



# Variable Baseline *Papio cynocephalus* Endogenous Retrovirus (PcEV) Expression Is Upregulated in Acutely SIV-Infected Macaques and Correlated to STAT1 Expression in the Spleen

## OPEN ACCESS

**Emmanuel Atangana Maze**<sup>1,2</sup>, **Claire Ham**<sup>2</sup>, **Jack Kelly**<sup>1</sup>, **Lindsay Ussher**<sup>1</sup>, **Neil Almond**<sup>2</sup>, **Greg J. Towers**<sup>3</sup>, **Neil Berry**<sup>2\*†</sup> and **Robert Belshaw**<sup>1\*†</sup>

<sup>1</sup> School of Biomedical Sciences, Faculty of Medicine and Dentistry, University of Plymouth, Plymouth, United Kingdom,

<sup>2</sup> Division of Infectious Disease Diagnostics, National Institute of Standards and Control (NIBSC), Potters Bar, United Kingdom, <sup>3</sup> Division of Infection and Immunity, University College London, London, United Kingdom

### Edited by:

Gkikas Magiorkinis,  
National and Kapodistrian University  
of Athens, Greece

### Reviewed by:

Mani Larjani,  
Memorial University of Newfoundland,  
Canada  
Claudia Matteucci,  
University of Rome Tor Vergata, Italy

### \*Correspondence:

Neil Berry  
neil.berry@nibsc.org  
Robert Belshaw  
robert.belshaw@plymouth.ac.uk

† Joint senior authors

### Specialty section:

This article was submitted to  
Viral Immunology,  
a section of the journal  
Frontiers in Immunology

**Received:** 28 January 2019

**Accepted:** 08 April 2019

**Published:** 15 May 2019

### Citation:

Maze EA, Ham C, Kelly J, Ussher L,  
Almond N, Towers GJ, Berry N and  
Belshaw R (2019) Variable Baseline  
*Papio cynocephalus* Endogenous  
Retrovirus (PcEV) Expression Is  
Upregulated in Acutely SIV-Infected  
Macaques and Correlated to STAT1  
Expression in the Spleen.  
Front. Immunol. 10:901.  
doi: 10.3389/fimmu.2019.00901

Retroviral replication leaves a DNA copy in the host cell chromosome, which over millions of years of infection of germline cells has led to 5% of the human genome sequence being comprised of endogenous retroviruses (ERVs), distributed throughout an estimated 100,000 loci. Over time these loci have accrued mutations such as premature stop codons that prevent continued replication. However, many loci remain both transcriptionally and translationally active and ERVs have been implicated in interacting with the host immune system. Using archived plasma and tissue samples from past macaque studies, experimentally infected with simian immunodeficiency virus (SIV), the expression of one macaque ERV in response to acute viral infection was explored together with a measure of the innate immune response. Specifically, RNA levels were determined for (a) *Papio cynocephalus* Endogenous Retrovirus (PcEV), an ERV (b) STAT1, a key gene in the interferon signaling pathway, and (c) SIV, an exogenous pathogen. Bioinformatic analysis of DNA sequences of the PcEV loci within the macaque reference genome revealed the presence of open reading frames (ORFs) consistent with potential protein expression but not ERV replication. Quantitative RT-PCR analysis of DNase-treated RNA extracts from plasma derived from acute SIV-infection detected PcEV RNA at low levels in 7 of 22 macaques. PcEV RNA levels were significantly elevated in PBMC and spleen samples recovered during acute SIV infection, but not in the thymus and lymph nodes. A strong positive correlation was identified between PcEV and STAT1 RNA levels in spleen samples recovered from SIV-positive macaques. One possibility is that SIV infection induces PcEV expression in infected lymphoid tissue that contributes to induction of an antiviral response.

**Keywords:** endogenous retrovirus, STAT1, macaque, PcEV, SIV, RNA, innate immunity

## INTRODUCTION

Endogenous Retroviruses (ERV) are descendants of ancient retroviral infections that have become established in the germline and proliferated to now represent ~5% of the human genome and other mammals, rising to ~8% if the older group of Mammalian Apparent LTR-Retrotransposons (MaLRs) are included. Individual proviruses, termed loci, accumulate mutations over time during their Mendelian transmission that eventually render them replication-defective. In the human reference genome, no locus has retained full-length Open Reading Frames (ORFs) for all genes which precludes the possibility of yielding infectious, cell-free progeny virions (1, 2). However, ERVs and other retroelements have been shown to play a role in epigenetic gene regulation (3–7), with many ERV loci having been co-opted. Examples include HERV-E Long Terminal Repeat (LTR)-driven tissue-specific expression of a human salivary amylase gene (8), ERV-derived syncytins that play a role in placentation (9), and promoter-containing LTRs of ERVs that influence epigenetic gene control and pluripotency (10).

The study of ERVs in non-human primates (NHP) has potential to broaden the wealth of data generated from mouse studies. Smaller mammals tend to have more ERV loci integrating into their genome than larger mammals (11) and, consistent with this general pattern, replication-competent ERV loci have been found in the mouse (12) but not in humans (1, 2, 12–14). In terms of the number of recently integrated ERV loci, the macaque genome is more similar to the human than the mouse (11). Macaques thus represent an opportunity to derive additional relevant data that could augment more conventional mouse studies. To help bridge this gap we have utilized archived materials derived from previous simian immunodeficiency virus (SIV) studies.

Recently, it has been suggested ERVs and other retroelements play a role in innate immune signaling by acting as Pathogen Associated Molecular Patterns (PAMPs) (15–21). Indeed, it has been proposed that retroelement PAMPs, such as cytosolic DNA resulting from reverse transcription, set an activation threshold for triggering the innate immune response (22). These authors speculate that the level at which activation occurs is an evolutionary trade-off between avoidance of tolerance of exogenous pathogens on one hand and constant triggering by endogenous elements on the other. The latter appears to occur in the inherited autoimmune disease Aicardi-Goutieres syndrome where deficiency of TREX1 DNA endonuclease may lead to a build-up of retroelement PAMPs, which triggers inappropriate innate signaling (23). If ERVs were involved in the early immune response, it might be expected that a correlation would exist between ERV expression and measures of innate immune activation in archived macaque samples taken during acute infection.

Our study focuses on *Papio cynocephalus* Endogenous Retrovirus (PcEV), a recently integrated ERV first described in baboons (24). PcEV is one of three ERV lineages copying within the macaque genome during the last 5 million years (25). Since a range of components in the ERV replication cycle can potentially be detected by the innate immune system (26), we argue that the

likelihood of an ERV locus being biologically relevant increases if its sequence is free of inactivating mutations such as premature stop codons. Induction of type-1 interferons (IFN-1), via the JAK/STAT pathway, represents a key part of the innate response to viral infections (27, 28). Interferon-stimulated genes (ISGs) act to limit viral replication and play a role in the control of the viral burden (29). However, the immunomodulatory role of ISG induction in acute-phase and chronic SIV infection is complex and relatively poorly understood, contributing to both dampening and inflammatory immune responses (30, 31). IFN-1 signaling results in the transcription of hundreds of ISGs, an important member being STAT1 (32, 33). Basal IFN-1 levels set the abundance of STAT1, which influences the susceptibility toward infection (34). STAT1 therefore provides a measure of IFN-1 activation induced during acute SIV infection in macaques (35), hence its expression was explored in this study.

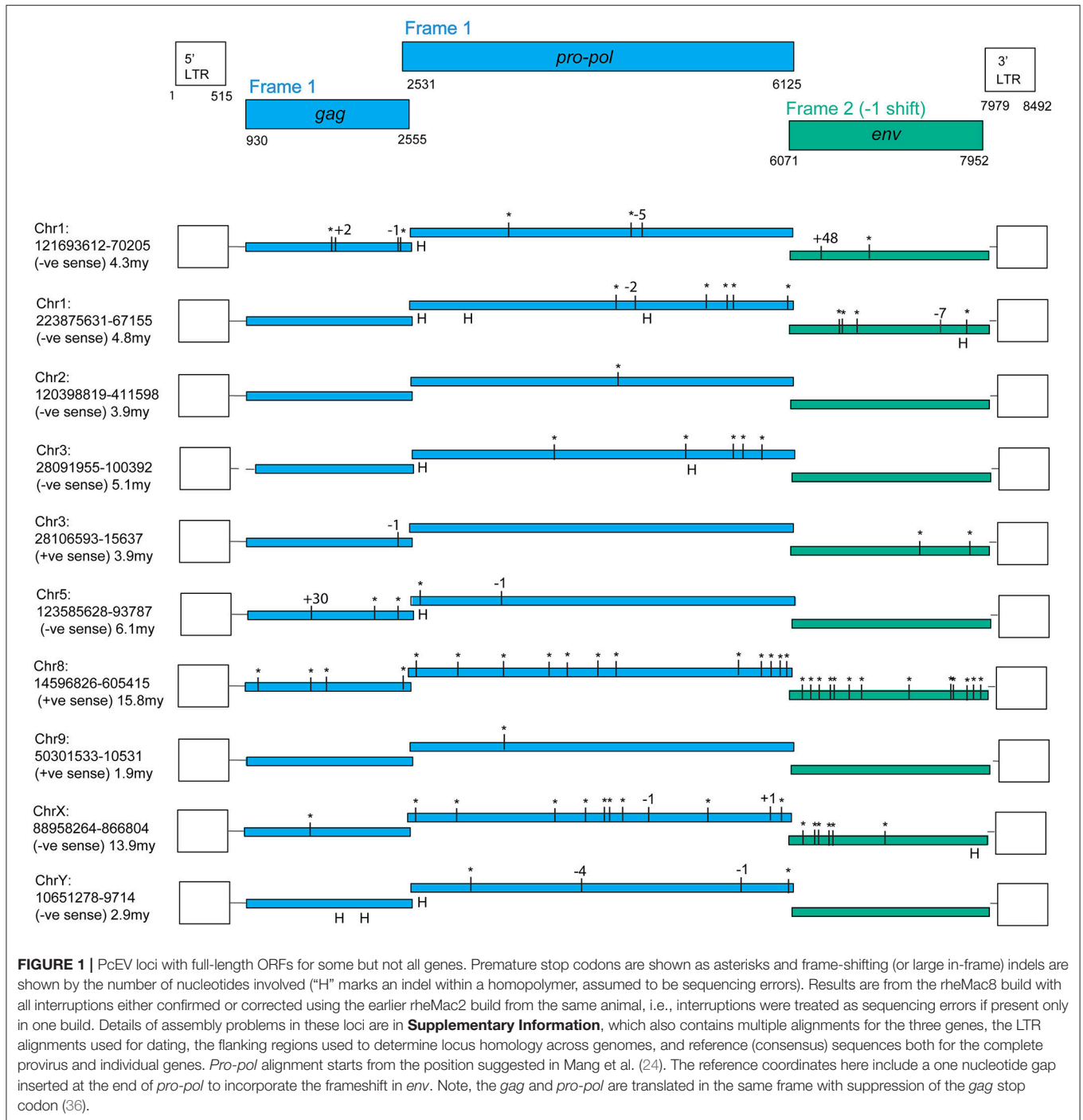
Archived macaque plasma and tissue samples were analyzed to determine PcEV activity during acute SIV infection when *de novo* responses are likely to be highest, particularly lymphoid tissues targeted during primary infection (35). This activity was also compared with STAT1 RNA levels in the same RNA preparations. Our study represents the first analysis of PcEV loci in the macaque reference genome sequences and demonstrates the presence of multiple potentially protein-coding, but not replicating, PcEV loci. In conjunction with identification of low levels of PcEV RNA in plasma of a proportion of acutely SIV-infected macaques, PcEV was actively transcribed in tissues with a level of cell-associated gene expression that is upregulated in response to acute SIV infection. In the spleen, this appeared to be directly correlated to STAT1 RNA expression. The potential significance of these findings linking ERV and innate immunity are discussed.

## RESULTS

### Bioinformatic Analysis Suggests PcEV Protein Production but Not Replication

An online search of the most recent rhesus macaque (*Macaca mulatta*, RM) genome assembly (rheMac8) using the constructed PcEV reference sequence revealed 72 PcEV loci (plus another 76 matches in unassembled regions), most of which are represented only by fragments. **Figure 1** illustrates the only 10 loci that do not have large regions (more than a few hundred nucleotides) missing from the reference genome. Among these loci are examples of full-length ORFs for all genes but no single locus has full-length ORFs in all genes.

The fragmentation observed is largely the result of genome assembly errors. Four of these 10 most intact loci contain scaffold gaps, where sequenced regions either side have not been joined. One example is locus chr9:50301533-10531, which in the rheMac8 assembly has its *pro-pol* reading frame interrupted by a single premature stop codon and by a long insertion which consists of a tandem duplication at a region marked as a scaffold gap. However, an examination of this locus sequence from an earlier assembly of the RM genome (rheMac2), which was from the same individual animal, and from the homologous



locus in the cynomolgus macaque (*Macaca fascicularis*, CM) genome (macFas5), a species that diverged from the RM ~1–2 million years ago (mya) (37, 38), demonstrates that neither the premature stop codon nor the insertion belong to this locus. These three *pro-pol* sequences are not homologous along their entire length and have a clearly visible transition, a “breakpoint,” near to position 3,937 in the reference genome: the ~40% of the rheMac8 *pro-pol* sequence before this breakpoint is no more

similar to the other two sequences than it is to a range of other PcEV loci, while downstream the sequences of all three loci are very similar.

This can also be demonstrated phylogenetically (**Figure 2**), where a tree of *pro-pol* sequences before the breakpoint shows the rheMac8 sequence is not recovered in the same clade as the rheMac2 and macFas5 sequences, while a tree of the *pro-pol* sequences after this break-point shows all three sequences

recovered together in a well-supported clade (as seen in the other loci). This represents a clear assembly error. Indeed, the locus in rheMac2 and in macFas5 are in the antisense direction while the locus in rheMac8 is in the sense direction. Consistent with widespread assembly problems, only approximately half of the 'intact' loci in rheMac8 are intact in macFas5, and *vice versa*. There is also much more fragmentation in PcEV loci than in the human HERV-K(HML2) loci (39), which belong to an older ERV lineage (40). The human ERVs were only sequenced after being cloned within bacterial artificial chromosomes (41), which avoided the problem with the macaque genomes of trying to assemble simultaneously multiple very similar ERV loci from short next generation sequencing reads.

Evidence was identified of insertionally polymorphic PcEV loci, i.e., loci present in some individuals but not others (42). The intact loci recovered are mostly several millions of years old (Figure 1), the most recently integrated locus being chr9:50301533-10531. The LTRs of this locus differ by two substitutions which, given an estimated rate of nucleotide substitution of  $\sim 1 \times 10^{-9}$ /nucleotide/year (25), indicates an age of  $\sim 1.9$  million years. All these loci are also found in the CM reference genome and are therefore expected to be fixed in the RM population. However, two loci with large scaffold gaps were estimated to have integrated within at least the last 700,000 years; both were insertionally polymorphic. Locus chr1:55452680-60247 is heterozygous for presence/absence (=provirus/pre-integration site) in the reference genome and locus chr5:114918771-21894 is represented solely by the pre-integration site in the rheMac3 genome sequence, which is from a different individual (from China rather than India; further details of these loci are included in **Supplementary Information**).

We infer from these observations that PcEV transcription arising from some of these 10 loci may result in PcEV protein production but is unlikely to result in PcEV replication.

## PcEV Loci Are Likely to be Transcribed and Lack STAT1 Binding Sites

The LTRs of the above loci, excluding the two very old ones, all have an undisrupted TATA binding protein (TBP) binding site (=TATA box) at position 382 (Figure 3; Table S1). This correlates to the position of the conserved TATA box found in PcEV in the baboon genome (24). Since the TATA box usually has at least minimal promoting activity (43), loci without disruption in the TBP binding site are likely to be transcribed. Another TBP binding site was predicted at position 51 but is unlikely to be real since it is poorly conserved among the loci and not found in other  $\gamma$ -retroviruses, such as HERV-W and MuLV (43, 44). The LTRs also contain many undisrupted transcription factor binding sites. The regions containing these sites correlated with ones already described in the baboon (24), so were named in accordance with that earlier study. Direct-repeat enhancers (DR1, DR2, DR3) contained sites for GR $\alpha$ , GATA1, and GATA2. Three CCAAT box-associated binding regions (CCAAT box 1, 2, and 3) contained sites for GR $\beta$ , CEBP $\beta$ , and NFY. A fourth CCAAT box region (box 4) contains these three sites plus the site for CEBP $\alpha$ . Such CCAAT boxes directly

upstream of a TATA box have been shown to efficiently promote transcription activity of the HERV-W LTR (43). Other retroviral LTRs similarly harbor binding sites for many transcription factors (45).

The screening of PcEV loci for putative transcription factor binding sites included key inflammation-related transcription factors: STAT1, NF $\kappa$ B (46), and STAT3 (47). Predicted binding sites for STAT1 and NF $\kappa$ B were not identified, so it is therefore unlikely that STAT1 or NF $\kappa$ B directly regulates PcEV expression through binding to the retroviral promoter. The only inflammation-related transcription factor binding site found was for STAT3 at position 149 (Figure 3).

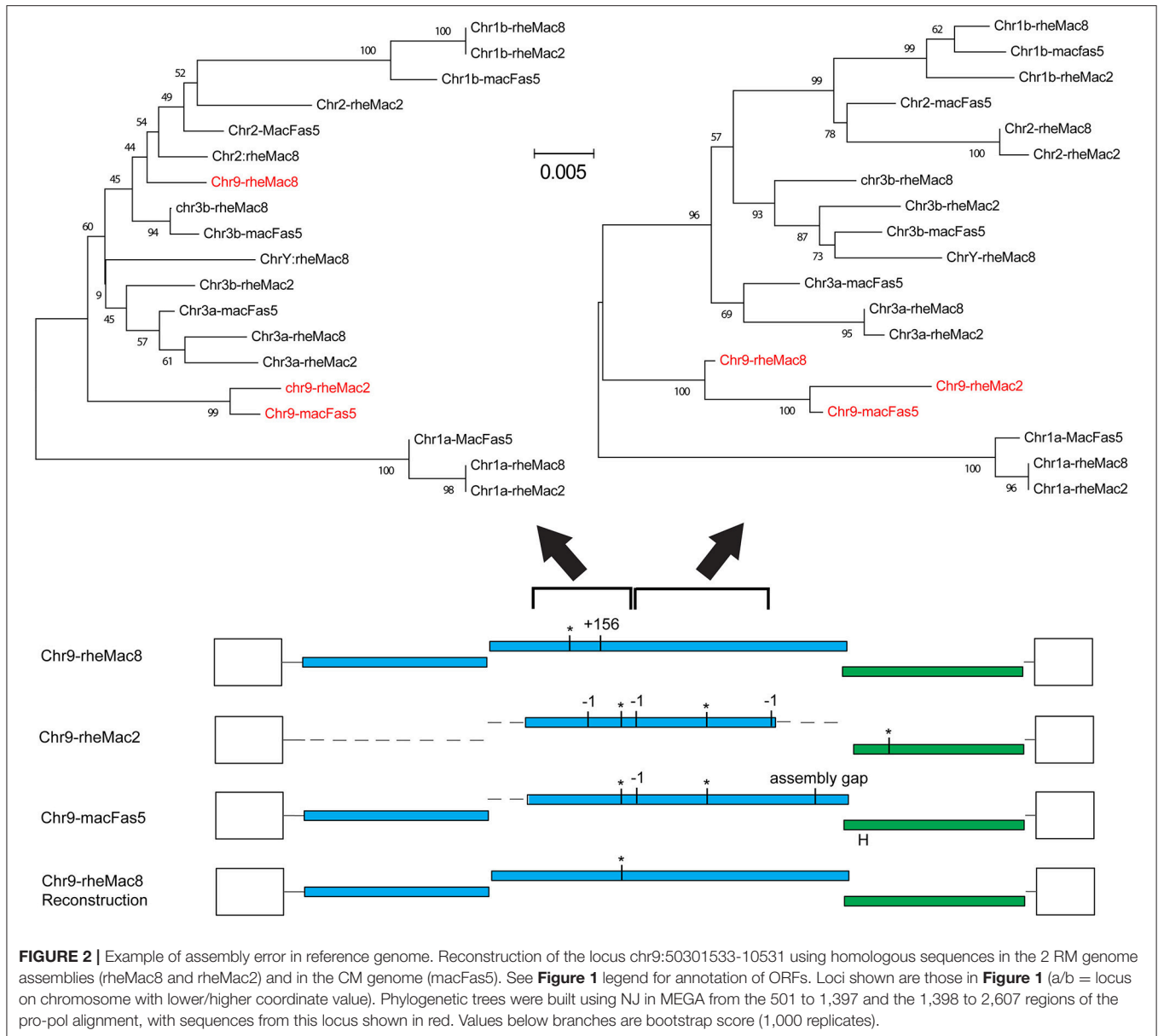
## Low Levels of Plasma PcEV RNA in One-Third of Acutely SIV-infected Macaques

Cell-free PcEV RNA levels were measured by RT-qPCR in plasmas from unchallenged control (naïve) macaques ( $n = 5$ ) or macaques at or around peak SIV viraemia (SIV+) challenged with one of four different SIV strains; SIVmac239 ( $n = 5$ ), SIVmac251 ( $n = 5$ ), SIVmacC8 ( $n = 6$ ), or SIVsmE660 ( $n = 6$ ). SIVmacC8 represents a minimally attenuated *nef*-disrupted variant of wild-type SIVmac251/32H (35), while all others are wild-type SIVs belonging to the SIVmac/SIVsm lineage. All viral challenges were performed in CM, except for SIVmac239 conducted in RM. Three naïve animals were CM, two were RM. As shown in Figure 4, 7 of 22 macaques (32%) exhibited low but detectable levels of PcEV in their plasma (39; 84; 94; 139; 70; 35; 165 PcEV copies/ml). These seven macaques include instances of infection with each of the four SIV strains. PcEV RNA was not detected in the plasma of naïve macaques, although the small sample size does not enable any inference about the relationship between PcEV detection in plasma and SIV status (Fisher's Exact Test comparing presence vs. absence of PcEV in plasma, from naïve and pooled SIV+ macaques,  $p=0.27$ ). Stringent protocols were employed for DNase treatment to remove contaminating genomic DNA, including paired +/- RT reaction steps in each experiment (Figure S1). Neither PcEV nor SIV were amplified in RT minus reactions (Table 1), providing additional confidence that all PcEV molecules originated from an RNA template. Hence, low but detectable PcEV RNA was detectable in around one-third of SIV+ plasmas during acute infection.

## Variable Baseline Tissue PcEV Transcriptional Activity Is Upregulated During Acute SIV Infection

Levels of detectable cellular PcEV transcripts in a range of tissues targeted by SIV were determined, isolating RNA from PBMC, spleen, thymus, peripheral and mesenteric lymph nodes (PLN/MLN) from naïve and acutely SIV+ macaques. Data generated for cell-associated PcEV transcriptional activity were compared for each preparation, normalized by co-amplification of the constitutively expressed cellular GAPDH gene with PcEV. Baseline PcEV RNA transcriptional activity was identified in the majority of tissues from all naïve macaques (Figure 5),





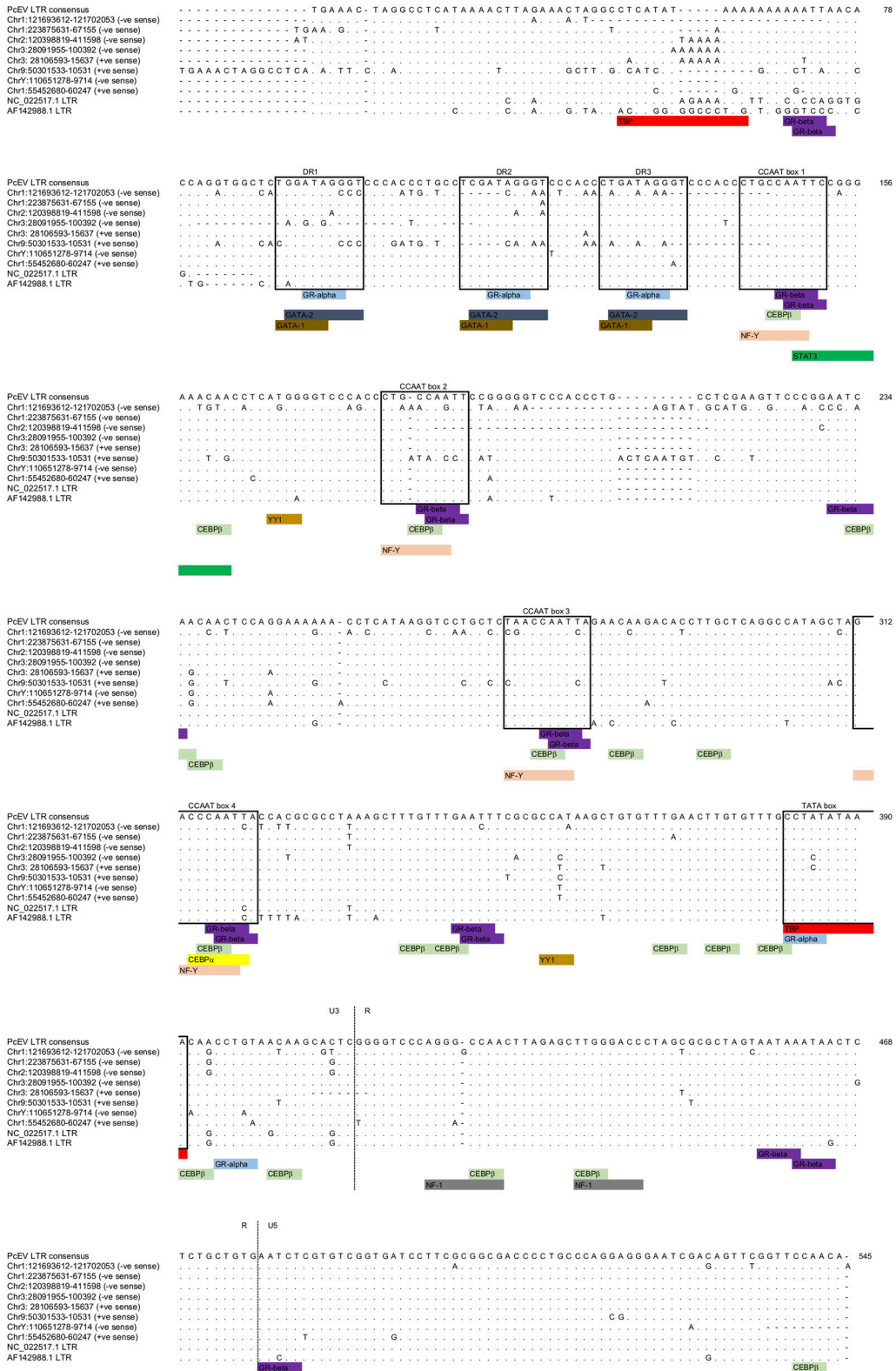
though many were at low levels and none exceeded 60 PcEV RNA copies/1,000 copies GAPDH. In these naïve tissues, basal PcEV transcriptional activity for MLN, thymus, PBMC, spleen and PLN was determined as 19, 20, 25, 33, and 52 PcEV RNA copies/ml, respectively. Differences between these tissue means were not statistically significant (Kruskal-Wallis test,  $p = 0.07$ ).

Mean PcEV levels were higher in tissues recovered from SIV+ macaques during acute infection (10–14 days post-infection) compared to naïve macaques (**Figure 5**). PcEV RNA was significantly increased in PBMC (Mann-Whitney test,  $p = 0.01$ ) and thymus ( $p = 0.03$ ), but increases were not significantly different in spleen ( $p = 0.23$ ), MLN ( $p = 0.38$ ), and PLN ( $p = 0.20$ ). Differences between tissue means were not statistically significant (Kruskal-Wallis test,  $p = 0.83$ ). Taken

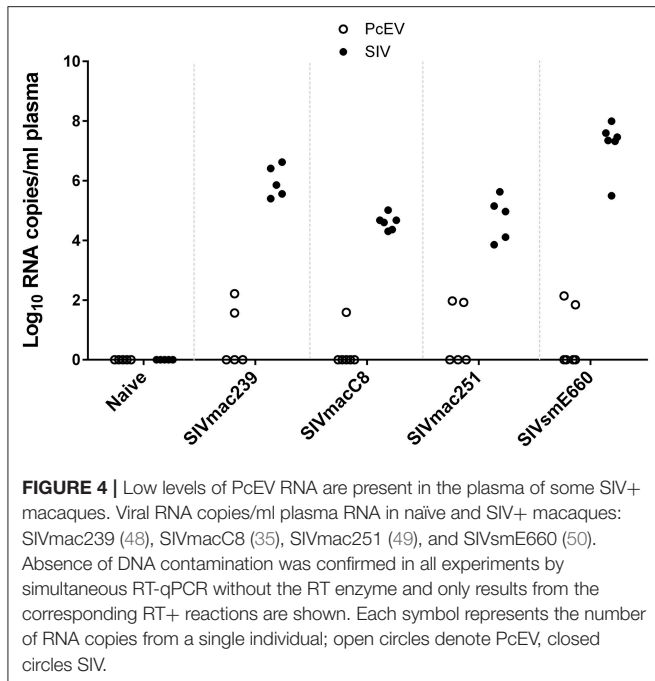
together, the data indicate that SIV infection *per se* upregulates the expression of PcEV RNA from a variable baseline level of transcriptional activity in a range of tissues targeted by SIV during acute infection.

### Cellular STAT1 RNA Levels Are Elevated During Acute SIV Infection

STAT1 represents an independent variable in these studies and its relationship with SIV status and tissue type was determined in the same manner as for PcEV (**Figure 6**). As with PcEV, a global STAT1 induction was observed during acute SIV infection compared to naïve macaques. Considering each tissue separately, STAT1 was significantly increased in PBMC (Mann-Whitney test,  $p < 0.001$ ), thymus ( $p < 0.001$ ), and MLN ( $p = 0.03$ ), but not spleen ( $p = 0.15$ ) or PLN ( $p = 0.60$ ). Hence, during acute SIV



**FIGURE 3** | PcEV LTR TATA box and transcription factor binding sites. Prediction of transcription factor binding sites and TATA box in the eight PcEV loci described in the macaque genome in this study. LTR sequences from PcEV loci in the baboon (accessions NC\_022517 and AF142988) are also shown.



infection cellular STAT1 levels are elevated, though at variable levels for individual tissues, as SIV infection stimulates a robust immune response.

### Cell-Associated PcEV Transcription Levels Are Correlated to STAT1 in SIV+ Spleens

Finally, expression levels of the two genes were compared in the same RNA preparations from each tissue sample. **Figure 7** shows tissue-specific patterns identifying the relationship between localized PcEV and STAT1 expression. Cellular PcEV and STAT1 RNA levels were strongly correlated in the spleen of SIV+ macaques (Spearman test,  $r = 0.90$ ,  $p = 0.005$ ) and showed a positive though statistically non-significant relationship in spleens derived from naïve macaques ( $r = 0.68$ ,  $p = 0.11$ ). Two SIV+ macaques (H19 and H20) with elevated splenic PcEV levels were PcEV+ in plasma, but the macaque with the highest level (G24) was PcEV RNA- in plasma. Four PBMC samples from macaques that were PcEV+ in plasma (including H19 and H20) exhibited a wide range of PcEV transcription levels.

STAT1 and PcEV levels were not significantly correlated in the other tissues but most relationships were positive. Pooling all tissue data from the 22 SIV+ macaques (a total of 30 samples) yielded a strong positive correlation (Spearman test,  $r = 0.47$ ;  $p = 0.0091$ ). Pooled PcEV and STAT1 levels in the 13 naïve macaques were also strongly correlated ( $r = 0.61$ ;  $p = 0.0004$ ), albeit at a lower magnitude of induction for both PcEV and STAT1. Such correlations in pooled data might be artifactual but it is interesting that no such correlation was found between STAT1 and cellular SIV levels in these total RNA preparations ( $r = 0.29$ ,  $p = 0.21$ ); there was also no correlation between STAT1 and plasma SIV RNA levels ( $r = 0.25$ ,  $p = 0.22$ ).

**TABLE 1** | PcEV and SIV RNA levels in plasma as determined by qPCR with (+) and without (-) reverse transcriptase (RT).

SIV strain	Individual	PcEV RNA copies/ml		SIV RNA copies/ml		
		RT+	RT-	RT+	RT-	
Naïve <sup>a</sup>	E79	No Ct	No Ct	No Ct	No Ct	
	E80	No Ct	No Ct	No Ct	No Ct	
	E81	No Ct	No Ct	No Ct	No Ct	
SIVmac239 <sup>a</sup>	E79	No Ct	No Ct	$2.60 \times 10^6$	No Ct	
	E80	No Ct	No Ct	$7.24 \times 10^5$	No Ct	
	E81	No Ct	No Ct	$4.23 \times 10^6$	No Ct	
SIVmacC8 <sup>a</sup>	E5	No Ct	No Ct	$1.04 \times 10^5$	No Ct	
	E6	No Ct	No Ct	$2.03 \times 10^4$	No Ct	
	G3	39	No Ct	$4.01 \times 10^4$	No Ct	
	G4	No Ct	No Ct	$4.75 \times 10^4$	No Ct	
	G5	No Ct	No Ct	$5.14 \times 10^3$	No Ct	
	G6	No Ct	No Ct	$2.31 \times 10^4$	No Ct	
SIVmac251 <sup>a</sup>	J41	No Ct	No Ct	$1.42 \times 10^5$	No Ct	
	J42	No Ct	No Ct	$9.35 \times 10^4$	No Ct	
	J44	84	No Ct	$4.29 \times 10^5$	No Ct	
	J17	No Ct	No Ct	$7.22 \times 10^3$	No Ct	
	J18	94	No Ct	$1.29 \times 10^4$	No Ct	
SIVsmE660 <sup>a</sup>	G25	No Ct	No Ct	$2.25 \times 10^7$	No Ct	
	G26	139	No Ct	$3.13 \times 10^5$	No Ct	
	G27	No Ct	No Ct	$2.12 \times 10^7$	No Ct	
	G28	No Ct	No Ct	$3.97 \times 10^7$	No Ct	
	G7	70	No Ct	$9.91 \times 10^7$	No Ct	
	G8	No Ct	No Ct	$2.89 \times 10^7$	No Ct	
	Naïve <sup>b</sup>	H17	No Ct	No Ct	No Ct	No Ct
		H18	No Ct	No Ct	No Ct	No Ct
SIVmac239 <sup>b</sup>	H19	37	No Ct	$3.64 \times 10^5$	No Ct	
	H20	165	No Ct	$2.52 \times 10^5$	No Ct	

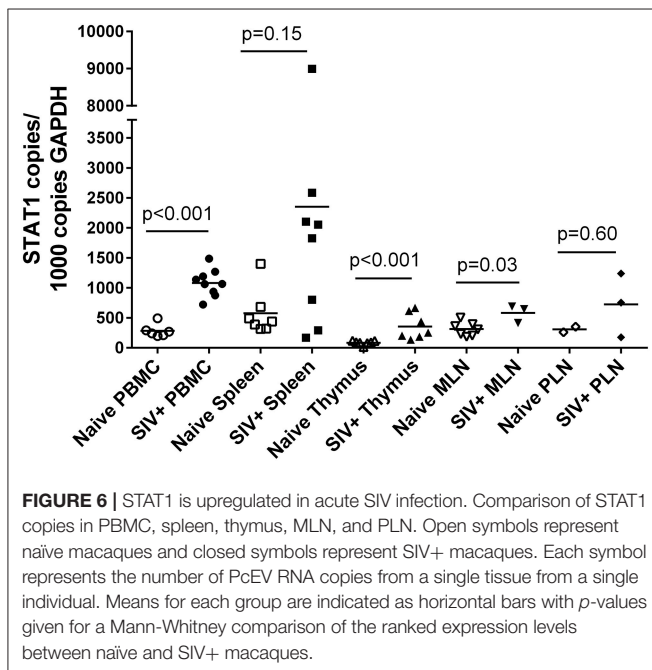
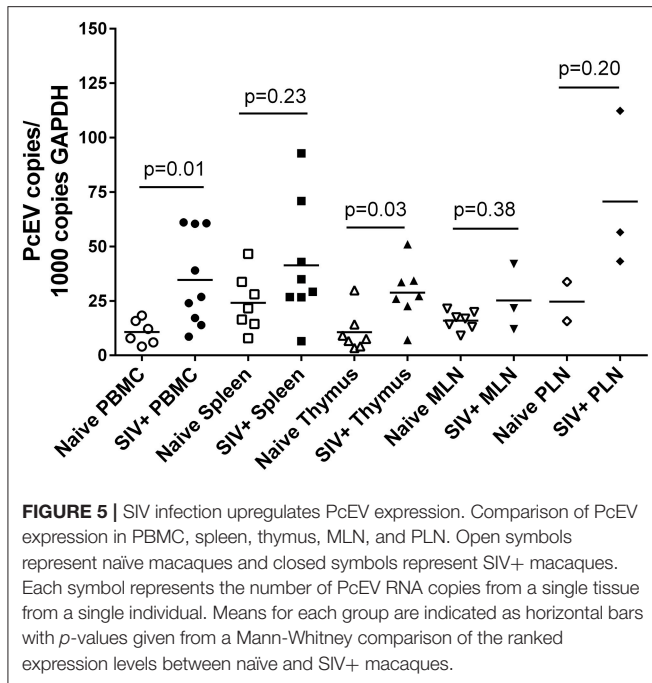
Species of macaque is indicated; CM, cynomolgus macaque; RM, rhesus macaque. PcEV RNA levels are expressed as simple integers; SIV RNA levels use a power of ten notation. Ct, cycle threshold.

<sup>a</sup>In CM.

<sup>b</sup>In RM.

## DISCUSSION

The transcriptional activity of an ERV in macaques has been investigated using a combination of bioinformatic prediction analyses and biological measures of *in vivo* activity, exploiting archived biological materials collected from previous macaque challenge studies. PcEV expression was upregulated in several key lymphoid tissues of acutely SIV-infected macaques, when *de novo* immune responses are likely to be highest. In the spleen PcEV expression was positively correlated to cellular STAT1 RNA levels. Analysis of baseline data indicated low, variable cell-associated PcEV RNA levels in the absence of a clear external stimulus, such as SIV. Introduction of SIV induced PcEV in all tissues measured, attaining statistical significance in the thymus and PBMC. While caution needs to be exercised in not over-interpreting our findings given a number of caveats (e.g., inherent variations in SIV strain dynamics, tissues targeted during acute



infection, macaque species), these data clearly suggest a global upregulation of PcEV transcriptional activity in response to acute SIV infection.

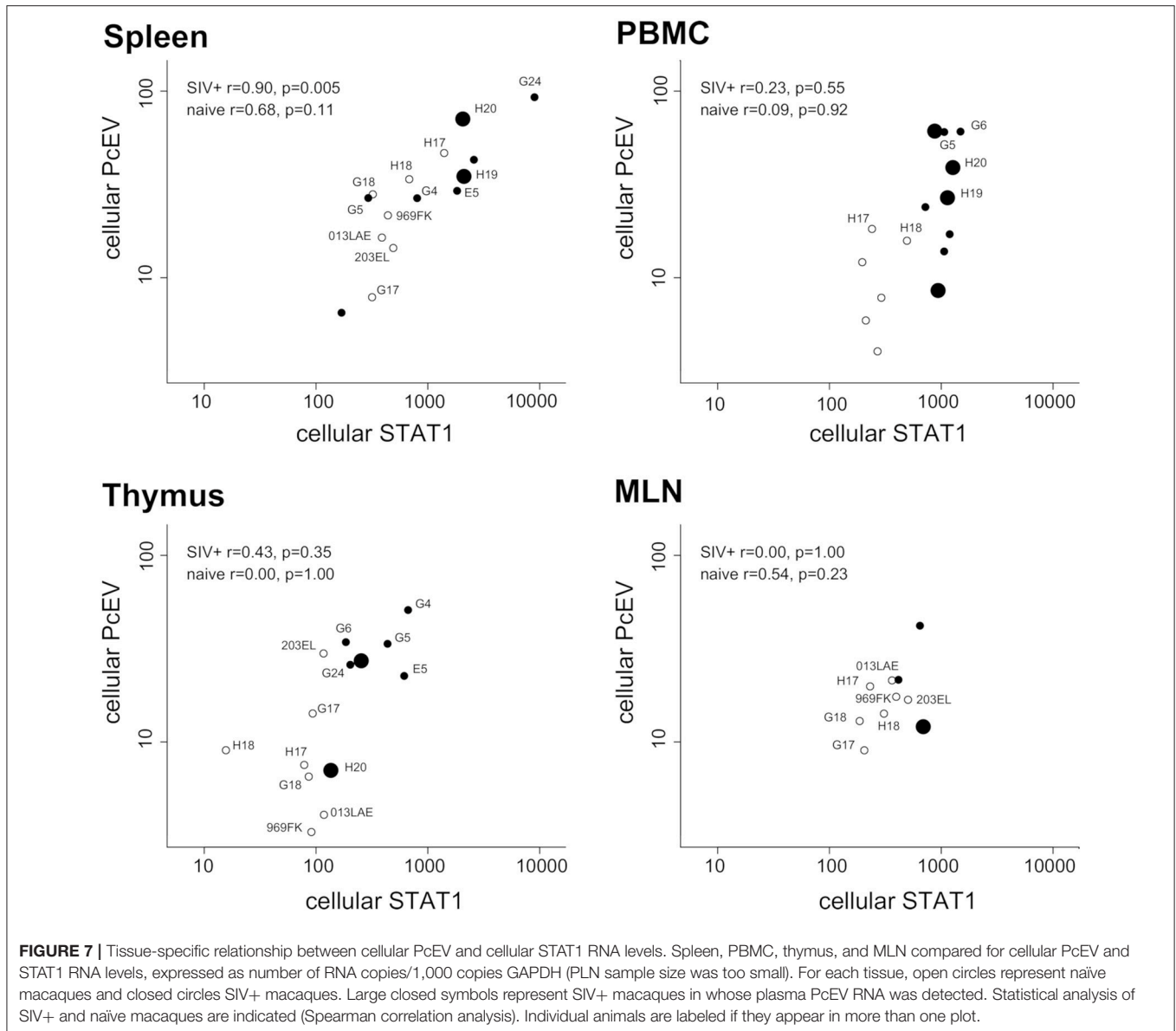
It is widely recognized that SIV establishes productive infection in multiple lymphoid tissues during the primary viraemic phase. STAT1, part of the IFN-1 cascade, represents one component of the dynamic response to SIV, triggering a range of innate immunity genes during acute infection (35), but this

response fails to clear the virus in the long term. This study confirmed induction of STAT1, part of the IFN-1 cascade in response to SIV, highlighting tissue-specific differences. In the spleen, PcEV and STAT1 RNA levels were positively correlated but this relationship was less clear in PBMC, thymus and MLN and warrants further scrutiny, especially where small sample sizes precluded a comprehensive analysis. With both PcEV and STAT1, exposure to SIV *per se* appears sufficient to trigger a response independent of the level of viral replication.

Although little is known about the macaque, in the mouse model it is well-established that microbial triggering of innate and adaptive sensors leads to increased ERV transcription (51). Treatment with both lipopolysaccharide (LPS), a Toll-like receptor 4 (TLR4) agonist, and polyinosinic-polycytidylic acid [poly(I:C)], a TLR3 agonist, significantly induced expression of two mouse ERV proviruses (52). Examples of human ERVs whose expression is enhanced in the presence of IFN or viral infection include HERV-K(HML2), reported to be increased in PBMC from HIV-1-infected individuals (13, 53, 54), and HERV-W was reported to be increased by influenza virus infection *in vitro* (55, 56). The HERV-K(HML2) LTR has two interferon-stimulated response elements (ISREs) and binding sites for inflammatory transcription factors such as NF $\kappa$ B (57, 58). PcEV lacks such motifs and the absence of clear STAT1 binding sites suggests that STAT1 proteins do not directly drive PcEV transcription. However, prediction of transcription binding sites is complex and lowering the dissimilarity margin of transcription binding site prediction from 5% to 15% allocates the STAT3 binding sites in PcEV also to NF $\kappa$ B and STAT1. STAT3 is activated at the protein level by IFN-1 (59) and binding motifs in PcEV LTR for CEPB $\beta$ , whose RNA expression is elevated by type-II IFN (60), were identified in the sequence analyses. Alternatively, PcEV might be upregulated by accessory proteins of exogenous retroviruses as HIV-1 Tat increases HERV-K(HML2) transcription (61, 62). Another possibility is inflammation-induced demethylation of LTRs, where the HIV-1 5'LTR undergoes demethylation following administration of TNF $\alpha$  *in vitro* and LPS to transgenic mice harboring latent HIV genomes in their lymphocytes (63, 64).

With regard to how PcEV might impact on innate immune signaling pathways, two main mechanisms can be envisaged (26). The first, via an exogenous route, implies binding of ERV particles or glycoproteins to innate receptors such as TLRs, as demonstrated *in vitro* where recombinant Env glycoprotein from HERV-W triggers secretion of pro-inflammatory cytokines via TLR4 in PBMC (65). The second, via an endogenous route, is that ERV transcription and reverse transcription may generate intermediate reaction products such as dsRNA, possibly by transcription and hybridization of ERV sense and antisense transcripts, and RNA:DNA hybrids generated during reverse transcription detected by TLRs and cytosolic RIG-I-like receptors, such as MDA5 (18–20). Only low PcEV RNA levels were detected in plasma from around one-third of SIV+ macaques, suggesting PcEV particles are unlikely to be produced in large quantities. It seems plausible, however, that PcEV might trigger induction mechanisms through released viral components generated in a transient nature, supported





by the finding of full-length ORFs for the *env* gene in several loci. In humans, ERV-encoded reverse transcriptase has been suggested to reverse transcribe viral RNA intracellularly (66, 67). In autoimmune TREX1-deficient mice, ERV DNA accumulates in the cytoplasm, associated with inflammatory phenotypes (23) which can be treated using RT inhibitors (68). Hence, the correlation between PcEV and STAT1 expression in this *in vivo* model requires further clarification.

One study limitation is that we restricted our analyses to data from only a single time point. This might not reveal any delayed correlations between PcEV, SIV, and STAT1 levels, especially given differences in cellular and anatomical localization of SIV (35). Cell sorting experiments to derive specific cell sub-sets would help address this issue and refine analyses. Differences in viral kinetics and tropism might also play a confounding role,

although all SIV strains studied, even the minimally attenuated SIVmacC8, exhibit robust early replication kinetics *in vivo* as determined by plasma viral RNA levels 10–14 days post-infection.

To assess the relevance of the macaque as a model system, similarities and differences between human and macaque ERVs require consideration. Humans have a single ERV lineage that has been replicating within the last few million years: HERV-K (HML2). Early studies reported HERV-K(HML2) RNA in plasma from HIV patients (53), but more recent studies failed to reproduce this (13, 14). Evidence for low level PcEV RNA was identified in the plasma of a proportion of SIV+ macaques. Detection of retroviral RNA in plasma presumes some form of protein protection, whether as fully or partially packaged particles, so some macaques in the population may harbor a

replication-competent PcEV locus. Although we did not find any potentially replication-competent PcEV in the reference macaque genome sequences, this may reflect a combination of genome assembly problems with the likelihood that such loci would be unfixed in the macaque population, i.e., present only in the genomes of some individual macaques (42). Evidence for at least two unfixed PcEV loci was found. This phenomenon occurs in HERV-K(HML2), where loci with full-length ORFs in all genes are present only in some non-reference genomes (69, 70). Replication-competent PcEV loci could also be generated by recombination: in the mouse model, recombination between replication-defective loci reconstituted replication-competent loci with resulting infectious virions (71). Although macaques might differ from humans in having as yet unrecovered replication-competent PcEV loci, levels of cell-associated HERV-K(HML2) transcription appear similar: ~30 copies/1,000 copies GAPDH in PBMC of non-acute HIV infection cases and ~15 RNA/1,000 copies GAPDH in PBMC from uninfected individuals (13).

Assuming SIV induces PcEV expression *in vitro*, knock-down of PcEV transcription and individual sensors followed by measurement of the IFN-1 response would determine whether PcEV is involved in the IFN-1 response. The other two recently active ERV lineages in the macaque, Chimpanzee Endogenous Retrovirus (CERV), and Simian Endogenous type D Retrovirus (SERV) (25), also require study since within both lineages single examples of loci with full-length ORFs in all genes were identified. Another important group of retroelements, Long Interspersed Nuclear Elements (LINEs), also require scrutiny because they, unlike ERVs, are known to be copying in the human population (72) and potentially replication-competent loci have been found in the macaque genome (73). LINEs have an entirely intra-cellular replication cycle that would help is distinguishing between endogenous and exogenous routes of innate triggering. Considering several ERVs and LINEs together would provide a wider perspective on whether retroelements are involved in fundamental immune response induction.

In conclusion, the current study demonstrates that PcEV retains transcriptional activity despite the apparent lack of replication. Expression appears to be upregulated by acute SIV infection and in some tissues a correlation exists between the expression levels of PcEV and STAT1, the latter being part of the IFN-1 cascade. The mechanistic processes at play warrant detailed investigation to determine whether ERVs are overlooked components of innate immune responses and activation signaling mechanisms.

## MATERIALS AND METHODS

### Extraction and Alignment of PcEV Locus Sequences

Bioinformatic analyses were performed using UCSC's Genome Browser website and its BLAT tool (74) on the most recent reference genome sequence of the RM, rheMac8 (Mmul\_8.0.1). This assembly derives from short-read Illumina whole genome sequencing (MacaM) (75)

with later closing of some gaps using long-read PacBio sequencing of the same individual animal (comment in GenBank accession GCA\_000772875.3).

Sequences of loci were compared with those from an earlier build of the RM genome, rheMac2 (76) (GenBank accession GCA\_000002255.2). This build is an earlier short-read whole genome sequencing project largely from the same (female) individual (ID 17573), with some sequencing of an unrelated male used in finishing and to provide a Y chromosome. Sequences were also compared to those of a CM, a closely related species sharing a common ancestor only 0.91 ( $\pm 0.11$ ) million years ago, which is more recent than the common ancestor of all rhesus macaques (38). The reference genome sequence of the CM, macFas5, appears to have been built independently of the RM genome using a range of short-read technologies (comment in GenBank accession GCA\_000364345.1). The other RM genome assembly available on the UCSC website, rheMac3, is very fragmentary.

PcEV was described initially from the baboon (24) (GenBank reference AF142988), erroneously referred to as BaEV in an earlier analysis of the macaque genome (25) (see **Supplementary Information**). This GenBank sequence was used to locate and download the sequences of several PcEV loci in the macaque using the UCSC Genome Browser. A multiple alignment was then made of the downloaded sequences using MEGA (77, 78) and a new consensus reference sequence built with full-length ORFs in all genes (sequences from different loci differed only by a few percent so alignments were unambiguous). This reference was used to re-search the macaque genome. Multiple alignments were made manually from downloaded sequences and individual loci examined visually for full-length ORFs and other motifs.

Previously published data on recently active macaque ERVs lineages (11, 25) are presented here in more detail (**Figures S2–S4**).

### Prediction of Transcription Factor Motifs

A consensus PcEV LTR sequence was built based on an alignment of LTR sequences from the loci displayed in **Figure 1**, excluding the two much older loci, plus NC\_022517 (36) and AF142988 (24). This sequence was screened with the ALGGEN-PROMO online tool (79, 80) for the presence of putative binding sites for a range of transcription factors linked to either inflammation or retroviral infection (43, 44). The dissimilarity between query and binding site matrices from the database was fixed at a threshold of 5%. The following transcription factors and binding sites were selected for NHPs only, with the accession number in the TRANSFAC public database given in square brackets: AP1 [T00029]; NFI/CTF [T00094]; C/EBP $\alpha$  [T00105]; NFY [T00150]; CREB [T00163]; cRel [T00168]; CTF [T00174]; GATA1 [T00306]; GATA2 [T00308]; GATA3 [T00311]; GR $\alpha$  [T00337]; NF1 [T00539]; NFAT1 [T00550]; C/EBP $\beta$  [T00581]; C/EBP $\Delta$  [T00583]; NFkB [T00590]; NFkB1 [T00593]; RelA [T00594]; POU2F1 [T00641]; Sp1 [T00759]; TBP [T00794]; YY1 [T00915]; STAT1 $\alpha$ , [T01492]; STAT3 [T01493]; STAT1 $\beta$  [T01573]; NFYA [T01804];

GR $\beta$  [T01920]; NFAT2 [T01945]; NFAT1 [T01948]; STAT1 [T04759]; GR [T05076].

## Dating the Integration of ERV Loci

Integrations were dated using the nucleotide divergence between the two LTRs of the proviruses. These LTRs form the flanks of the provirus (the complete integrated DNA form of a retrovirus) and are identical at the time of integration, subsequently accumulating substitutions at the host background rate. The following equation was used to calculate locus age (81): integration time = sequence distance between 5' and 3' LTR/(rate in 5' LTR + rate in 3' LTR).

Thus, estimated age = (a/b)/r\*2 where,

a = number of mismatches between the two LTRs,

b = length of the LTR,

r = estimated rate of nucleotide substitution.

Note, because substitutions will have occurred along the two branches leading to the 5' and the 3' LTRs it is possible to alternatively multiply the LTR length by two to reflect the effect of this. The rate of nucleotide substitution was taken as  $1.0 \times 10^{-9}$  substitutions/nucleotide/year based on analyses of primate ERV sequences (25). This value is about half the rate estimated for mammals generally (82, 83) but is close to estimates for the rate of neutral molecular evolution in Old World Monkeys and Apes (38).

For loci with no substitutions within their LTRs, the probability ( $P_0$ ) of observing this if the locus had integrated at time "age" can be found from the Poisson distribution with  $P_0 = e^{-\lambda}$  where  $\lambda$  is the average number of events in the time period. Thus, using the notation  $P_0 = e^{(-2*b*r*age)}$ , the probability of having no substitutions in the LTRs is just <0.5 at age 700,000 years.

## Study Population and RNA Preparation

Plasma and cells were derived from acutely SIV-infected Indian RM and CM challenged with different SIV isolates in historical studies of SIV pathogenesis and vaccination (35, 48, 50, 84). Total RNA was extracted from 140  $\mu$ L to 1 mL plasmas using QiaAmp (Qiagen Ltd) according to the manufacturer's instruction. Plasmas obtained from CM were extracted from a volume of 140  $\mu$ L, the four RM plasmas were extracted from 1 mL plasma. Each RNA preparation underwent a DNase step to remove contaminant DNA and all qPCR samples were run with an RT minus control, an essential step when working with ERVs. Healthy human plasma contains about 1,000 genome copies/mL, in the form of cell-free DNA (85), with many ERV copies in every cell. By contrast, exogenous retroviruses have only a single DNA copy in a small minority of PBMC (86), so the danger of DNA contamination is greatly reduced. For quantification of cell-associated transcripts, RNA was extracted from frozen cells, washed once in PBS and lysed with guanidine isothiocyanate (Sigma Ltd). Chloroform (200  $\mu$ L) was added followed by centrifugation at 13,000 rpm for 2 min, aqueous phase collected, 1 volume 100% ethanol added and samples loaded onto an RNEasy silica column (Qiagen Ltd). Columns were washed and RNA eluted according to the manufacturer's instruction to a final volume of 50  $\mu$ L and quantitated using

nanodrop. Each RNA preparation was diluted to 10 ng/ $\mu$ L, and 50 ng (5  $\mu$ L) added to each qPCR reaction.

## Primers and Probe Sequences

Primer sequences for qPCR are shown below in 5'-3' orientation. PcEV primers were based on an alignment of PcEV *pro-pol* sequences (Figure S5) designed to anneal to all loci. SIV primers are based on conserved regions in *gag* (49) except for SIVsmE660 based on conserved LTR sequences (50). PcEV, GAPDH, and STAT1 sequences were as follows:

PcEV: CCGTGTCTATCAAGCAATATCC (forward), GGCAGAAGAGGAGTGTCCAGG (reverse) and AACTCGGAGTGTTGCGAC (probe).

GAPDH: GGCTGAGAACGGGAAGCTC (forward), AGG GATCTCGCTCCTGGAA (reverse), and TCATCAATGGAA GCCCCATCACCA (probe).

STAT1: CAATACCTCGCACAGTGGTTAGAAAA (forward) and CGGATGGTGGCAAATGAAAC (reverse).

SIVsmE660: CTCCACGCTTGCTTGCTTAA (forward), AGGGTCCTAACAGACCAGGG (reverse), TCCCATATCTCT CCTAGYCGCCGC (probe).

Other SIVs: AGTGCCAACAGGCTCAGAAAA (forward), TGCGTGAATGCACCAGATG (reverse) and TTA AAAAGC CTTTATAATACTGTCTGCG (probe).

## Plasmids and *in vitro* Transcription

pBluescript II SK+ (pBS, 3.0 kb) was from Stratagene Ltd (1  $\mu$ g/ $\mu$ L). pBS-PcEV was derived separately by inserting a PcEV amplicon into the multiple cloning site of a pBS vector using *Apal* (nt 604) restriction sites (GENEWIZ, Inc). pBS-PcEV was linearized by *HindIII* restriction located 16nt downstream of the PcEV sequence. Reaction mixes included 1  $\mu$ g of pBS-PcEV plasmid (0.2  $\mu$ g/ $\mu$ L), 2 U/ $\mu$ L final concentration of enzyme and 15  $\mu$ L of water/enzyme buffer. Linearization was at 37°C for 1 h. *In vitro* transcription and purification were carried using MAXIscript kit (Ambion Ltd) following manufacturer's instructions.

## Standard Curves

PcEV transcripts were DNase-treated and 10-fold serial dilutions made. PcEV copy number was determined by Poisson end-point calculations, with Ct values obtained for each plasma sample compared to the standard curve to estimate number of PcEV copies per reaction (for 5  $\mu$ L total RNA input). The number of PcEV copies/mL for each plasma sample was determined taking into account extraction starting volume. SIV standards used were derived from a SIVmac251 plasma RNA reference series (49, 87). For intracellular RT-PCR assessments, standards for GAPDH and STAT1 were as previously described (35).

## DNase Treatment of RNA Preparations

Removal of DNA contaminant was conducted using DNA-free kit (Ambion Ltd). 5  $\mu$ L of 10X DNase buffer was added to 50  $\mu$ L RNA preparation. 2  $\mu$ L TURBO DNase was added and incubated for 1 h at 37°C. 10  $\mu$ L Inactivation Reagents were added and incubated for 5 min with intermittent flicking of tubes. Inactivation reagents were pelleted by centrifugation at

10,000 g for 2 min. Supernatant containing RNA was harvested, assayed or stored at  $-80^{\circ}\text{C}$  for further analysis. In the absence of DNAase treatment, similar copy numbers between RT+/RT- reactions were noted: H17(RT+ =  $1.18\text{E}+04$ ; RT- =  $1.27\text{E}+04$ ); H18(RT+ =  $7.06\text{E}+03$ ; RT- =  $9.12\text{E}+03$ ); H19(RT+ =  $2.39\text{E}+04$ ; RT- =  $2.37\text{E}+04$ ); H20(RT+ =  $6.96\text{E}+03$ ; RT- =  $8.99\text{E}+03$ ). Observed levels of DNA contamination are consistent with theoretical expectations: assuming macaques have similar levels of cell-free DNA in plasma, given at least 50 intact PcEV copies exist in each macaque cell, with an expectation of at least  $5 \times 10^4$  DNA copies in plasma.

## qPCR and qRT-PCR

Quantification of PcEV and SIV transcripts in plasma were determined with  $5 \mu\text{L}$  of DNase-treated RNA per qPCR reaction, assayed in triplicate using RNA Ultrasense One-Step Quantitative RT-PCR kit (Invitrogen Ltd). For PcEV, qPCR reactions consisted of  $25.9 \mu\text{L}$  nuclease-free water,  $1 \mu\text{L}$  forward and reverse primers ( $5 \mu\text{M}$  each, final concentration:  $100 \text{ nM}$ ),  $2.5 \mu\text{L}$  PcEV-specific Taqman probe ( $10 \mu\text{M}$ ),  $10 \mu\text{L}$  of 5X master mix,  $0.1 \mu\text{L}$  ROX reference dye,  $2.5 \mu\text{L}$  Superscript III RT and Platinum Taq polymerase enzyme mix. Cycling parameters were  $48^{\circ}\text{C}$  for 30 min RT step,  $95^{\circ}\text{C}$  for 5 min,  $95^{\circ}\text{C}$  for 15 s (40 cycles);  $61^{\circ}\text{C}$  extension 1 min. For SIV, mix contained  $25.9 \mu\text{L}$  nuclease-free water,  $1 \mu\text{L}$  each forward and reverse primer ( $5 \mu\text{M}$  each, final concentration:  $100 \text{ nM}$ ),  $2.5 \mu\text{L}$  SIV-specific probe ( $10 \mu\text{M}$ ). Cycling for  $52^{\circ}\text{C}$  60 min RT step,  $95^{\circ}\text{C}$  10 min,  $95^{\circ}\text{C}$  for 30 s (40 cycles) and  $61^{\circ}\text{C}$  extension for 90 s.

For quantification of cell-associated transcripts, each RNA preparation was diluted to  $10 \text{ ng}/\mu\text{L}$ , so that  $50 \text{ ng}$  ( $5 \mu\text{L}$ ) was added in each qPCR reaction. For PcEV and SIV, qPCR was conducted as stated above. For GAPDH, the mix contained  $25.9 \mu\text{L}$  nuclease-free water,  $1 \mu\text{L}$  each forward and reverse primer ( $5 \mu\text{M}$  each, final concentration:  $100 \text{ nM}$ ),  $2.5 \mu\text{L}$  GAPDH-specific probe ( $10 \mu\text{M}$ ). Cycling parameters were  $52^{\circ}\text{C}$  60 min RT step,  $95^{\circ}\text{C}$ , 10 min, 40 cycles of  $95^{\circ}\text{C}$ , 30 s (40 cycles) and  $61^{\circ}\text{C}$  extension for 90 s. For STAT1, the Power SYBR Green RNA-to-Ct 1-step kit was used (Applied Biosystems Ltd). Mix consisted of  $15.6 \mu\text{L}$  nuclease-free water,  $4 \mu\text{L}$  each forward and reverse primer ( $10 \mu\text{M}$  each, final concentration:  $400 \text{ nM}$ ),  $25 \mu\text{L}$  2X Master Mix (containing SYBR Green and polymerase) and  $0.4 \mu\text{L}$  RT-enzyme mix. Cycling parameters were  $48^{\circ}\text{C}$  30 min RT,  $95^{\circ}\text{C}$  for 10 min,  $95^{\circ}\text{C}$  for 15 s (40 cycles) and  $60^{\circ}\text{C}$  extension for 1 min. Melt curve was  $95^{\circ}\text{C}$  for 15 s,  $60^{\circ}\text{C}$  for 15 s and  $95^{\circ}\text{C}$  for 15 s. RNA preparations were confirmed to lack detectable DNA by performing a parallel quantitative PCR lacking reverse transcriptase for PcEV, SIV and GAPDH on all RNA extracts. In the RNA Ultrasense system, enzyme mix was replaced by  $0.2 \mu\text{L}$  Platinum Taq polymerase ( $10 \text{ U}/\mu\text{L}$ ) (Invitrogen Ltd).

## REFERENCES

1. Stoye JP. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol.* (2012) 10:395–406. doi: 10.1038/nrmicro2783
2. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet.* (2008) 42:709–32. doi: 10.1146/annurev.genet.42.110807.091501

## Statistical Analysis

GraphPad Prism version 6.01 was used for all statistical analyses. The Mann-Whitney test was performed for comparing non-paired samples in naïve vs. SIV+ groups; the Kruskal-Wallis test was used to compare expression levels between macaques infected with different strains of SIV. Since most datasets had distributions that were significantly different from normal, correlation coefficient  $r$  and  $p$ -values are presented from a non-parametric Spearman test. Fisher's Exact Test was used to assess whether SIV is associated with PcEV in the plasma of SIV+ animals.

## ETHICS STATEMENT

All biological samples used in this study were derived from previously conducted experiments, which received the appropriate ethical approval from the local NIBSC ethical committee. Animal procedures were performed in strict accordance with UK Home Office guidelines under a license granted by the Secretary of State for the Home Office which approved the work described. These previous experiments have been published (35, 48, 81–83, 85) along with full details of experimental design, procedures and ethical approval.

## AUTHOR CONTRIBUTIONS

EM, NB, and RB conceived and designed the experiments. EM and CH performed the experiments. RB, JK, and LU performed the bioinformatic analyses. EM, NA, GT, NB, and RB wrote the paper. EM and RB carried out the statistical analyses. All authors approved the final draft of the paper.

## FUNDING

The project was supported by NIBSC and a University of Plymouth Ph.D. studentship.

## ACKNOWLEDGMENTS

We are grateful to Giada Mattiuzzo (NIBSC, Division of Virology) for providing GAPDH and STAT1 standards. Amelia Ellisson and David Camp helped with bioinformatic analysis of PcEV loci.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00901/full#supplementary-material>

3. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature.* (2010) 463:237–40. doi: 10.1038/nature08674
4. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jönsson ME, Männe C, et al. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep.* (2015) 10:20–8. doi: 10.1016/j.celrep.2014.12.004



5. Brattås PL, Jönsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, et al. TRIM28 controls a gene regulatory network based on endogenous retroviruses in human neural progenitor cells. *Cell Rep.* (2017) 18:1–11. doi: 10.1016/j.celrep.2016.12.010
6. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K (HML-2). *J Virol.* (2005) 79:876. doi: 10.1128/JVI.79.2.876
7. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature.* (2014) 516:242–5. doi: 10.1038/nature13760
8. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* (1992) 6:1457–65. doi: 10.1101/gad.6.8.1457
9. Lavalie C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of “syncytins,” retroviral env genes exapted for a role in placentation. *Philos Trans R Soc B Biol Sci.* (2013) 368:20120507. doi: 10.1098/rstb.2012.0507
10. Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol.* (2014) 21:423–5. doi: 10.1038/nsmb.2799
11. Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, Gifford R. Larger mammalian body size leads to lower retroviral activity. *PLoS Pathog.* (2014) 10:e1004214. doi: 10.1371/journal.ppat.1004214
12. Stocking C, Kozak CA. Endogenous retroviruses: murine endogenous retroviruses. *Cell Mol Life Sci.* (2008) 65:3383–98. doi: 10.1007/s00018-008-8497-0
13. Bhardwaj N, Maldarelli F, Mellors J, Coffin JM. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses *in vivo* but not to increased virion production. *J Virol.* (2014) 88:11108–20. doi: 10.1128/JVI.01623-14
14. Karamitros T, Paraskevis D, Hatzakis A, Psychogiou M, Elefsiniotis I, Hurst T, et al. A contaminant-free assessment of Endogenous Retroviral RNA in human plasma. *Sci Rep.* (2016) 6:1–12. doi: 10.1038/srep33598
15. Bannert N, Hofmann H, Block A, Hohn O. HERVs new role in cancer: from accused perpetrators to cheerful protectors. *Front Microbiol.* (2018) 9:178. doi: 10.3389/fmicb.2018.00178
16. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science.* (2016) 351:1083–7. doi: 10.1126/science.aad5497
17. Hung T, Pratt GA, Sundaraman B, Townsend MJ, Chaivorapol C, Bhangale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science.* (2015) 350:455–9. doi: 10.1126/science.aac7442
18. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell.* (2015) 162:974–86. doi: 10.1016/j.cell.2015.07.011
19. Strick R, Strissel PL, Baylin SB, Chiappinelli KB. Unraveling the molecular pathways of DNA-methylation inhibitors: human endogenous retroviruses induce the innate immune response in tumors. *Oncotarget.* (2016) 5:e1122160. doi: 10.1080/2162402X.2015.1122160
20. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell.* (2015) 162:961–73. doi: 10.1016/j.cell.2015.07.056
21. Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol.* (2018) 9:2039. doi: 10.3389/fimmu.2018.02039
22. Volkman HE, Stetson DB. The enemy within: Endogenous retroelements and autoimmune disease. *Nat Immunol.* (2014) 15:415–22. doi: 10.1038/ni.2872
23. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell.* (2008) 134:587–98. doi: 10.1016/j.cell.2008.06.032
24. Mang R, Goudsmit J, van der Kuyl AC. Novel endogenous type C retrovirus in baboons: complete sequence, providing evidence for baboon endogenous virus gag-pol ancestry. *J Virol.* (1999) 73:7021–6.
25. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: Extinction and survival. *Retrovirology.* (2015) 12:8. doi: 10.1186/s12977-015-0136-x
26. Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. *J Gen Virol.* (2015) 96:1207–18. doi: 10.1099/jgv.0.000017
27. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nat Rev Immunol.* (2014) 14:36–49. doi: 10.1038/nri3581
28. Platanius LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol.* (2005) 5:375–86. doi: 10.1038/nri1604
29. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature.* (2011) 472:481–5. doi: 10.1038/nature09907
30. Sandler NG, Bosinger SE, Estes JD, Zhu RTR, Tharp GK, Boritz E, et al. Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature.* (2014) 511:601–5. doi: 10.1038/nature13554
31. Utay NS, Douek DC. Interferons and HIV Infection: the good, the bad, and the ugly. *Pathog Immun.* (2016) 1:107. doi: 10.20411/pai.v1i1.125
32. Der SD, Zhou A, Williams BRG, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci USA.* (1998) 95:15623–8. doi: 10.1073/pnas.95.26.15623
33. Lanford RE, Guerra B, Lee H, Chavez D, Brasky KM, Bigger CB. Genomic response to interferon- $\alpha$  in chimpanzees: implications of rapid downregulation for hepatitis C kinetics. *Hepatology.* (2006) 43:961–72. doi: 10.1002/hep.21167
34. Gough DJ, Messina NL, Clarke CJP, Johnstone RW, Levy DE. Constitutive type I interferon modulates homeostatic balance through tonic signaling. *Immunity.* (2012) 36:166–74. doi: 10.1016/j.immuni.2012.01.011
35. Ferguson D, Mattiuzzo G, Ham C, Stebbings R, Li B, Rose NJ, et al. Early biodistribution and persistence of a protective live attenuated SIV vaccine elicits localised innate responses in multiple lymphoid tissues. *PLoS ONE.* (2014) 9:e104390. doi: 10.1371/journal.pone.0104390
36. Kato S, Matsuo K, Nishimura N, Takahashi N, Takano T. The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retroviruses. *Japanese J Genet.* (1987) 62:127–37. doi: 10.1266/jgg.62.127
37. Jiang J, Yu J, Li J, Li P, Fan Z, Niu L, et al. Mitochondrial genome and nuclear markers provide new insight into the evolutionary history of macaques. *PLoS ONE.* (2016) 11:e0154665. doi: 10.1371/journal.pone.0154665
38. Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, et al. Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. *BMC Genomics.* (2008) 9:90. doi: 10.1186/1471-2164-9-90
39. 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) 8:90. doi: 10.1186/1742-4690-8-90
40. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA.* (2004) 101:4894–9. doi: 10.1073/pnas.0307800101
41. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* (2001) 409:860–921. doi: 10.1038/35057062
42. Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. *J Virol.* (2014) 88:9529–37. doi: 10.1128/JVI.00919-14
43. Lee WJ. Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat. *J Gen Virol.* (2003) 84:2229–35. doi: 10.1099/vir.0.19076-0
44. Speck N, Baltimore D. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol.* (1987) 7:1101–10. doi: 10.1128/MCB.7.3.1101
45. Luciw PA, Leung NJ. Mechanisms of retrovirus replication. In: Levy JA, editor. *The Retroviridae*. Boston, MA: The Viruses. Springer (1992) 159–298. doi: 10.1007/978-1-4615-3372-6\_5
46. Uleri E, Mei A, Mameli G, Poddighe L, Serra C, Dolei A. HIV Tat acts on endogenous retroviruses of the W family and this occurs via

- Toll-like receptor 4: inference for neuroAIDS. *AIDS*. (2014) 28:2659–70. doi: 10.1097/QAD.0000000000000477
47. Fan Y, Timani K, He J. STAT3 and its phosphorylation are involved in HIV-1 tat-induced transactivation of glial fibrillary acidic protein. *Curr HIV Res*. (2015) 13:55–63. doi: 10.2174/1570162X13666150121115804
  48. Berry N, Manoussaka M, Ham C, Ferguson D, Tudor H, Mattiuzzo G, et al. Role of occult and post-acute phase replication in protective immunity induced with a novel live attenuated SIV vaccine. *PLoS Pathog*. (2016) 12:e1006083. doi: 10.1371/journal.ppat.1006083
  49. Berry N, Stebbings R, Ferguson D, Ham C, Alden J, Brown S, et al. Resistance to superinfection by a vigorously replicating, uncloned stock of simian immunodeficiency virus (SIVmac251) stimulates replication of a live attenuated virus vaccine (SIVmacC8). *J Gen Virol*. (2008) 89:2240–51. doi: 10.1099/vir.0.2008/001693-0
  50. Berry N, Ham C, Mee ET, Rose NJ, Mattiuzzo G, Jenkins A, et al. Early potent protection against heterologous SIVsmE660 challenge following live attenuated SIV vaccination in mauritian cynomolgus macaques. *PLoS ONE*. (2011) 6:e23092. doi: 10.1371/journal.pone.0023092
  51. Kassiotis G, Stoye JP. Immune responses to endogenous retroelements: taking the bad with the good. *Nat Rev Immunol*. (2016) 16:207–19. doi: 10.1038/nri.2016.27
  52. Young GR, Mavrommatis B, Kassiotis G. Microarray analysis reveals global modulation of endogenous retroelement transcription by microbes. *Retrovirology*. (2014) 11:59. doi: 10.1186/1742-4690-11-59
  53. Contreras-Galindo R, Kaplan MH, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Ferlenghi I, Giusti F, et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J Virol*. (2012) 86:262–76. doi: 10.1128/JVI.00602-11
  54. Contreras-Galindo R, López P, Vélez R, Yamamura Y. HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) *in vitro*. *AIDS Res Hum Retroviruses*. (2007) 23:116–22. doi: 10.1089/aid.2006.0117
  55. Nellåker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology*. (2006) 3:44. doi: 10.1186/1742-4690-3-44
  56. Li F, Nellaker C, Sabunciyar S, Yolken RH, Jones-Brando L, Johansson A-S, et al. Transcriptional depression of the ERVWE1 locus following influenza A virus infection. *J Virol*. (2014) 88:4328–37. doi: 10.1128/JVI.03628-13
  57. Manghera M, Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology*. (2013) 10:16. doi: 10.1186/1742-4690-10-16
  58. Manghera M, Ferguson-Parry J, Lin R, Douville RN. NF- $\kappa$ B and IRF1 induce endogenous retrovirus K expression via interferon-stimulated response elements in its 5' long terminal repeat. *J Virol*. (2016) 90:9338–49. doi: 10.1128/JVI.01503-16
  59. Ho HH, Ivashkiv LB. Role of STAT3 in type I interferon responses: Negative regulation of STAT1-dependent inflammatory gene activation. *J Biol Chem*. (2006) 281:14111–8. doi: 10.1074/jbc.M511797200
  60. Roy SK, Wachira SJ, Weihua X, Hu J, Kalvakolanu DV. CCAAT/enhancer-binding protein- $\beta$  regulates interferon-induced transcription through a novel element. *J Biol Chem*. (2000) 275:12626–32. doi: 10.1074/jbc.275.17.12626
  61. Gonzalez-Hernandez MJ, Swanson MD, Contreras-Galindo R, Cookinham S, King SR, Noel RJ, et al. Expression of human endogenous retrovirus type K (HML-2) is activated by the tat protein of HIV-1. *J Virol*. (2012) 86:7790–805. doi: 10.1128/JVI.07215-11
  62. Gonzalez-Hernandez MJ, Cavalcoli JD, Sartor MA, Contreras-Galindo R, Meng F, Dai M, et al. Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 tat protein. *J Virol*. (2014) 88:8924–35. doi: 10.1128/JVI.00556-14
  63. Ishida T, Hamano A, Koiwa T, Watanabe T. 5' long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals. *Retrovirology*. (2006) 3:69. doi: 10.1186/1742-4690-3-69
  64. Tanaka J, Ishida T, Choi B-I, Yasuda J, Watanabe T, Iwakura Y. Latent HIV-1 reactivation in transgenic mice requires cell cycle-dependent demethylation of CREB/ATF sites in the LTR. *AIDS*. (2003) 17:167–75. doi: 10.1097/e01.aids.0000042971.95433.b0
  65. Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN. The envelope protein of a human endogenous retrovirus-w family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J Immunol*. (2006) 176:7636–44. doi: 10.4049/jimmunol.176.12.7636
  66. Dube D, Contreras-Galindo R, He S, King SR, Gonzalez-Hernandez MJ, Gitlin SD, et al. Genomic flexibility of human endogenous retrovirus type K. *J Virol*. (2014) 88:9673–82. doi: 10.1128/JVI.01147-14
  67. Contreras-Galindo RA, Dube D, Fujinaga K, Kaplan MH, Markovitz DM. Susceptibility of human endogenous retrovirus type-K to reverse transcriptase inhibitors. *J Virol*. (2017) 91:e01309-17. doi: 10.1128/JVI.01309-17
  68. Beck-Engeser GB, Eilat D, Wabl M. An autoimmune disease prevented by anti-retroviral drugs. *Retrovirology*. (2011) 8:91. doi: 10.1186/1742-4690-8-91
  69. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci USA*. (2016) 113:E2326–34. doi: 10.1073/pnas.1602336113
  70. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol*. (2001) 11:1531–5. doi: 10.1016/S0960-9822(01)00455-9
  71. Young GR, Eksmond U, Salcedo R, Alexopoulou L, Stoye JP, Kassiotis G. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature*. (2012) 491:774–8. doi: 10.1038/nature11599
  72. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 retrotransposition activity in human genomes. *Cell*. (2010) 141:1159–70. doi: 10.1016/j.cell.2010.05.021
  73. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, et al. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science*. (2007) 316:238–40. doi: 10.1126/science.1139462
  74. Kent WJ. BLAT — The BLAST-like alignment tool. *Genome Res*. (2002) 12:656–64. doi: 10.1101/gr.229202
  75. Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S, et al. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct*. (2014) 9:20. doi: 10.1186/1745-6150-9-20
  76. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*. (2007) 316:222–34. doi: 10.1126/science.1139247
  77. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. (2013) 30:2725–9. doi: 10.1093/molbev/mst197
  78. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. (2011) 28:2731–9. doi: 10.1093/molbev/msr121
  79. Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. (2002) 18:333–4. doi: 10.1093/bioinformatics/18.2.333
  80. Farré D, Roset R, Huerta M, Adsua JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. *Nucleic Acids Res*. (2003) 31:3651–3. doi: 10.1093/nar/gkg605
  81. Martins H, Villesen P. Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS ONE*. (2011) 6:e14745. doi: 10.1371/journal.pone.0014745
  82. Subramanian S, Kumar S. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res*. (2003) 13:838–44. doi: 10.1101/gr.1152803
  83. Yi S, Ellsworth DL, Li WH. Slow molecular clocks in old world monkeys, apes, and humans. *Mol Biol Evol*. (2002) 19:2191–8. doi: 10.1093/oxfordjournals.molbev.a004043
  84. Mattiuzzo G, Rose NJ, Almond N, Towers GJ, Berry N. Upregulation of TRIM5 $\alpha$  gene expression after liveattenuated simian immunodeficiency virus vaccination in Mauritian cynomolgus macaques, but TRIM5 $\alpha$  genotype has no impact on virus acquisition or vaccination outcome. *J Gen Virol*. (2013) 94:606–11. doi: 10.1099/vir.0.047795-0

85. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal Bioanal Chem.* (2014) 406:6499–512. doi: 10.1007/s00216-014-7835-3
86. Parkin NT. Measurement of HIV-1 viral load for drug resistance surveillance using dried blood spots: literature review and modeling of contribution of DNA and RNA. *Aids Rev.* (2014) 16:160–71.
87. Ham C, Srinivasan P, Thorstensson R, Verschoor E, Fagrouche Z, Sernicola L, et al. International multicenter study to assess a panel of reference materials for quantification of Simian immunodeficiency virus RNA in plasma. *J Clin Microbiol.* (2010) 48:2582–5. doi: 10.1128/JCM.00082-10

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Maze, Ham, Kelly, Ussher, Almond, Towers, Berry and Belshaw. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.