Virology Journal

CrossMark

The sense behind retroviral anti-sense transcription

Mamneet Manghera[1], Alycia Magnusson[2] and Renée N. Douville[1,2*]

Abstract

Retroviruses are known to rely extensively on the expression of viral proteins from the sense proviral genomic strand. Yet, the production of regulatory retroviral proteins from antisense-encoded viral genes is gaining research attention, due to their clinical significance. This report will discuss what is known about antisense transcription in *Retroviridae,* and provide new information about antisense transcriptional regulation through a comparison of Human Immunodeficiency Virus (HIV), Human T-cell Lymphotrophic Virus (HTLV-1) and endogenous retrovirus-K (ERVK) long terminal repeats (LTRs). We will attempt to demonstrate that the potential for antisense transcription is more widespread within retroviruses than has been previously appreciated, with this feature being the rule, rather than the exception.

Keywords: Viral genomes, Retrovirus, Human endogenous retrovirus-K, Antisense transcription, Long-terminal repeat (LTR), Transcription factors, Conserved protein domains

Main text

Retroviruses share a common genomic organization in which the 5′ long terminal repeat (LTR) is followed by the *gag*, *pro*, *pol* and *env* genes, and terminates with the 3′ LTR. Accessory genes are encoded in ways unique to each viral species. The majority of viral protein products stem from the translation of sense-strand RNA transcripts. Until recently, retroviral antisense transcription has been largely overlooked as a source of viral RNA and proteins. However, there is accumulating evidence of antisense transcription in numerous exogenous retroviral genera, including lentiviruses, deltaretroviruses, gammaretroviruses and betaretroviruses. Thus, the expression of antisense proteins may be a broad phenomenon occurring across *Retroviridae*, suggesting that antisense encoded genes are an integral part of the viral genome. This report contributes to our understanding of antisense transcription by characterizing exogenous and endogenous retroviral 3′ (antisense) promoters. Our results highlight that antisense transcription may be more widespread than previously appreciated, with endogenous retroviruses (ERVs) incapable of antisense transcription being the exception, rather than the rule.

Antisense transcription among exogenous retroviruses

Antisense transcription is much better understood in exogenous retroviruses, as compared to their endogenous counterparts. Human Immunodeficiency Virus (HIV) and Human T-cell Lymphotrophic Virus (HTLV) are the characterized retroviruses exhibiting this phenomenon. The proteins encoded by their antisense strands serve important functions, including control of viral sense transcription, as well as viral latency, pathology, and spread [1–4].

HIV antisense transcription

The ability of HIV-1 to encode an antisense protein was suspected as early as 1988, when a conserved ORF, later named *asp-1*, was identified in the region complementary to the *env* gene in many HIV-1 strains [5]. Since then, many studies have confirmed the expression of ASP-1 RNA and protein *in vitro* in various HIV-expressing cell types, including monocyte-derived macrophages, dendritic cells, and T cells [6–10]. Antibodies recognizing an antisense protein derived from the *env* gene have also been identified in HIV[+] patients [11]. Recently, ASP-1 has been shown to induce autophagy, which may explain its low abundance in HIV-1 infected cells and the inherent difficulty in detecting this protein

* Correspondence: r.douville@uwinnipeg.ca
[1]Department of Immunology, University of Manitoba, Winnipeg, MB, Canada
[2]Department of Biology, The University of Winnipeg, Winnipeg, MB, Canada

Manghera *et al. Virology Journal* (2017) 14:9

Page 2 of 12

[12]. ASP-1 has been postulated to utilize this pathway to enhance HIV-1 replication, as wild-type, but not mutated forms of this antisense protein, resulted in optimal viral replication through stimulation of autophagy [12]. In contrast, antisense transcript variants corresponding to proviral *asp* have been shown to inhibit HIV-1 replication *in vitro* in acutely and chronically infected cell lines, as well as in HIV-1 infected human PBMCs [8]. This suggests that HIV-1 antisense transcription may have a critical role in establishing viral latency. Moreover, ASP-1 can form stable cytosolic aggregates, which may sequester essential cellular proteins, and thus modulate the function of cellular pathways in HIV-1 infected cells [12]. Nonetheless, the precise functions of viral antisense RNA and protein during HIV-1 infection *in vivo* remain to be clearly elucidated.
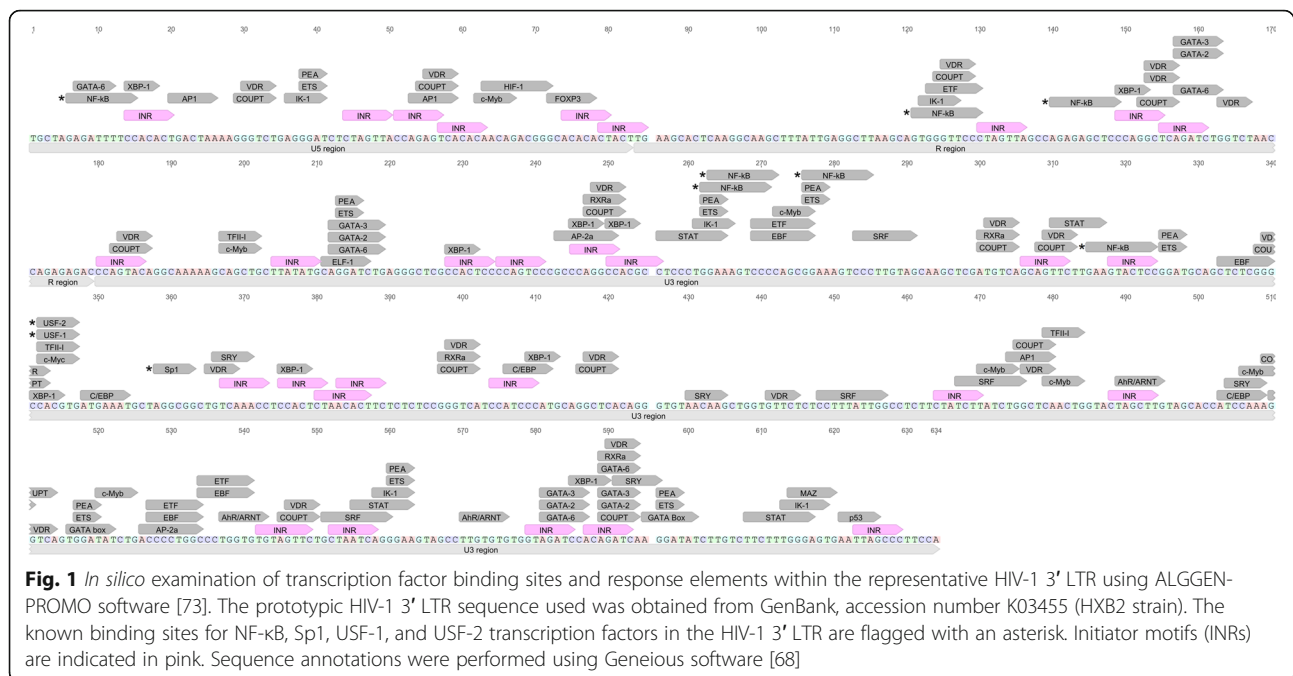
The 3′ LTR promoter is crucial for driving antisense transcription in HIV-1. It has been experimentally demonstrated that HIV-1 antisense transcription initiates at multiple positions in the U3 region of the 3′ LTR, as well as at other downstream regions within the antisense strand [8]. Other studies have reported that HIV-1 antisense transcription is initiated in the U5 region of the 3′ LTR [7, 13]. The multiplicity of transcription initiation sites may be a consequence of the lack of a TATA box in the HIV-1 3′ LTR [6], in which case transcription is initiated through alternative core promoter elements called initiator (INR) motifs (YYANWYY) [7, 14]. The presence of these multiple INR motifs serves to explain the variability observed in transcription initiation sites reported by different studies.

In comparison to that of its sense transcription, the regulation of HIV-1 antisense transcription is not well understood. Nonetheless, several host transcription factors have been shown to play key roles in inducing transcription from the antisense strand of HIV-1, including Specificity Protein-1 (Sp1) [15, 16] Upstream Stimulating Factor (USF) [17], and Nuclear Factor-kappa B (NF-κB) [8, 16]. Mutagenesis of a conserved USF site has been shown to diminish the activity of the HIV-1 antisense promoter in reporter constructs [17]. The HIV-1 3′ LTR also contains several conserved NF-κB binding sites, which are known to drive HIV-1 antisense transcription. Point mutations in these sites have been demonstrated to down-regulate the activity of the antisense promoter in HIV-1 LTR reporter constructs [8, 15, 17]. NF-κB-activating agents, such as phorbol 12-myristate 13-acetate (PMA) and tumor necrosis factor α (TNFα), are also known to induce HIV-1 antisense transcription, likely through increased binding of NF-κB on the viral 3′ LTR [6, 8, 15]. In line with this finding, HIV-1 antisense LTR reporter plasmids containing mutated κB sites demonstrate decreased responsiveness towards PMA and TNFα stimulation [6, 8]. PMA stimulation of antisense transcription in a luciferase-expressing HIV-1 proviral DNA clone has also been demonstrated in both transfection and infection experiments [6]. In comparison, TNFα-mediated induction of HIV-1 antisense transcription remains debatable, as other studies have failed to replicate this phenomenon [6]. Thus, accumulating evidence illustrates an important role of NF-κB in the induction of HIV-1 antisense transcription. However, our bioinformatics analysis of the HIV-1 antisense promoter suggests many additional transcription factors likely contribute to the overall regulation of negative-sense transcripts in this retrovirus (Fig. 1, Table 1).

In addition, there is evidence of single nucleotide polymorphisms in transcription factor binding sites within the LTRs of different HIV-1 subtypes [18–21]. This sequence variation leads to subtype-specific differences in proviral gene expression, thereby imparting unique biological characteristics to a given strain. This is well illustrated for the 5′ LTR of HIV-1 subtype E, which harbors a shift from an NF-κB to a GABP binding site [20]. Abolished NF-κB binding to this LTR does not lead to a loss of promoter function *in vitro*; instead, gain of GABP binding to this mutated NF-κB site enhances Tat-mediated HIV-1 gene expression in several cell types, as well as improves virus replication in SuPT1 cell line [20]. Thus, variability in transcription factor binding sites in retroviral LTRs can have a positive impact on proviral gene expression under certain conditions – this may serve to enhance retroviral fitness and spread. Likewise, subtype-specific variations in the HIV-1 3′ LTR may also exert another layer of control over the proviral antisense transcription.

Further, retroviral proteins can also modulate proviral antisense transcription. Conflicting results have been reported on the potential role of the HIV-1 accessory protein Tat in regulating antisense transcription. Tat has been reported to enhance HIV-1 antisense transcription in cell lines co-transfected with a Tat expression vector and luciferase reporter plasmids containing HIV-1 3′ LTR but not 5′ LTR [6]. However, in studies utilizing luciferase reporter constructs containing both HIV-1 3′ LTR and 5′ LTR, Tat was not shown to alter antisense transcription [22, 23]. Likewise, in a separate study utilizing HIV-1 3′ LTR luciferase reporter plasmids, overexpression of Tat did not influence antisense luciferase activity [8]. Thus, the role of Tat in regulating HIV-1 antisense transcription remains controversial and should be confirmed in the context of full length proviruses integrated during HIV-1 infection, rather than in artificial reporter assays. As there is no evidence of TAR RNA synthesis during antisense transcription, the mechanism by which Tat influences the activity of the HIV-1 antisense promoter remains unknown. It is possible that the interaction of Tat with cellular transcription factors,

Manghera et al. Virology Journal (2017) 14:9

Page 3 of 12

Fig. 1 In silico examination of transcription factor binding sites and response elements within the representative HIV-1 3′ LTR using ALGGEN-PROMO software [73]. The prototypic HIV-1 3′ LTR sequence used was obtained from GenBank, accession number K03455 (HXB2 strain). The known binding sites for NF-κB, Sp1, USF-1, and USF-2 transcription factors in the HIV-1 3′ LTR are flagged with an asterisk. Initiator motifs (INRs) are indicated in pink. Sequence annotations were performed using Geneious software [68]

such as Sp1, modulates their binding to the HIV-1 3′ LTR [6], which may affect the extent of HIV-1 antisense transcription.

HTLV Antisense transcription

A large portion of our knowledge on retroviral antisense transcription stems from studies of Human T-Lymphotropic Viruses (HTLV), particularly HTLV-1. The HTLV-1 antisense genomic strand encodes a basic leucine zipper (bZIP)-containing protein, designated HBZ. Although HTLV-1 is capable of infecting different cell types in vitro, HBZ protein is mainly detected in CD4+ T cells in vivo [3, 24, 25]. This cell-type specific expression of HBZ has been shown to play a variety of roles in the pathogenesis of HTLV-mediated T-cell leukemia (reviewed in [3, 26]). For instance, HBZ transforms T-cells into a cancerous phenotype, in part by enhancing the expression of chemokine receptor CCR4 in this cell type, which promotes T-cell proliferation and migration [27]. HBZ also inhibits HTLV-1 sense transcription by recruiting essential transcription factors, such as CREB, away from the proviral sense promoter – this process facilitates HTLV-1 latency in infected T cells [3]. HBZ also affects many other cellular processes, including host gene expression, innate immune signaling, apoptosis, autophagy, and DNA repair – all of which further influence the pathology of the HTLV-1 infection (reviewed in [3]). Similar to HTLV-1, HTLV-2, HTLV-3, and HTLV-4 are equally capable of producing

antisense proteins – APH-2, APH-3, and APH-4, respectively – though their functions have not been clearly elucidated [4, 26, 28–30].

Despite the extensive research focused on deciphering the role of HTLV-encoded antisense proteins in disease pathogenesis, there are a limited number of studies aimed at understanding the regulation of HTLV antisense transcription at the level of the proviral 3′ LTR. It has been demonstrated that HTLV-1 hbz is transcribed starting from the 3′ LTR of the HTLV-1 provirus [31–33]. Initiation of transcription is possible at several different positions within the R and U5 regions of the 3′ LTR [31]. Like HIV-1 antisense promoter, the HTLV-1 3′ LTR is a TATA-less promoter harboring many INR motifs, thus leading to a multitude of antisense transcription initiation sites [31, 34, 35]. The transcription of hbz relies heavily on three Sp1 sites in the U5 region of the proviral 3′ LTR [23, 31, 34, 36]. In luciferase assays, HTLV-1 antisense promoter activity is markedly reduced upon mutation of single or multiple Sp1 sites [34]. The same study identified binding sites for GATA binding protein-2 (GATA-2), cAMP responsive element binding protein (CREB), activating protein 1 (AP-1), and nuclear factor-1 (NF-1) in the HTLV-1 3′ LTR. However, mutations of each of these sites only reduced promoter activity slightly in luciferase assays [34]. Other cellular transcription factors, including activating transcription factor (ATF), CCAAT-enhancer binding protein (C/EBP), and histone acetyltransferase p300 have also been shown to bind the HTLV-1

Manghera *et al. Virology Journal* (2017) 14:9

Page 4 of 12

Table 1 Comparison of the types of cellular transcription factors and the number of their cognate binding sites on the antisense promoters of human-specific ERVK, HIV-1, and HTLV-1. Sequences of the consensus binding sites for cellular transcription factors predicted to bind these retroviral antisense LTRs are also shown

Transcription Factor	Consensus binding site sequence	Number of putative binding sites predicted in 3′ LTR		
		ERVK HS	HIV-1	HTLV-1
Activating Protein 1 (AP-1) c-Jun/c-Fos	TGA (G/C) TCA	5	3	2
Activating Protein 2α (AP-2α)	GCCNNNGGC	5	2	14
Activating Transcription Factor-3 (ATF-3)	T (T/G) ACGT (A/C) (A/G)	1	—	3
Androgen Receptor (AR)	GG (A/T) ACANNNTGTTCT (**ARE**)	1	—	—
Aryl hydrocarbon receptor / AhR nuclear translocator (AhR/ARNT)	TNGCGTG	2	3	4
cAMP Response Element Binding Protein (CREB)	TGACGTCA	—	—	2
CCAAT-Enhancer Binding Protein (C/EBP)	A (A/G/T) C (C/A) AAT	7	3	3
Cellular Myeloblastosis virus protein (c-Myb)	(T/C) AAC (G/T) G	12	7	8
Cellular Myelocytomatosis virus protein (c-Myc)	NNNCACGTGNN (**E-box**)	10	1	2
Chicken Ovalbumin Upstream Promoter Transcription Factor-1 (COUPT or COUPTF-1)	(A/G) G (G/T) TCA	7	15	13
E2 Factor-1 (E2F-1)	TTT (C/G) (C/G) CGC	—	—	2
Early B-cell Factor (EBF)	CCCNNGGG	9	4	7
EGFR-specific Transcription Factor (ETF)	G/C rich regions	14	4	19
Enkephalin Transcription Factor-1 (ENKTF-1)	TGGCGTA	1	—	—
Estrogen Receptor (ER-α; ER-β)	GGTCANNNTGACC (**ERE**)	1	—	—
E-twenty six (ETS-1; ETS-2)	GGA (A/T)	12	8	7
ETS-like Transcription Factor 1 (Elk-1; ELF-1)	CAGGAA (C/G) (**PU Box**)	5	1	2
Forkhead Box P3 (FOXP3)	(G/A) (T/C) AAACA	14	1	3

GATA binding proteins
  GATA - 1, 4, 5
  GATA - 2, 3
  GATA - 6
	(A/T) GATA (A/G) (**GATA Box**) (A/T) GAT (A/C) (A/G) (A/T) GAT (T/C) (A/G)	3
		6
		7
			2
			6
			5
				1
				2
				2

GC Binding Factor (GCF)	GCGGGGC	—	—	1
Glucocorticoid Receptor (GR-α; GR- β)	GGTACANNNTGTTC (**GRE**)	1	—	—
Hepatocyte Nuclear Factor-3/4 (HNF3; HNF4)	GGTCA repeats	3	—	3
Hepatocyte Nuclear Factor-1 (HNF1)	GTTAATNATTAAC	1	—	—
Hypoxia Inducible Factor-1 (HIF-1)	(A/G) CGTG (flanked by (A/C) ACAG)	3	1	2
Ikarose-1 (IK-1)	TGGGA (A/T)	16	5	6
Interferon Regulatory Factors (IRF)	GAAANN repeats (**ISRE**)	2	—	2
Lymphoid Enhancer-binding Factor 1 (LEF-1)	CTTTGAA	3	—	—
Myc Associated Zinc finger protein (MAZ)	GGGAGGG	9	1	6
Nuclear Factor of Activated T cells (NFAT)	GGAGAA	10	—	7
Nuclear Factor I (NFI; NFI/CTF)	TTGGCNNNNNGCCAA	—	—	2
Nuclear Factor Kappa B (NF-κB)	GG (G/A) (G/A) NN (C/T) (C/T) CC	9	7	6
Nuclear Factor-Y (NF-Y)	CCAAT	2	—	3
Polyomavirus Enhancer Activator 3 (PEA or PEA-3)	GGA (A/T)	12	8	7
Progesterone Receptor (PR-A; PR-B)	GNACANNNTGTNC (**PRE**)	1	—	—
Protein 53 (p53)	CATTAG	8	1	1
Recombination signal Binding Protein-Jκ (RBP-Jκ)	(C/T) GTGGGAA	1	—	—
Retinoid X Receptorα (RXRa)	AGGTCA	11	4	7
Serum Response Factor (SRF)	C (C/T) (A/T)$_6$GG	2	4	2

Manghera et al. Virology Journal (2017) 14:9

Page 5 of 12

Table 1 Comparison of the types of cellular transcription factors and the number of their cognate binding sites on the antisense promoters of human-specific ERVK, HIV-1, and HTLV-1. Sequences of the consensus binding sites for cellular transcription factors predicted to bind these retroviral antisense LTRs are also shown (Continued)

Sex-determining Region Y (SRY)	(A/T) (A/T) CAA (A/T)	12	4	3
Signal Transducers and Activators of Transcription (STAT)	TTCNNNNGAA	10	4	8
Specificity Protein-1 (Sp1)	GGGCGG	15	1	14
TATA Binding Protein (TBP)	TATAAA	1	—	1
T cell Factor-4E (TCF-4E)	(G/C) ATCAAAGG	1	—	—
Thyroid Hormone Receptor (TRβ)	TGAGGTCA (**TRE**)	1	—	1
Transcription Factor II-D (TFII-D)	TATAAA	1	—	1
Transcription Factor II-I (TFII-I)	CANNTG	13	3	1
Upstream Transcription Factor (USF-1, USF-2)	CACGTG (**E-box**)	10	1	2
Vitamin D Receptor (VDR)	G (G/T) TCA	25	20	21
X-box binding protein (XBP-1)	CCACG	9	9	9
Yin Yang 1 (YY1)	GCCATNTT	1	—	1

— indicates no putative binding sites
ARE androgen response element, E-box enhancer box, ERE estrogen response element, PU Box purine box, GRE glucocorticoid response element, ISRE interferon stimulated response element, PRE progesterone response element, TRE thyroid hormone response element

antisense promoter in HTLV-1 transformed cell lines, as well as in cells derived from patients with Adult T-cell Leukemia/Lymphoma (ATL) [3, 37]. Thus, these transcription factors are postulated to play a role in regulating antisense HTLV-1 transcription. Whether they promote or inhibit antisense gene expression remains to be elucidated. In addition, T-cell factor 1 (TCF-1) and Lymphoid enhancing factor 1 (LEF-1) have been shown to slightly enhance hbz transcription and HTLV-1 3′ LTR activation in luciferase assays [38]. In line with these studies, bioinformatics analysis of the consensus HTLV-1 3′ LTR has not only confirmed the presence of intact binding sites for the aforementioned transcription factors, but has also revealed putative sites for numerous other antisense transcriptional regulators (Fig. 2, Table 1). Interestingly, some of the identified binding sites for transcription factors, notably ATF, CREB, and NF-I, are unique to the 3′ LTR of HTLV-1, and are not predicted within the HIV-1 3′ LTR. Thus, HTLV-1 antisense gene expression is likely regulated by a multitude of cellular and retroviral transcription factors. There is an evident need for future research characterizing the transcriptional regulators that broadly and selectively modulate antisense gene expression in the various tissue types targeted by retroviruses.

The transcription of HTLV-1 proviruses is further modulated by the antisense-encoded HBZ protein. HBZ binding to the 5′ LTR of HTLV-1 promotes viral latency by suppressing sense transcription [39]. Conversely, HTLV-1 antisense transcription is positively regulated by HBZ. HBZ has the capacity to form heterodimers with a

cellular transcription factor JunD [36]. Co-expression of JunD and HBZ has been shown to significantly increase HTLV-1 3′ LTR activity in luciferase assays as compared to the expression of JunD or HBZ alone [36]. Also, luciferase activity was not enhanced with HBZ overexpression in knockout cells lacking JunD [36]. It was further shown that HBZ/JunD dimers are recruited to Sp1-bound regions of the HTLV-1 3′ LTR, due to the interaction of JunD with Sp1 [36]. Accordingly, mutation of one of these Sp1 sites in the HTLV-1 reporter construct, or the overexpression of Sp1 mutants lacking DNA-binding ability, resulted in a significant decrease in luciferase expression [36]. Therefore, HTLV-1 antisense transcription is regulated through interactions between HBZ, JunD, and Sp1 at the 3′ LTR.

As suggested for HIV-1 Tat interaction with its 3′ LTR, the HTLV-1 accessory protein Tax can also upregulate the proviral antisense transcription. Overexpression of Tax has been shown to markedly enhance luciferase activity from transiently expressed, as well as stably integrated, HTLV-1 3′ LTR reporter constructs in human cell lines [40]. Tax responsive elements (TxREs) containing near-consensus CREB binding sites have been reported in the HTLV-1 antisense promoter [3, 34, 40]. Mutations of these TxREs, which render them incapable of interacting with CREB, exhibited a dramatically reduced luciferase activity from the 3′ LTR in the presence of Tax [34, 40]. Thus, viral Tax protein has been shown to drive HTLV-1 antisense transcription by cooperating with CREB at TxREs at the 3′ LTR. In stark contrast, several reports using similar methodology, but different host cells, have not detected Tax-mediated

Manghera et al. Virology Journal (2017) 14:9

Page 6 of 12

Fig. 2 In silico examination of transcription factor binding sites and response elements within the representative HTLV-1 3′ LTR using ALGGEN-PROMO software [73]. The prototypic HTLV-1 3′ LTR sequence used was obtained from GenBank, accession number AB513134 (B1033-2009 isolate). The known binding sites for AP-1, ATF, C/EBP, CREB, GATA-2, NFI, and Sp1 in the HTLV-1 3′ LTR are flagged with an asterisk. Transcription factor binding sites unique to HTLV-1 are indicated in green. Initiator motifs (INRs) are indicated in pink. ISRE sites are indicated in blue. Abbreviations used include: INR = initiator motif, ISRE = interferon stimulated response element, and TRE = thyroid hormone response element, and TxRE = tax responsive element. Sequence annotations were performed using Geneious software [68]

regulation of viral antisense transcription [23, 32, 41]. Thus, the discrepancies between these studies suggests that Tax-mediated regulation of antisense gene expression likely depends on the cell type being investigated, and consequently, the availability of cell-specific transcription factor complexes required for this process. This would be consistent with similar cell-type specific observations surrounding Tax-dependent transactivation of HTLV-1 sense transcription [41].

Antisense transcription among other exogenous retroviruses

Antisense transcription is not exclusive to HIV and HTLV, and has also been reported in the deltaretroviruses bovine leukemia virus (BLV) [42] and simian T-cell leukemia virus (STLV) [43], the lentiviruses feline immunodeficiency virus (FIV) [44] and bovine immunodeficiency virus (BIV) [45], as well as the gammaretrovirus murine leukemia virus (MLV) [46]. However, the regulation of antisense transcription remains poorly studied in these retroviruses. A recent report has demonstrated that the antisense transcription of BLV, a close relative of HTLV-1, is regulated by an Interferon Regulatory Factor (IRF) binding site and an E-box in its 3′ LTR [42]. Through BLV 3′ LTR luciferase reporter assays, mutation of this IRF binding site or the E-box resulted in modest to significant downregulation of antisense luciferase activity, respectively. Bioinformatics analysis has revealed the presence of two putative intact IRF binding sites in the HTLV-1, but not HIV-1, representative 3′ LTR, as well the presence of intact E-boxes in both antisense promoters (Fig. 2, Table 1). This suggests that IRF may regulate the antisense transcription of select retroviruses, whereas E-boxes may be a broader feature of retroviral 3′ LTRs.

Manghera *et al. Virology Journal* (2017) 14:9

Page 7 of 12

Antisense transcription among endogenous retroviruses

Antisense transcription has been emerging as a common, but generally underappreciated, feature of ERV gene expression patterns. Several human ERVs, particularly ERV9 and ERVK loci, exhibit transcription from the antisense strand. Above and beyond the potential of antisense products to modulate endogenous retrovirus expression patterns, the impact of antisense viral products on human biology is becoming apparent. Most notably, antisense transcription of ERVs may play important roles in the regulation of human gene expression or modulation of cellular pathways.

ERV9 antisense transcription

Among human endogenous retroviruses, antisense transcriptional regulation of ERV9 loci is the best understood. Several cellular transcription factors are known to induce the expression of antisense RNA from the U3 region (referred to as the U3 AS RNA) of the ERV9 LTR. This includes CREB, glucocorticoid receptors (GR), IRF, signal transducers and activators of transcription (STAT), and activating protein 2 (AP-2) [47]. Interestingly, the AUUGG motifs within the ERV9 antisense transcripts have been experimentally demonstrated to interact with and sequester select cellular transcription factors – NF-Y, p53 and Sp1 [47]. We have predicted the presence of similar motifs in the antisense RNA originating from the ERVK 3' LTR (data not shown). By sequestering the aforementioned cellular transcription factors, ERV9 U3 AS RNA serves to repress the expression of genes involved in cell cycle activation, thereby inhibiting uncontrolled cellular proliferation. Accordingly, deregulation of this ERV-derived antisense RNA has the potential to promote tumor formation and propagation [47]. Thus, the production of endogenous AS RNA decoys may be an important phenomenon among endogenous retroviruses, and may serve essential regulatory and protective functions for their human hosts.

ERVK antisense transcription

The human genome is ubiquitously populated with ERVK sequences including solitary LTRs and partial proviral sequences, as well as full-length proviruses. Solitary LTRs are the most abundant ERVK elements within the human genome, and are estimated to number over 25,000 [48]. They are frequently present in close proximity to our genes, and therefore may be involved in the regulation of neighbouring genes by acting as promoters or enhancers. It is estimated that at least 50% of human-specific ERVK (HML-2) LTRs serve as promoters for the transcription of human genes [49]. ERVK LTRs have been experimentally shown to activate the expression of promoter-less reporter genes in luciferase assays when

inserted in both forward and reverse orientations, indicating their bidirectional promoter activity [48]. Such bidirectional activity lends plausibility to antisense viral RNA transcription mediated by the 3' LTR of ERVK.

Recently, several ERVK loci present outside human intronic regions have been demonstrated to exhibit transcription of the proviral antisense strand in prostate cancer cell lines. These include ERVK(I), ERVK-106, an un-named ERVK within locus 7q34, and multiple loci of solo LTRs [50]. When inserted in an opposite transcriptional orientation to that of their host intron, antisense transcription of ERVK proviruses can be explained as a consequence of host gene transcription. In contrast, the basis of transcription of the antisense strands of ERVK loci, such as ERVK-106, situated outside of human genes remains unclear. Though the regulation of antisense transcription driven by the ERVK 3' LTR is poorly understood, it is likely mediated by a complex of TFs binding to the 3' LTR (Fig. 3).

As a first step to better understand putative antisense ERVK transcription from its 3' LTR, we performed an extensive bioinformatics analysis of 92 full-length ERVK HML-2 sequences, and predicted intact and conserved binding sites for numerous human transcription factors within 3' LTRs of human-specific ERVK HML-2 proviruses (Fig. 3 shows transcription factor binding sites on five prototypic ERVK 3' LTRs, and Table 1). Similar to antisense promoters of other retroviruses, conserved signatures of ERVK 3' LTR include absence of a TATA-box and the presence of multiple conserved INR motifs scattered throughout the LTR (Fig. 3). This suggests that these putative alternative core promoter elements may initiate transcription from the ERVK proviral antisense strand at multiple sites. Additionally, in the absence of a TATA box (TATAAA), select subtypes of exogenous retroviruses, such as HIV-1 subtype E, have been shown to utilize a CATA box (CATAAA) to initiate proviral gene transcription [51]. Since the ERVK 3' LTR contains a conserved putative CATA box (Fig. 3), ERVK may similarly use this promoter element to initiate antisense transcription.

It is noteworthy that we identified multiple Sp1 binding sites in the ERVK 3' LTR, as they are critical for inducing transcription from TATA-less promoters. This is due to Sp1 recruitment of transcription factor II-D (TFII-D) which promotes the formation of transcription initiation complexes [16]. Since Sp1 binding to both HIV-1 and HTLV-1 antisense LTRs activates expression of their respective antisense proteins, this ubiquitous transcription factor may also have a key role in driving antisense transcription from the TATA-less ERVK 3' LTR. Indeed, the ERVK 3' LTR is laden with multiple conserved potential Sp1 binding sites; a total of 15 were identified, a similar density to that found of the HTLV-1
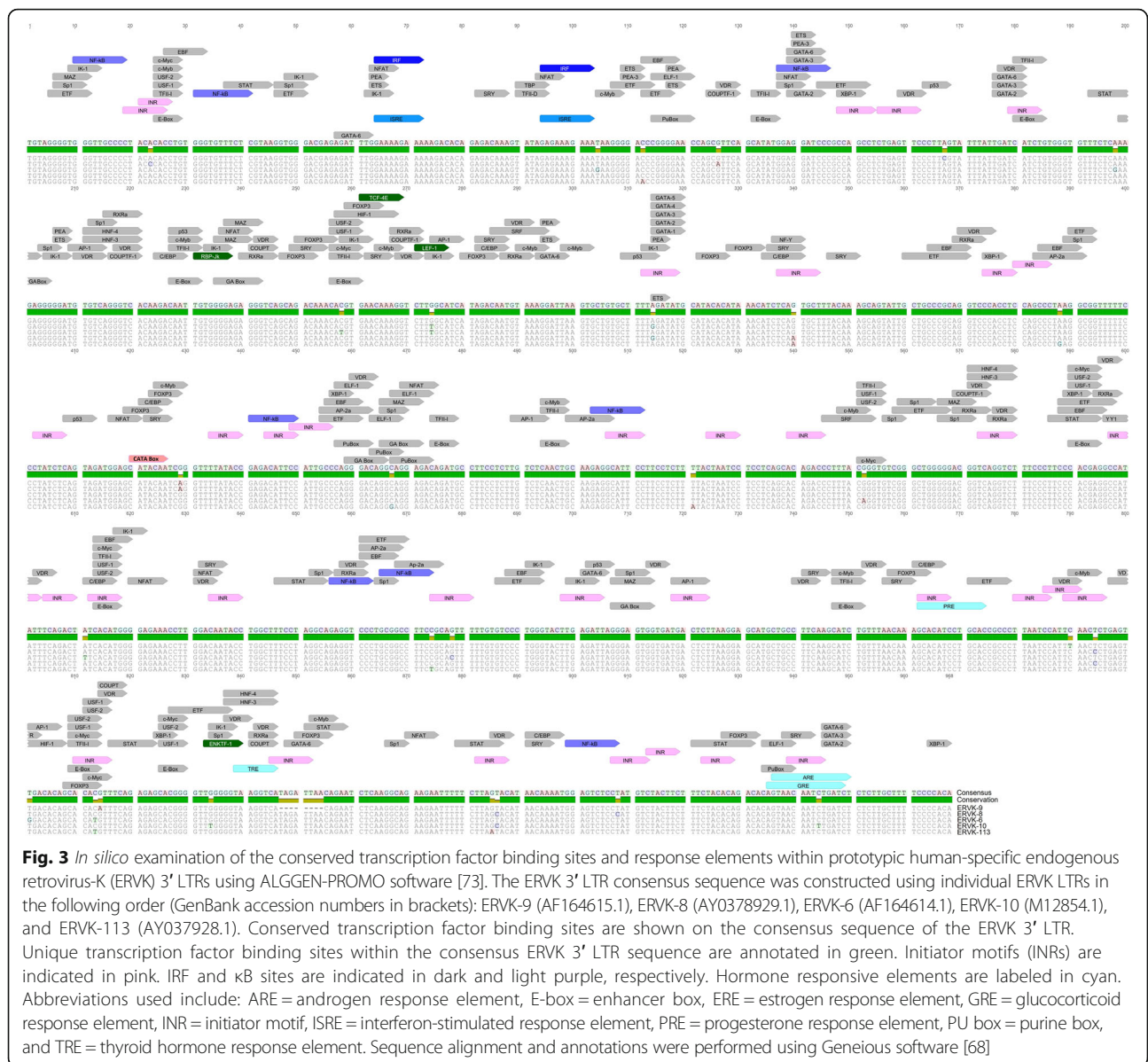
Manghera *et al. Virology Journal* (2017) 14:9

Page 8 of 12

Fig. 3 *In silico* examination of the conserved transcription factor binding sites and response elements within prototypic human-specific endogenous retrovirus-K (ERVK) 3′ LTRs using ALGGEN-PROMO software [73]. The ERVK 3′ LTR consensus sequence was constructed using individual ERVK LTRs in the following order (GenBank accession numbers in brackets): ERVK-9 (AF164615.1), ERVK-8 (AY0378929.1), ERVK-6 (AF164614.1), ERVK-10 (M12854.1), and ERVK-113 (AY037928.1). Conserved transcription factor binding sites are shown on the consensus sequence of the ERVK 3′ LTR. Unique transcription factor binding sites within the consensus ERVK 3′ LTR sequence are annotated in green. Initiator motifs (INRs) are indicated in pink. IRF and κB sites are indicated in dark and light purple, respectively. Hormone responsive elements are labeled in cyan. Abbreviations used include: ARE = androgen response element, E-box = enhancer box, ERE = estrogen response element, GRE = glucocorticoid response element, INR = initiator motif, ISRE = interferon-stimulated response element, PRE = progesterone response element, PU box = purine box, and TRE = thyroid hormone response element. Sequence alignment and annotations were performed using Geneious software [68]

antisense promoter (Fig. 3). The ERVK antisense promoter harbors putative binding sites for other transcription factors known to induce the activity of 3′ LTRs of exogenous retroviruses and ERV9. These include docking sequences for STAT, AP-2, AP-1, USF, GATA, NF-Y, ATF, CBP, TCF, LEF, and E-box (Fig. 3).

In addition, the ERVK 3′ LTR contains multiple putative NF-κB binding sites (Fig. 3). Thus, this pro-inflammatory transcription factor may drive ERVK antisense transcription under select conditions, as documented for exogenous retroviruses [17, 26]. Another interesting feature of the ERVK 3′ LTR is the presence of two conserved consensus interferon stimulated response elements (ISREs). We have recently demonstrated the ability of IRF1 and NF-κB p65/p50 to synergistically enhance transcription

from the ERVK sense promoter in the presence of select pro-inflammatory cytokines [52, 53]. Of note, ERVK 3′ LTR ISRE sequences are more similar to canonical ISREs as compared to their 5′ LTR counterparts, suggesting stronger IRF/NF-κB binding potential [54, 55]. Thus, inflammatory stimuli that enhance the activity of IRFs and NF-κB have the potential to provide an additional level of regulation on the ERVK antisense transcriptome.

Bioinformatics analysis further revealed select transcription factor binding profiles unique to ERVK antisense promoters, notably the presence of hormone responsive elements that were absent from both HIV-1 and HTLV-1 3′ LTRs. This includes the presence of putative binding sites for androgen (AR), estrogen (ER), glucocorticoid (GR), and progesterone (PR) receptors (Fig. 3, Table 1).

Manghera *et al. Virology Journal* (2017) 14:9

Page 9 of 12

Since these hormonal receptors are known to drive the activity of the ERVK 5′ LTR [53, 56, 57], they may also modulate the activity of its 3′ LTR.

The TF profile of the ERVK 3′ LTR may also point to cell-specific activation of antisense transcription. As with the HTLV-1 3′ LTR, the ERVK 3′ LTR contains putative binding sites for FOXP3, a transcription factor specific to regulatory T cells. Expression of antisense HTLV-1 transcripts in T regulatory cells is associated with the development of Adult T-cell Leukemia (ATL) [3, 24], potentially suggesting a shared mechanism for ERVK-associated leukemia [58]. In addition, several GATA family transcription factor binding sites are found within the ERVK antisense LTR. Owing to the importance of GATA family transcription factors in regulating immune cells [59], our data indicate that antisense ERVK expression may be modulated in hematopoietic cells, Moreover, impairment of GATA transcription factors are a hallmark of many cancers [60]. If ERVK were to employ an antisense product whose expression was i) driven by GATA proteins, and ii) held a similar latency-inducing function as HTLV-1 HBZ, the lack of GATA protein expression in cancers could explain the enhanced expression of sense-encoded ERVK protein products in transformed cells [58, 61].

Understanding the activity of the ERVK 3′ LTR promoter may be the key to elucidating the basis of antisense transcription of endogenous proviruses; however, it should be noted that not all ERVK LTRs are equally intact. Evaluation of human-specific and older 3′ ERVK LTRs in the HML-2 family reveals conserved, alternative and unique TF binding site profiles, when comparing recent and older provirus LTRs (data not shown). Therefore, developing an understanding of ERVK antisense transcription, especially in the context of specific genomic loci, is an area of research that clearly requires more investigation.

ERVK genome harbors ORFs for putative antisense proteins

Since the ERVK antisense promoter contains conserved putative enhancer elements and consensus binding sites for numerous human transcription factors, it puts forth the question as to whether the ERVK antisense genomic strand contains open reading frames (ORFs) for putative antisense proteins. We have been able to identify conserved regions of ERVK antisense genome that resemble motifs found within glycosyltransferases (GTs) and thioredoxin/thioredoxin reductase (TRX) complexes (data not shown). Interestingly, one of these conserved motifs is located in a region complementary to the sense strand of ERVK env – a position similar to that of the open reading frames for *hbz* in HTLV-1, *aph-2* in HTLV-2, and *asp*

in HIV-1 [9]. Interestingly, the production of viral-derived GTs or TRXs would be consistent with the needs of viruses [62], and more specifically retroviruses [63–65].

However, due to several limitations, it is currently difficult to predict with confidence whether the ERVK genome encodes antisense products. Notably, the primary structures of GTs and TRXs are extremely diverse and lack signature features [66, 67]. This lack of specificity in the predicted protein domains creates further issue for sequence alignment and does not lend assurance to current predictions without further bioinformatic and experimental investigation. It would further be worthwhile to employ whole transcriptome sequencing to examine the production of ERVK antisense transcripts in tissue specimens from patients with ERVK-associated diseases versus healthy controls. In the future, techniques developed to study antisense transcription in exogenous retroviruses will be useful in characterizing the expression of ERV antisense genomes.

Conclusions
In the light of this report, further research on antisense transcription in endogenous retroviruses is warranted. We have shown that the exogenous and endogenous antisense LTRs share many regulatory similarities. Thus, it would be interesting to examine whether regulatory and pathological processes associated with exogenous retroviral antisense transcription are also applicable to ERVs. The presence of potentially new antisense-encoded transcripts and proteins would provide a more complete understanding of the biology of endogenous retroviruses, such as ERVK, and their roles in health and disease. A reconsideration of the nature of exogenous, as well as endogenous, retroviral transcription is required for a better understanding of *Retroviridae* as a whole.

Methodology
The sequences of antisense promoters (3′ LTR) of exogenous retroviruses (HIV-1 and HTLV-1) and endogenous retrovirus-K (HML2) were obtained from GenBank, and reverse complemented in Geneious [68]. For HIV-1 and HLTV-1 3′ LTRs, the prototypic sequences used were that of the HXB2 and B1033-2009 strains, respectively, as these are the most commonly used reference sequences for these exogenous retroviruses [69–71]. The 92 ERVK (HML2) 3′ LTRs analyzed were grouped into human-specific or old sequences [72]. These were aligned separately in Geneious and a consensus sequence was obtained for each of the two groups. The human specific ERVK 3′ LTRs were further refined into five prototypic sequences, as each of the remaining ERVK 3′ LTRs exhibited transcription factor binding site patterns similar to one of these prototypic antisense promoters. These prototypic LTRs were aligned in

Manghera *et al. Virology Journal* (2017) 14:9

Page 10 of 12

Geneious-R6® software (version 6.1.7), and a consensus sequence was obtained. The binding sites for human-specific transcription factors within consensus HIV-1, HTLV-1, and ERVK (HML2) 3′ LTR sequences were predicted through ALGGEN PROMO database, which uses version 8.3 of TRANSFAC [73]. PROMO can be accessed at http://alggen.lsi.upc.es/cgi-bin/promo_v3/promo/promoinit.cgi?dirDB=TF_8.3. The search parameters used were: factor's species – *Homo sapiens*, and site's species – *Homo sapiens*. Each binding site for a given transcription factor was compared to the sequence of its known consensus binding site (listed in Table 1). The consensus binding sites for transcription factors predicted to interact with these retroviral promoters have been previously described [53]. The consensus binding sites for ATF3, AhR/ARNT, COUPTF1, E2F1, ETF, ENKTF1, FOXP3, GATA family, GCF, HNF, HIF, NF-Y, RxRα, SRF, TCF-4E, and TRβ were obtained from [74–90]. Only those sites with a maximum of one base pair deviation from the consensus binding sequence (or two for large hormonal response elements) were annotated on the target retroviral 3′ LTR. All annotations were performed in Geneious-R6®.

Abbreviations
ASP: Antisense protein; ATL: Adult T-cell leukemia; BIV: Bovine immunodeficiency virus; BLV: Bovine leukemia virus; ERV: Endogenous retrovirus; FIV: Feline immunodeficiency virus; GT: Glycosyltransferase; HBZ: HTLV-1 bZIP factor; HIV: Human immunodeficiency virus; HTLV: Human T-cell lymphotropic virus; INR: Initiator motif; ISRE: Interferon stimulated response element; LTR: Long terminal repeat; MLV: Murine leukemia virus; STLV: Simian T-cell leukemia virus; TNFα: Tumor necrosis factor α; TRX: Thioredoxin reductase

Authors' contributions
MM performed transcription factor binding site analyses in ALGGEN PROMO and annotations in Geneious. AM performed the analyses of ORFs in ERVK antisense genomic strands. RND conceived the study. MM, AM, and RND wrote the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Not applicable.

References
1. Bet A, Maze EA, Bansal A, Sterrett S, Gross A, Graff-Dubois S, et al. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. Retrovirology. 2015;12:15.
2. Cassan E, Arigon-Chifolleau A-M, Mesnard J-M, Gross A, Gascuel O. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. Proc Natl Acad Sci USA. 2016;113(41):1–6.
3. Ma G, Yasunaga J-I, Matsuoka M. Multifaceted functions and roles of HBZ in HTLV-1 pathogenesis. Retrovirology BioMed Central. 2016;13:1–9.
4. Panfil AR, Dissinger NJ, Howard CM, Murphy BM, Landes K, Fernandez SA, et al. Functional comparison of HBZ and the related APH-2 protein provides insight into human T-cell leukemia virus type 1. J Virol. 2016;90:3760–72.
5. Miller R. Human immunodeficiency virus may encode a novel protein on the genomic DNA plus strand. Science. 1988;239:1420–2.
6. Landry S, Halin M, Lefort S, Audet B, Vaquero C, Mesnard J-M, et al. Detection, characterization and regulation of antisense transcripts in HIV-1. Retrovirology. 2007;4:71.
7. Ludwig LB, Ambrus JL, Krawczyk KA, Sharma S, Brooks S, Hsiao C-B, et al. Human immunodeficiency virus-type 1 LTR DNA contains an intrinsic gene producing antisense RNA and protein products. Retrovirology. 2006;3:80.
8. Kobayashi-Ishihara M, Yamagishi M, Hara T, Matsuda Y, Takahashi R, Miyake A, et al. HIV-1-encoded antisense RNA suppresses viral replication for a prolonged period. Retrovirology. 2012;9:38.
9. Clerc I, Laverdure S, Torresilla C, Landry S, Borel S, Vargas A, et al. Polarized expression of the membrane ASP protein derived from HIV-1 antisense transcription in T cells. Retrovirology. 2011;8:74.
10. Torresilla C, Mesnard J, Barbeau B. Reviving an old HIV-1 gene: the HIV-1 antisense protein. Curr HIV Res. 2015;13:117–24.
11. Vanhee-Brossollet C, Thoreau H, Serpente N, D'Auriol L, Levy J-P, Vaquero C. A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. Virology. 1995;206:196–202.
12. Torresilla C, Larocque É, Landry S, Halin M, Coulombe Y, Masson J-Y, et al. Detection of the HIV-1 minus-strand-encoded antisense protein and its association with autophagy. J Virol. 2013;87:5089–105.
13. Barbeau B, Devaux C, Mesnard J-M. Antisense transcription in human T-cell leukemia virus type 1: discovery of a new viral gene. In: Lever AM, Jeang K-T, Berkhout B, editors. Recent Adv. Hum. retroviruses Princ. replication Pathog. Singapore: World Scientific Publishing Company; 2010. p. 105–27.
14. Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat Rev Genet. 2007;8:424–36.
15. Peeters A, Lambert PF. A fourth Sp1 site in the human immunodeficiency virus type 1 long terminal repeat is essential for negative-sense transcription. J Virol. 1996;70:6665–72.
16. Lin S, Zhang L, Luo W, Zhang X. Characteristics of antisense transcript promoters and the regulation of their activity. Int J Mol Sci. 2015;17:1–17.
17. Michael NL, Vahey MT, Arcy LD, Ehrenberg PK, Mosca JD, Rappaport JAY, et al. Negative-strand RNA transcripts are produced in human immunodeficiency virus type 1-infected cells and patients by a novel promoter downregulated by Tat. J Virol. 1994;68:979–87.
18. Nonnemacher MR, Pirrone V, Feng R, Moldover B, Passic S, Aiamkitsumrit B, et al. HIV-1 promoter single nucleotide polymorphisms are associated with clinical disease severity. PLoS One. 2016;11:e0150835.
19. Shah S, Alexaki A, Pirrone V, Dahiya S, Nonnemacher MR, Wigdahl B. Functional properties of the HIV-1 long terminal repeat containing single-nucleotide polymorphisms in Sp site III and CCAAT / enhancer binding protein site I. Virol J. 2014;11:92.
20. Jeeninga RE, Hoogenkamp M, Armand-ugon M, Baar MDE, Verhoef K, Berkhout BEN. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes a through G. J Virol. 2000;74:3740–51.
21. Verhoef K, Sanders RW, Fontaine V, Kitajima S, Berkhout BEN. Evolution of the human immunodeficiency virus type 1 long terminal repeat promoter by conversion of an NF-kB enhancer element into a GABP binding site. J Virol. 1999;73:1331–40.

Manghera *et al. Virology Journal* (2017) 14:9

Page 11 of 12

22. Laverdure S, Gross A, Arpin-André C, Clerc I, Beaumelle B, Barbeau B, et al. HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells. J Virol. 2012;86:13785–9.

23. Arpin-André C, Laverdure S, Barbeau B, Gross A, Mesnard J-M. Construction of a reporter vector for analysis of bidirectional transcriptional activity of retrovirus LTR. Plasmid. 2014;74:45–51.

24. Satou Y, Yasunaga J-I, Zhao T, Yoshida M, Miyazato P, Takai K, et al. HTLV-1 bZIP factor induces T-cell lymphoma and systemic inflammation in vivo. PLoS Pathog. 2011;7:e1001274.

25. Miyazato P, Matsuo M, Katsuya H, Satou Y. Transcriptional and epigenetic regulatory mechanisms affecting HTLV-1 provirus. Viruses. 2016;8:1–14.

26. Barbeau B, Mesnard J-M. Does chronic infection in retroviruses have a sense? Trends Microbiol Elsevier Ltd. 2015;23:367–75.

27. Sugata K, Yasunaga J-I, Kinosada H, Mitobe Y, Furuta R, Mahgoub M, et al. HTLV-1 viral factor HBZ induces CCR4 to promote T-cell migration and proliferation. Cancer Res. 2016;76:5068–79.

28. Barbeau B, Mesnard J-M. Making sense out of antisense transcription in human T-cell lymphotropic viruses (HTLVs). Viruses. 2011;3:456–68.

29. Larocque É, Halin M, Landry S, Marriott SJ, Switzer WM, Barbeau B. Human T-cell lymphotropic virus type 3 (HTLV-3)- and HTLV-4-derived antisense transcripts encode proteins with similar Tax-inhibiting functions but distinct subcellular localization. J Virol. 2011;85:12673–85.

30. Switzer WM, Salemi M, Qari SH, Jia H, Gray RR, Katzourakis A, et al. Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. Retrovirology. 2009;6:1–20.

31. Cavanagh M-H, Landry S, Audet B, Arpin-André C, Hivin P, Paré M-E, et al. HTLV-I antisense transcripts initiating in the 3′LTR are alternatively spliced and polyadenylated. Retrovirology. 2006;3:15.

32. Larocca D, Chao L, Seto H, Brunck T. Human T-cell leukemia virus minus strand transcription in infected T-cells. Biochem Biophys Res Commun. 1989;163:1006–13.

33. Murata K, Hayashibara T, Sugahara K, Uemura A, Yamaguchi T, Harasawa H, et al. A novel alternative splicing isoform of human T-cell leukemia virus type 1 bZIP factor (HBZ-SI) targets distinct subnuclear localization. J Virol. 2006;80:2495–505.

34. Yoshida M, Satou Y, Yasunaga J-I, Fujisawa J-I, Matsuoka M. Transcriptional control of spliced and unspliced human T-cell leukemia virus type 1 bZIP factor (HBZ) gene. J Virol. 2008;82:9359–68.

35. Satou Y, Yasunaga J, Yoshida M, Matsuoka M. HTLV-I basic leucine zipper factor gene mRNA supports proliferation of adult T cell leukemia cells. PNAS. 2006;103:1–6.

36. Gazon H, Lemasson I, Polakowski N, Césaire R, Matsuoka M, Barbeau B, et al. Human T-cell leukemia virus type 1 (HTLV-1) bZIP factor requires cellular transcription factor JunD to upregulate HTLV-1 antisense transcription from the 3′ long terminal repeat. J Virol. 2012;86:9070–8.

37. Lemasson I, Polakowski NJ, Laybourn PJ, Nyborg JK. Transcription regulatory complexes bind the human T-cell leukemia virus 5 J and 3 J long terminal repeats to control gene expression. Mol Cell Biol. 2004;24:6117–26.

38. Ma G, Yasunaga J, Akari H, Matsuoka M. TCF1 and LEF1 act as T-cell intrinsic HTLV-1 antagonists by targeting Tax. Proc Natl Acad Sci U S A. 2015;112: 2216–21.

39. Gaudray G, Gachon F, Basbous J, Biard-piechaczyk M, Devaux C, Mesnard J. The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription. J Virol. 2002;76:12813–22.

40. Landry S, Halin M, Vargas A, Lemasson I, Mesnard J-M, Barbeau B. Upregulation of human T-cell leukemia virus type 1 antisense transcription by the viral tax protein. J Virol. 2009;83:2048–54.

41. Laverdure S, Polakowski N, Hoang K, Lemasson I. Permissive sense and antisense transcription from the 5′ and 3′ long terminal repeats of human T-cell leukemia virus type 1. J Virol. 2016;90:3600–10.

42. Durkin K, Rosewick N, Artesi M, Hahaut V, Griebel P, Arsic N, et al. Characterization of novel Bovine Leukemia Virus (BLV) antisense transcripts by deep sequencing reveals constitutive expression in tumors and transcriptional interaction with viral microRNAs. Retrovirology. 2016;13:33.

43. Miura M, Yasunaga J, Tanabe J, Sugata K, Zhao T, Ma G, et al. Characterization of simian T-cell leukemia virus type 1 in naturally infected Japanese macaques as a model of HTLV-1 infection. Retrovirology. 2013;10:118.

44. Vaquero C, Briquet S, Richardson J, Vanhe C. Natural antisense transcripts are detected in different cell lines and tissues of cats infected with feline immunodeficiency virus. Gene. 2001;267:157–64.

45. Liu B, Zhao X, Shen W, Kong X. Evidence for the antisense transcription in the proviral R29-127 strain of bovine immunodeficiency virus. Virol Sin. 2015;30:224–7.

46. Rasmussen MH, Ballarín-González B, Liu J, Lassen LB, Füchtbauer A, Füchtbauer E-M, et al. Antisense transcription in gammaretroviruses as a mechanism of insertional activation of host genes. J Virol. 2010;84:3780–8.

47. Xu L, Elkahloun AG, Candotti F, Grajkowski A, Beaucage SL, Petricoin EF, et al. A novel function of RNAs arising from the long terminal repeat of human endogenous retrovirus 9 in cell cycle arrest. J Virol. 2013;87:25–36.

48. Domansky AN, Kopantzev EP, Snezhkov EV, Lebedev YB, Leib-mosch C, Sverdlov ED. Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. FEBS Lett. 2000;472:191–5.

49. Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, Sverdlov E. At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active promoters for host nonrepetitive DNA transcription. J Virol. 2006;80:10752–62.

50. Agoni L, Guha C, Lenz J. Detection of human endogenous retrovirus K (HERV-K) transcripts in human prostate cancer cell lines. Front Oncol. 2013;3:180.

51. Van Opijnen T, Kamoschinski J, Jeeninga RE, Berkhout B. The human immunodeficiency virus type 1 promoter contains a CATA Box instead of a TATA box for optimal transcription and replication. J Virol. 2004;78:6883–90.

52. Manghera M, Ferguson-Parry J, Lin R, Douville RN. NF-κB and IRF1 induce endogenous retrovirus K expression via interferon-stimulated response elements in its 5′ long terminal repeat. J Virol. 2016;90:9338–49.

53. Manghera M, Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? Retrovirology. 2013;10:16.

54. Grandvaux N, Servant MJ, Sen GC, Balachandran S, Barber GN, Lin R, et al. Transcriptional profiling of interferon regulatory factor 3 target genes: direct involvement in the regulation of interferon-stimulated genes. J Virol. 2002; 76:5532–9.

55. Lin R, Heylbroeck C, Genin P, Pitha PM, Hiscott J, Al LINET, et al. Essential role of interferon regulatory factor 3 in direct activation of RANTES chemokine transcription. Mol Cell Biol. 1999;19:959–66.

56. Ono M, Kawakami M, Ushikubo H. Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D. J Virol. 1987;61:2059–62.

57. Hanke K, Chudak C, Kurth R, Bannert N. The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT). Int J Cancer. 2013; 132:556–67.

58. Downey RF, Sullivan FJ, Wang-Johanning F, Ambs S, Giles FJ, Glynn SA. Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? Int J Cancer. 2015;137:1249–57.

59. Gao J, Chen Y-H, Peterson LC. GATA family transcriptional factors: emerging suspects in hematologic disorders. Exp Hematol Oncol BioMed Central. 2015;4:28.

60. Zheng R, Blobel GA. GATA transcription factors and cancer. Genes Cancer. 2010;1:1178–88.

61. Suntsova M, Garazha A, Ivanova A, Kaminsky D, Zhavoronkov A, Buzdin A. Molecular functions of human endogenous retroviruses in health and disease. Cell Mol life Sci Springer Basel. 2015;72:3653–75.

62. Markine-Goriaynoff N, Gillet L, Van Etten JL, Korres H, Verma N, Vanderplasschen A. Glycosyltransferases encoded by viruses. J Gen Virol. 2004;85:2741–54.

63. Ueno M, Masutani H, Arai RJ, Yamauchi A, Hirota K, Sakai T, et al. Thioredoxin-dependent redox regulation of p53-mediated p21 activation. J Biol Chem. 1999;274:35809–15.

64. Masutani H, Ueda S, Yodoi J. The thioredoxin system in retroviral infection and apoptosis. Cell Death Differ. 2005;12:991–8.

65. Crispin M, Doores K. Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. Curr Opin Virol. 2015;11:63–9.

66. Atkinson HJ, Babbitt PC. An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. PLoS Comput Biol. 2009;5: e1000541.

67. Varki A, Cummings R, Esko J, Freeze H, Stanley P, Bertozzi C, et al. Essentials of glycobiology. 2nd ed. Cold Spring Harbor: Cold Spring Harbor Larboratory Press; 2009.

68. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for

Manghera *et al. Virology Journal* (2017) 14:9

Page 12 of 12

the organization and analysis of sequence data. Bioinformatics. 2012;28: 1647–9.

69. Ko GM, Reddy S, Kumar S, Bailey BA, Garg R. Computational analysis of HIV-1 protease protein binding pockets. J Chem Inf Model. 2010;50:1759–71.

70. Korber B, Gaschen B, Yusim K, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull. 2001;58:19–42.

71. Pessôa R, Watanabe JT, Nukui Y, Pereira J, Casseb J, Kasseb J, et al. Molecular characterization of human T-cell lymphotropic virus type 1 full and partial genomes by Illumina massively parallel sequencing technology. PLoS One. 2014;9:e93374.

72. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology. 2011;8:1–22.

73. Messeguer X, Escudero R, Farre D, Nunez O, Martinez J, Alba MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinforma Appl Note. 2002;18:333–4.

74. Zhao J, Li X, Guo M, Yu J, Yan C. The common stress responsive transcription factor ATF3 binds genomic sites enriched with p300 and H3K27ac for transcriptional regulation. BMC Genomics. 2016;17:335.

75. Patel RD, Kim DJ, Peters JM, Perdew GH. The aryl hydrocarbon receptor directly regulates expression of the potent mitogen epiregulin. Toxicol Sci. 2006;89:75–82.

76. Montemayor C, Montemayor OA, Ridgeway A, Lin F, Wheeler DA, Pletcher SD, et al. Genome-wide analysis of binding sites and direct target genes of the orphan nuclear receptor NR2F1/COUP-TFI. PLoS One. 2010;5:e8910.

77. Bieda M, Xu X, Singer MA, Green R, Farnham PJ. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. Genome Res. 2006;16:595–605.

78. Kageyama R, Merlino GT, Pastan I. Nuclear factor ETF specifically stimulates transcription from promoters without a TATA Box. J Biol Chem. 1989;264: 15508–14.

79. Comb M, Mermod N, Hyman SE, Pearlberg J, Ross ME, Goodman HM. Proteins bound at adjacent DNA elements act synergistically to regulate human proenkephalin cAMP inducible transcription. EMBO J. 1988;7:3793–805.

80. Koh KP, Sundrud MS, Rao A. Domain requirements and sequence specificity of DNA binding for the forkhead transcription factor FOXP3. PLoS One. 2009;4:1–9.

81. Lowry JA, Atchley WR. Molecular evolution of the GATA family of transcription factors : conservation within the DNA-binding domain. J Mol Evol. 2000;50:103–15.

82. Murakami A, Ishida S, Dickson C. GATA-4 interacts distinctively with negative and positive regulatory elements in the Fgf-3 promoter. Nucleic Acids Res. 2002;30:1056–64.

83. Wang L, Hunt KE, Martin GM, Oshima J. Structure and function of the human Werner syndrome gene promoter: evidence for transcriptional modulation. Nucleic Acids Res. 1998;26:3480–5.

84. Gardner-Stephen DA, Gregory PA, Mackenzie PI. Identification and characterization of functional hepatocyte nuclear factor 1-binding sites in UDP-glucuronosyltransferase genes. Methods Enzymol. 2005;400:22–46.

85. Mojsilovic-Petrovic J, Callaghan D, Cui H, Dean C, Stanimirovic DB, Zhang W. Hypoxia-inducible factor-1 (HIF-1) is involved in the regulation of hypoxia-stimulated expression of monocyte chemoattractant protein-1 (MCP-1/CCL2) and MCP-5 (Ccl12) in astrocytes. J Neuroinflammation. 2007;4:12.

86. Xu H, Fu J, Ha S-W, Ju D, Zheng J, Li L, et al. The CCAAT box-binding transcription factor NF-Y regulates basal expression of human proteasome genes. Biochim Biophys Acta. 2012;1823:818–25.

87. Rastinejad F, Wagner T, Zhao Q, Khorasanizadeh S. Structure of the RXR –RAR DNA-binding complex on the retinoic acid response element DR1. EMBO J. 2000;19:1045–54.

88. Kasza A, Wyrzykowska P, Horwacik I, Tymoszuk P, Mizgalska D, Palmer K, et al. Transcription factors Elk-1 and SRF are engaged in IL1-dependent regulation of ZC3H12A expression. BMC Mol Biol. 2010;11:14.

89. Cadigan KM, Waterman ML. TCF / LEFs and Wnt signaling in the nucleus. Cold Spring Harb Perspect Biol. 2012;4:a007906.

90. Ayers S, Switnicki MP, Angajala A, Lammel J, Arumanayagam AS, Webb P. Genome-wide binding patterns of thyroid hormone receptor beta. PLoS One. 2014;9:e81186.