **i:** 569–571.

Kalyanaraman VS, Sarngadharan MG, Poiesz B, Ruscetti FW & Gallo RC (1981) Immunological properties of a type C retrovirus isolated from cultured human T-lymphoma cells and comparison to other mammalian retroviruses. *J Virol* **38:** 906–915.

Kawasaki T, Kosaki F, Okawa S, Shigematsu I & Yanagawa H (1974) A new infantile acute febrile mucocutaneous lymph node syndrome (MLNS) prevailing in Japan. *Pediatrics* **54:** 271–276.

Kiley MP, Bowen ET, Eddy GA et al (1982) *Filoviridae*: a taxonomic home for Marburg and Ebola viruses? *Intervirology* **18:** 24–32.

Kolykhalov AA, Feinstone SM & Rice CM (1996) Identification of a highly conserved sequence element at the 3′ terminus of hepatitis C virus genome RNA. *J Virol* **70:** 3363–3371.

Kuo G, Choo QL, Alter HJ et al (1989) An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science* **244:** 362–364.

Kurtzke JF (1993) Epidemiologic evidence for multiple sclerosis as an infection [published erratum appears in *Clin Microbiol Rev* (1994) **7(1):** 141]. *Clin Microbiol Rev* **6:** 382–427.

Lee HW, Lee PW & Johnson KM (1978) Isolation of the etiologic agent of Korean hemorrhagic fever. *J Infect Dis* **137:** 298–308.

Le Guyader F, Estes MK, Hardy ME et al (1996) Evaluation of a degenerate primer for the PCR detection of human caliciviruses. *Arch Virol* **141:** 2225–2235.

Li MD, Lemke TD, Bronson DL & Faras AJ (1996) Synthesis and analysis of a 640-bp *pol* region of novel human endogenous retroviral sequences and their evolutionary relationships. *Virology* **217:** 1–10.

Liang P & Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–971.

Linnen J, Wages J, Zhang Keck ZY et al (1996) Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271:** 505–508.

Lipkin WI, Travis GH, Carbone KM & Wilson MC (1990) Isolation and characterization of Borna disease agent cDNA clones. *Proc Natl Acad Sci USA* **87:** 4184–4188.

Lisieux T, Coimbra M, Nassar ES et al (1994) New arenavirus isolated in Brazil. *Lancet* **343:** 391–392.

Lisitsyn N & Wigler M (1995) Representational difference analysis in detection of genetic lesions in cancer. *Methods Enzymol* **254:** 291–304.

Lisitsyn N, Lisitsyn N & Wigler M (1993) Cloning the differences between two complex genomes. *Science* **259:** 946–951.

Medstrand P, Lindeskog M & Blomberg J (1992) Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J Gen Virol* **73:** 2463–2466.

Myers G, MacInnes K & Korber B (1992) The emergence of simian/human immunodeficiency viruses. *AIDS Res Hum Retroviruses* **8:** 373–386.

Newman LS, Rose CS & Maier LA (1997) Sarcoidosis. *N Engl J Med* **336:** 1224–1234.

Nichol ST, Spiropoulou CF, Morzunov S et al (1993) Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* **262:** 914–917.

Patience C, Takeuchi Y & Weiss RA (1997) Infection of human cells by an endogenous retrovirus of pigs. *Nature Med* **3:** 282–286.

Perron H, Garson JA, Bedin F et al (1997) Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. *Proc Natl Acad Sci USA* **94:** 7583–7588.

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 USA* **77:** 7415–7419.

Reyes GR, Purdy MA, Kim JP et al (1990) Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* **247:** 1335–1339.

Rice GP (1992) Virus-induced demyelination in man: models for multiple sclerosis. *Curr Opin Neurol Neurosurg* **5:** 188–194.

Rose TM, Strand KB, Schultz ER et al (1997) Identification of two homologs of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. *J Virol* **71:** 4138–4144.

Salas R, de Manzione N, Tesh RB et al (1991) Venezuelan haemorrhagic fever. *Lancet* **338:** 1033–1036.

Schweinfest CW, Henderson KW, Gu JR et al (1990) Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer. *Genet Anal Tech Appl* **7:** 64–70.

Seman G, Gallager HS, Lukeman JM & Dmochowski L (1971) Studies on the presence of particles resembling RNA virus particles in human breast tumors, pleural effusions, their tissue cultures, and milk. *Cancer* **28:** 1431–1442.

Shih A, Misra R & Rush MG (1989) Detection of multiple, novel reverse transcriptase coding sequences in human nucleic acids: relation to primate retroviruses. *J Virol* **63:** 64–75.

Shimizu H, Shimizu C & Burns JC (1994) Detection of novel RNA viruses: morbilliviruses as a model system. *Mol Cell Probes* **8:** 209–214.

Simons JN, Leary TP, Dawson GJ et al (1995a) Isolation of novel virus-like sequences associated with human hepatitis. *Nature Med* **1:** 564–569.

Simons JN, Pilot Matias TJ, Leary TP et al (1995b) Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci USA* **92:** 3401–3405.

Snijder EJ & Horzinek MC (1993) Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. *J Gen Virol* **74:** 2305–2316.

Snijder EJ, Horzinek MC & Spaan WJ (1993) The coronaviruslike superfamily. *Adv Exp Med Biol* **342:** 235–244.

Snijders PJF, Meijer CJLM & Walboomers JMM (1991) Degenerate primers based on highly conserved regions of amino acid sequence in papillomaviruses can be used in a generalized polymerase chain reaction to detect productive human papillomavirus infections. *J Gen Virol* **72:** 2781–2786.

Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE & Fanning TG (1997) Initial genetic characterization of the 1918 'Spanish' influenza virus. *Science* **275:** 1793–1796.

Tobler K & Ackermann M (1996) [Identification and characterization of new and unknown coronaviruses using RT-PCR and degenerate primers]. *Schweiz Arch Tierheilkd* **138:** 80–86.

Uchida T, Suzuki K, Hayashi N et al (1992) Hepatitis E virus: cDNA cloning and expression. *Microbiol Immunol* **36:** 67–79.

VanDevanter DR, Warrener P, Bennett L et al (1996) Detection and analysis of diverse herpesviral species by consensus primer PCR. *J Clin Microbiol* **34:** 1666–1671.

VandeWoude S, Richt JA, Zink MC, Rott R, Narayan O & Clements JE (1990) A Borna virus cDNA encoding a protein recognized by antibodies in humans with behavioral diseases. *Science* **250:** 1278–1281.

Wan JS, Sharp SJ, Poirier GMC et al (1996) Cloning differentially expressed mRNAs. *Nature Biotech* **14:** 1685–1691.

Yamanishi K, Okuno T, Shiraki K et al (1988) Identification of human herpesvirus-6 as a causal agent for exanthem subitum. *Lancet* **i:** 1065–1067.

Zhang Z & Martineau D (1996) Single-tube heminested PCR coupled with 'touchdown' PCR for the analysis of the walleye dermal sarcoma virus *env* gene. *J Virol Methods* **60:** 29–37.