Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease

David Moyes¹, David J. Griffiths² and Patrick J. Venables¹

¹The Kennedy Institute of Rheumatology, Imperial College London, 1 Aspenlea Road, London, W6 8LH, UK
²Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik, Midlothian, EH26 0PZ, UK

Human endogenous retroviruses (HERVs) result from ancestral infection by infectious viruses over millions of years of primate evolution. Some are transcriptionally active, express proteins and therefore have the potential to cause disease. Here we review the controversial attempts to link them with cancer and autoimmunity. The main difficulty is that most HERVs investigated to date are present at the same locus in 100% of the population. However, a new class of insertionally polymorphic HERV-K family members, present in a minority of individuals, has recently been described. We propose that insertionally polymorphic HERVs could be novel genetic risk factors and hence provide a new lease of life for research into HERVs and disease.

Evolution and classification of human endogenous retroviruses

Endogenous retroviruses (ERVs) are sequences within the genome that closely resemble infectious retroviruses [1–3] (Figure 1). They are the result of ancestral infections by exogenous retroviruses that became incorporated into germ-line DNA. Consequently, they are transmitted vertically in a Mendelian fashion, rather than by the horizontal spread typical of an infectious virus. ERVs have been detected in all vertebrates studied, and in humans they comprise ~8% of the genome [4]. The majority of human ERVs (HERVs) are also found in most apes and Old World monkeys, indicating that they originated in our ancestors at least 25 million years ago.

In some species, exogenous retroviruses related to ERVs still exist as infectious agents; for example, jaagsiekte retrovirus causing lung cancer in sheep [5], and feline leukemia virus causing leukemia and lymphomas in cats [6]. Koala endogenous retrovirus is of particular interest, because not only is it threatening the species with extinction, but also it is an example of an infectious retrovirus that is currently undergoing endogenization [7]. In humans, however, infectious counterparts of HERVs have yet to be identified.

Because most infectious retroviruses are harmful, after they become established in the genome, evolutionary selection ensures that they are disabled by mutations that prevent expression of some or all proviral genes. In most cases, recombination of the two long terminal repeats (LTRs) has resulted in the excision of almost the entire proviral genome to leave a single (solitary) LTR. Where the viral genome is maintained, other mutations involving smaller deletions, frame shifts or premature stop codons prevent the expression of functional proteins from most HERV loci. Hypermethylation of HERV promoters provides a further mechanism for silencing provirus expression [8]. These mechanisms for inactivating retroviruses have led to the concept that HERVs are nothing more than ‘junk’ DNA or ‘fossils’ of ancestral infections by exogenous retroviruses. Although this view might be accurate for the majority of HERV insertions, recent evidence indicates that specific proviruses might have physiological and pathological functions.

Here, we will concentrate on those HERVs that are transcriptionally active, and have the potential to interact with host functions and cause disease. HERVs can be classified using the single letter amino acid code for the tRNA specificity of the primer binding site used to initiate reverse transcription. Thus, HERV-W would use a tryptophan tRNA if it were an infectious virus, and HERV-K would use lysine [2]. Within the HERV-K family there are several subgroups denoted HML-1 to HML-10 (human endogenous MMTV-like), each representing a separate germ-line infection. Members of the HML-2 subgroup have the greatest coding competence and have therefore received most attention in relation to disease [1]. Two forms of HERV-K(HML-2) have been described (Figure 1); type II proviruses are complete whereas type I proviruses have a 292-bp deletion at the boundary of the pol and env genes (reviewed in Ref. [1]).

Those HERVs that have been linked with human disease include HERV-K, HERV-W, HERV-R and HERV-H. We will argue that much of the evidence for a pathological role for these sequences could be flawed by their ubiquity in the population. We conclude that the emerging literature on insertionally polymorphic HERVs provides a new lease of life for research into HERVs and human pathology.

Physiological roles of HERVs

In recent years, evidence has accumulated suggesting that some HERV proteins have been co-opted into physiological roles in the host. The best example is that of HERV-W, a family of around 100 proviruses that are largely defective except for a provirus on chromosome 7 that encodes a
functional Env protein [9]. This protein (also known as syncytin) is highly fusogenic, and because of its placental expression is predicted to mediate cell-to-cell fusion during formation of the syncytiotrophoblast, the placental membrane that separates maternal from foetal tissues [10]. Recent reports have demonstrated fusogenic activity for other HERV Env proteins [11,12], and for the Env protein of endogenous jaagsiekte retrovirus in sheep [13], suggesting that conservation of endogenous retroviral env sequences for placentogenesis might be widespread amongst mammalian species. Additional physiological roles proposed for HERV proteins include localized immunosuppression and resistance to viral infection [14]. Furthermore, a recent genome survey for Gag-like proteins suggested that >100 human proteins might have originated in retroviruses or related retrotransposons [15]. Although further evidence is required to confirm these functions, these data suggest that the contribution of HERV-derived proteins to the human proteome is likely to be underappreciated at present.

HERVs and disease
The major human retroviral pathogens, HIV-1, HIV-2 and human T-lymphotropic virus type I (HTLV-I), cause...
Box 1. HERV-K and cancer

1970s–1980s: Particles resembling retroviruses are observed in various human cancers and tumour cell lines by electron microscopy.
1986: HERV-K cloned because of its sequence similarity with mouse mammary tumour virus [48].
1993: Particles in human teratocarcinoma identified as HERV-K by immunoelectron microscopy (see Ref. [1]).
1995: A high proportion of seminoma patients found to have antibodies to HERV-K Gag protein (reviewed in Ref. [1]).
1995–1996: Retrovirus particles in breast cancer cell lines found to encapsidate a mixture of HERV sequences [49,50].
2000: HERV-K Rec protein shown to impair spermatogenesis in a transgenic mouse model [51].
2003: HERV-K particles identified in melanoma [19].
2005: Rec-transgenic mice develop testis-specific carcinoma in situ [17].
2006: HERV-K (and other HERV) envelope proteins are expressed in ovarian cancer [18].
2007: HERV-K Rec and Np9 proteins shown to bind promyelocytic leukaemia zinc finger, a transcriptional repressor of c-myc [16].

Conclusion: Expression of HERV-K is activated in cancer but it is still unclear whether it contributes to tumourigenesis. HERV-K Rec and Np9 might have a direct role in seminoma.

Box 2. HERVs and multiple sclerosis

1986: Identification of RNA sequences from an HTLV-I-like retrovirus in cerebrospinal fluid (CSF)-derived cell lines from patients with MS and antibodies to HTLV-I in their CSF and serum (reviewed in Ref. [52]).
1987: No evidence of HTLV-I or HIV (then described as HTLV-III) in MS tissue or CSF or of antibodies to HTLV-I in their sera [52].
1991: Retroviral-like particles described in MS [52].
1997: RNA from part of the pol gene of ‘multiple sclerosis-associated retrovirus’ (MSRV) identified by reverse transcriptase PCR from the brains and B-cell lines of patients with MS [53].
1999: MSRV established as part of the multicopy HERV-W family [54].
2001–2003: HERV-H expression in MS including particle production, RNA detection and RT activity (reviewed in Ref. [52]).
2001: HERV-K and HERV-W RNA expression increased in inflammatory lesions from the brains of MS patients [55].
2005: Upregulated expression of HERV-W Gag proteins in brain lesions from patients with MS compared with normal brain [25].

Conclusion: Several studies have shown that HERVs are upregulated in inflammatory lesions from patients with MS, but none has shown that this is disease-specific (as opposed to inflammatory in general) or uniquely specific to a particular HERV. Further analysis of the pro-inflammatory properties of HERV-W Env protein is necessary.
Box 3. HERV-K18 and diabetes

1997: A ‘novel’ endogenous retrovirus, denoted IDDMK1,2,22 (now renamed HERV-K18), was proposed as a pathogenic agent based on the detection of mRNA in peripheral blood from patients with type 1 diabetes but not in controls. The N-terminal region of the Env protein was reported to have superantigenic activity resulting in oligoclonal expansion of a specific T-cell subset [56].

1998–1999: The specific detection of mRNA in plasma from patients with diabetes was not confirmed [57].


2004–2006: The possible superantigenic activity of this envelope protein is further investigated including its specific upregulation by Epstein–Barr virus (EBV) [59].

Conclusion: The initial study, which claimed that the mRNA of HERV-K18 was specifically detected in plasma from patients with diabetes, was not confirmed by four independent studies [57,60–62]. Uptregulation of the superantigen activity of HERV-K18 by EBV implies an aetiological role for EBV rather than HERV-K18. There is no epidemiological evidence to support this.

Other diseases in which HERVs have been reported to be upregulated at the site of the lesion include rheumatoid arthritis (RA), where increased expression of HERV mRNA and proteins has been reported in synovial membrane and fluid [29,30], and psoriasis, where upregulation of HERV-K, -W and -E mRNA [31] has been found in affected skin. In terms of an antibody response, systemic lupus erythematosus (SLE) is analogous to both Sjögren’s syndrome and MS, with raised levels of antibodies to a variety of HERV sequences including HERV-K [22].

Type 1 diabetes and its alleged link with HERV-K18 deserves a special mention here (Box 3). The original report that RNA from HERV-K18 was present in the plasma of patients but not in controls was challenged by four independent studies. However, the possible superantigenic activity of the Env protein, defined by its putative role in the expansion of a specific pathogenic T-cell subset, is still under investigation.

Possible pathogenic mechanism of HERVs

Infectious retroviruses produce several proteins that, to facilitate viral replication, stimulate the cells in which they are expressed and enable the virus to evade an immune response. Similar proteins can be encoded by HERVs, a property that confers on them the potential to cause disease. Such mechanisms are extensively reviewed elsewhere [1,2,32] and summarized here in Figure 2, but direct evidence that they occur in vivo is remarkably sparse. Briefly, HERVs could cause pathology by: (i) disrupting genes at their integration site; (ii) suppressing or stimulating the immune response; or (iii) expressing accessory proteins that have direct effects on the cells in which they are expressed. The overwhelming problem in pursing these hypotheses is that the integration sites of the vast majority of HERVs are identical in different individuals and it becomes increasingly difficult to propose that a given HERV can uniquely cause a disease.

HERV polymorphisms

The existence of polymorphisms provides one explanation of how a ubiquitous gene such as a HERV can cause disease in only a proportion of individuals. We will consider such polymorphisms subdivided into two broad categories: (i) sequence polymorphisms and (ii) insertional polymorphisms.

Sequence polymorphisms

As with many genes, HERVs have polymorphisms that could affect the function of an expressed product. Only a few such polymorphisms have been described, partly because of the difficulty in identifying them against the background of closely similar proviruses. The availability of the human genome sequence has provided a way to focus on specific HERV proviruses and to examine sequence variations in the context of their integration site. For example, HERV-K18 is located in the first intron of the CD48 gene on human chromosome 1q. Using specific PCR primers spanning the insertion site and the putative superantigen region, Marguerat et al. reported a link between one of three HERV-K18 haplotypes (haplotype 3) and a protective effect against the development of type 1 diabetes in a large family-based association study [33]. It is possible that allelic variation of either the HERV-K18 provirus or the flanking CD48 sequence could be responsible for the disease association observed. However, the association was weak and it is clear that if these observations are confirmed in independent studies, HERV-K18 cannot be more than a weak susceptibility gene for diabetes. There is a much older literature examining polymorphisms of HERV-R and the partially deleted retroviral element denoted ‘HTLV-related endogenous sequence-1’ (HRES-1) in a small number of patients with autoimmune diseases. A possible association of one of three haplotypes of HRES-1 with SLE was described [34]. We can therefore conclude that to date there is little convincing evidence that any sequence polymorphism of a HERV provides susceptibility to human disease.

Insertional polymorphisms

Insertionally polymorphic HERVs are proviruses that are present only in a proportion of the human population. For completeness we include the HERV-K(C4) gene [35]. It is insertionally polymorphic in that it is present in about ~70% of the population and integrated into intron 9 of the C4 complement gene. However, unlike the insertionally polymorphisms discussed below, the absence of HERV-K(C4) in some people is the result of its deletion rather than its recent integration into the genome. A feature of the HERV-K(C4) provirus is that it is in opposite orientation to the C4 gene, and it has been suggested that the generation of antisense retroviral RNA from the C4 promoter inhibits the production of viral RNA from an infecting retrovirus [35]. This hypothesis is supported by the demonstration that antisense HERV-Kic4 transcripts downregulate expression of retroviral constructs in vitro [36]. Whether this occurs in vivo as a protective effect against HIV or HTLV-I infection is yet to be examined.

There is an accumulating number of a new class of insertionally polymorphic retroviruses in the HML-2 subgroup of the HERV-K family. The first two of these, HERV-K113 and HERV-K115, were described in 2001 [37]. Both are recent additions to the human genome (estimated as
being <200,000 and 1.2 million years old, respectively). HERV-K113 was present in the genomes of 29% of individuals, whereas the prevalence of HERV-K115 was 16%. Both are full-length proviruses and have open reading frames for all of their genes with the exception of a frameshift mutation in the gag gene of HERV-K115. Thus HERV-K113 has the potential to encode all the elements necessary for a fully functional retrovirus, although it seems that the Env protein is not fusogenic [12]. Since this initial report in 2001, a further 10–13 insertional polymorphisms of HERV-K(HML-2) have been identified [38–40] (Table 1) and it is likely that others remain undiscovered.

Are polymorphic HERVs associated with disease?
We propose that insertionally polymorphic HERVs are more likely to be involved in disease than those elements that are present in all humans. First, their recent insertion might disrupt host genes, although this does not seem to be the case for the examples identified to date (Table 1) [37,39]. However, the enhancer elements of retroviral LTRs can influence expression of neighbouring genes over relatively large genomic distances [41] and this could be worthy of further investigation. Second, as recent integrations, polymorphic HERVs are more likely to have retained functional coding sequences with the capacity to modulate cellular proliferation or the immune response (Figure 2). Third, their presence in only some individuals raises the possibility that they might be pathogenic in a manner analogous to infectious retroviruses.

So far, only a few studies have examined the possible association of insertionally polymorphic HERVs and disease, all on HERV-K113 and HERV-K115. In our own laboratory, we found that the prevalence of HERV-K113

Figure 2. Potential mechanisms for HERVs and disease. HERVs can potentially cause disease by a variety of mechanisms (reviewed in Ref. [32]). They can disrupt or modulate the expression of host genes at or near their integration site through the promoter activity of the LTR [67,68]; viral proteins such as Gag (green) or Env (red) can suppress or stimulate both the adaptive and the innate immune response [23,24,58,69,70], or accessory proteins (yellow) can have direct effects on the cells in which they are expressed [16,17]. Although some of these mechanisms are supported by evidence obtained in vitro and in murine models, their significance in vivo in humans remains speculative. Currently, greatest support for a pathogenic role comes from work on HERV-K Rec and Np9 and the HERV-W Env protein.

Table 1. Polymorphic HERV-K insertions

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Designation</th>
<th>Alleles*</th>
<th>Gene frequency**</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>6p21.3c</td>
<td>HERV-KC4</td>
<td>Provirus, PI</td>
<td>[35]</td>
<td></td>
</tr>
<tr>
<td>19p13.11c</td>
<td>HERV-K113</td>
<td>Provirus, PI</td>
<td>0–0.16 [37,42]</td>
<td></td>
</tr>
<tr>
<td>8p23.1c</td>
<td>HERV-K115</td>
<td>Provirus, PI</td>
<td>0–0.22 [37,38,42]</td>
<td></td>
</tr>
<tr>
<td>9q12.c</td>
<td>sLTR, PI</td>
<td>[40]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5p15.31c</td>
<td>sLTR, PI</td>
<td>0.74–0.83 [38,63]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8c</td>
<td>Provirus, PI</td>
<td>[64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12c</td>
<td>sLTR, PI</td>
<td>[64]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3p25c</td>
<td>sLTR, PI</td>
<td>0.92 [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9q33c</td>
<td>sLTR, PI</td>
<td>0.92 [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q36c</td>
<td>sLTR, PI</td>
<td>0.26 [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11q22c</td>
<td>Provirus, sLTR, PI</td>
<td>0.94 [38,39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8p22c</td>
<td>sLTR, PI</td>
<td>0.49 [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12q13c</td>
<td>Provirus, sLTR, PI</td>
<td>0.59 [38]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7p22.1d</td>
<td>HERV-K108</td>
<td>Provirus, sLTR, TR</td>
<td>[39,65]</td>
<td></td>
</tr>
<tr>
<td>6q14.1d</td>
<td>HERV-K109</td>
<td>Provirus, sLTR</td>
<td>[66]</td>
<td></td>
</tr>
<tr>
<td>12q14d</td>
<td>Provirus, sLTR</td>
<td>[39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1p31d</td>
<td>Provirus, sLTR</td>
<td>[39]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10p12.1d</td>
<td>HERV-K103</td>
<td>Provirus, sLTR</td>
<td>[40]</td>
<td></td>
</tr>
<tr>
<td>3q13.2d</td>
<td>HERV-K106</td>
<td>Provirus, sLTR</td>
<td>[40]</td>
<td></td>
</tr>
</tbody>
</table>

*LTR, solitary LTR; PI, preintegration site; TR, tandem repeat.
**Combined frequency of inserted alleles; note that estimates are based on small sample sizes and are greatly influenced by ethnic origin of sample.
*Insertionally polymorphic HERV-K proviruses. All are members of the HML-2 subgroup, except HERV-KC4, which does not represent a recent integration.
*Additional polymorphic HERV-K loci that do not represent insertional polymorphisms.
and HERV-K115 was 4.16% [gene frequency (Gf) 0.02] and 1% (Gf 0.01), respectively, in 96 people in the UK [42] but increased to 21.8% and 34.1%, (Gf 0.18 and 0.25), respectively, in 174 individuals from Africa. The overall prevalence is in accordance with the original report [37] because 23% of the individuals in that study originated from central Africa. This geographical variation indicates that both viruses were recently integrated into the human genome, possibly after the migration of early humans out of Africa less than 100,000 years ago. Alternatively, these proviruses could be older (i.e. they integrated before the emergence from Africa) but differed in distribution because of evolutionary selection pressure; that is, the proviruses were advantageous in Africa but deleterious in the UK. In the same study, we found that the prevalence of HERV-K113 but not HERV-K115 is significantly increased in patients with MS (11.9%, Gf 0.06; P < 0.05) and Sjögren’s syndrome (15.6%, Gf 0.08; P < 0.01) compared with normal controls (4.16%, Gf 0.02) and patients with RA (5.7%, Gf 0.03). Given the widespread racial variation in prevalence of both proviruses, it is essential to confirm this association with family studies, which are currently in progress.

Other groups have investigated the link between insertionally polymorphic HERVs and disease. The presence of HERV-K113 and K115 was investigated in patients with breast cancer but no association was detected [43]. Similar studies on HERV-K115 in schizophrenia also found no statistically significant association with disease [44], although a possible link with early-onset schizophrenia was nevertheless suggested. The authors concluded that further larger-scale studies were required to address this point.

Is there an infectious variant of HERV-K?
The short evolutionary history of insertionally polymorphic retroviruses provides a plausible model for the existence of an infectious counterpart, at least for the HERV-K(HML-2) family. A recent study has provided tantalizing data suggesting that such a virus might be circulating among humans. By analysing nucleotide substitutions and the acquisition of stop codons in HERV-K sequences, Belshaw and co-workers demonstrated that the majority of these proviruses are the result of an infection rather than transposition, indicating the presence of a pool of exogenous HERV-K-like viruses throughout the 25 million years of recent HERV-K history [45]. Such reinfection could involve movement from somatic cells to germ cells within the same individual and does not necessarily require human–human spread.

Despite this recent evidence, attempts to identify infectious HERV-K proviruses in malignancies where particle formation has been described have so far been unsuccessful [21]. Recently, two groups have independently reconstructed prototype HERV-K clones that represent the consensus sequence of recently integrated HERV-K(HML-2) proviruses [46,47]. These viruses are infectious in vitro and although artificial, provide novel tools for future research on the pathogenic effects of HERV-K in cancer and autoimmunity.

Concluding remarks
Insertionally polymorphic HERVs are much more compelling candidates for causing disease than the ancient, disrupted sequences that have littered our genome throughout millions of years of prime evolution. The relatively small number of polymorphic insertions identified to date is based on the analysis of only a few human genomes, and as additional individuals are characterized it seems probable that more insertional polymorphisms will be discovered. It is conceivable that, as in animals, some recently integrated HERVs have infectious counterparts that have eluded detection to date because of their similarity to their endogenous relatives. Thus we propose that research on HERVs and human disease, which in the past has been dogged by controversy and unconfirmed findings, has now found a new lease of life with the discovery of insertional polymorphisms.

Acknowledgements
We thank Karin Lundberg for critical discussion of the manuscript and the Multiple Sclerosis Society, UK, the Wellcome Trust and the Arthritis Research Council for financial support.

References
17 Galli, U.M. et al. (2005) Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. Oncogene 24, 3223–3228
26 Green, J.E. et al. (1989) Exocrinopathy resembling Sjogren’s syndrome in HTLV-1 transgenic mice. Nature 341, 72–74
35 Dangel, A.W. et al. (1994) The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. Immunogenetics 40, 425–436
49 Seifarth, W. et al. (1995) Retrovirus-like particles released from the human breast cancer cell line T4-7 D display type B and C related endogenous retroviral sequence. J. Virol. 69, 6405–6416
57 Badenhoop, K. et al. (1999) IDDM patients neither show humoral reactivities against endogenous retroviral envelope protein nor do they differ in retroviral mRNA expression from healthy relatives or normal individuals. Diabetes 48, 215–218
60 Jaeckel, E. et al. (1999) No evidence for association between IDDMK1,2/22, a novel isolated retrovirus, and IDDM. Diabetes 48, 209–214
62 Muir, A. et al. (1999) The IDDM1,2/22 retrovirus is not detectable in either mRNA or genomic DNA from patients with type 1 diabetes. Diabetics 48, 219–222

Medstrand, P. *et al.* (2001) Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biol. Chem.* 276, 1896–1903
