nature communications

Article

https://doi.org/10.1038/s41467-025-56772-3

Non-AUG HIV-1 uORF translation elicits specific T cell immune response and regulates viral transcript expression

Received: 11 August 2023

Accepted: 29 January 2025

Published online: 18 February 2025

Check for updates

Emmanuel Labaronne[1,2], Didier Décimo[3,11], Lisa Bertrand[4,11], Laura Guiguettaz[1], Thibault J. M. Sohier[1], David Cluet[1], Valérie Vivet-Boudou[5], Ana Luiza Chaves Valadão[6], Clara Dahoui[3], Pauline François[4], Isabelle Hatin[4], Olivier Lambotte[7,8], Assia Samri[9], Brigitte Autran[9], Lucie Etienne[3], Caroline Goujon[6], Jean-Christophe Paillart[5], Olivier Namy[4], Bertha Cecilia Ramirez[4,10], Théophile Ohlmann[3], Arnaud Moris[4] & Emiliano P. Ricci[1]

Human immunodeficiency virus type-1 (HIV-1) is a complex retrovirus that relies on alternative splicing, translational, and post-translational mechanisms to produce over 15 functional proteins from its single ~10 kb transcriptional unit. Using ribosome profiling, nascent protein labeling, RNA sequencing, and whole-proteomics of infected CD4 + T lymphocytes, we characterized the transcriptional, translational, and post-translational landscape during infection. While viral infection exerts a significant impact on host transcript abundance, global translation rates are only modestly affected. Proteomics data reveal extensive transcriptional and post-translational regulation, with many genes showing opposing trends between transcript/ribosome profiling and protein abundance. These findings highlight a complex regulatory network orchestrating gene expression at multiple levels. Viral ribosome profiling further uncovered extensive non-AUG translation of small peptides from upstream open reading frames (uORFs) within the 5′ long terminal repeat, which elicit specific T cell responses in people living with HIV. Conservation of uORF translation among retroviruses, along with TAR sequences, shapes DDX3 dependency for efficient translation of the main viral open reading frames.

HIV-1 is a single-stranded, positive-sense, diploid and enveloped RNA virus that belongs to the Retroviridae family. Upon infection, the genomic RNA is reverse-transcribed into a double-stranded DNA molecule that is then integrated into the host cell genome as a provirus or in some instances remains unintegrated in a circular form. The provirus DNA (integrated or circular) is transcribed by host RNA polymerase II (Pol II) into a single capped and polyadenylated RNA. Like other complex retroviruses, this single RNA species undergoes extensive alternative splicing to produce a multitude of matured transcripts coding for regulatory proteins in addition to the canonical Gag, Pol and Env proteins[1–3]. Alternative splicing of the viral transcript and expression of the different viral proteins is temporally regulated during the course of infection. Fully spliced transcripts coding for regulatory proteins are preferentially exported to the cytoplasm during early stages of the replication cycle. Partially spliced mRNAs and the unspliced genomic RNA (the latter also coding for Gag and Gag-Pol) are preferentially exported at late stages of the replication cycle through a process that requires the viral encoded protein Rev[4].

Translation of viral transcripts is also tightly regulated during the replication cycle through multiple mechanisms of translation

A full list of affiliations appears at the end of the paper. e-mail: emiliano.ricci@ens-lyon.org

initiation for both the unspliced genomic RNA and several spliced variants[5–7]. Indeed, the 5′ untranslated region (UTR) of viral transcripts is relatively long (>300 nucleotides) and contains several extensive RNA structures (such as the Tat-responsive target (TAR), Poly(A), primer-binding site (PBS) and the packaging signal stem-loops) that are essential at different steps of the replication cycle but may represent an obstacle for 43S ribosomes to bind viral mRNAs and reach the translation start codon[8]. As such, internal ribosome entry sites (IRESes) have been described in the 5′UTR of unspliced and some spliced viral transcripts as well as in the coding region of Gag for unspliced RNAs[5,9–16]. Furthermore, the viral protease encoded from the full-length RNA is capable of mediating proteolytic cleavage of the translation initiation factors eIF4G, eIF4G2 and PABP, which inhibits cap-dependent translation and 43S ribosomal scanning[15,17–21]. Translation initiation on viral RNAs can also occur through a cap-dependent process that is facilitated by many host proteins including RNA helicases and RNA binding proteins that facilitate recruitment and scanning of 40S on viral transcripts[5,22–30]. Translation of the full-length RNA is also regulated at the elongation step, through a frameshifting signal which allows the expression of the structural Gag polyprotein and the fusion Gag-Pol polyprotein that contains viral proteins with enzymatic activity such as the protease, reverse-transcriptase and integrase[31].

The regulation of viral protein translation has a strong influence on the initiation of immune responses to the virus, specifically on T cell mediated immunity. T cells recognize, on infected cells, virus-derived peptides presented by major histocompatibility complex (MHC) molecules. In particular, CD8 + T cells recognize viral peptide, presented by MHC class-I molecules, that are derived from the processing, by proteasomes, of the native viral proteins and/or, misfolded or truncated viral polypeptides[32,33]. As a consequence, there is a direct link between translation, protein degradation and the loading of MHC-I molecules[32,34,35], in particular in HIV-infected cells[36,37]. Interestingly, CD4 + T cells can also recognize HIV peptides, presented by MHC class-II molecules, that are derived from newly synthesized viral proteins[38].

Translation of HIV-1 and closely related lentiviruses has been mainly studied using in vitro translation extracts or reporter constructs that have been instrumental to uncover the many mechanisms governing HIV-1 expression[8,10–12,15,19,20,39]. However, a global view of the expression and translation of viral transcripts during a productive replication cycle is still missing. Here, we have performed whole proteome mass-spectrometry analysis, ribosome profiling and RNA sequencing from cytoplasmic extracts obtained from infected cells and cells with a latent HIV-1 infection in which the replication cycle is induced in a coordinated manner within the cell population (with most cells responding simultaneously). Our results highlight the temporal regulation of expression and translation of viral and cellular transcripts and indicate a modest impact of infection on overall cellular translation rates. Ribosome profiling data reveals extensive non-AUG translation events from upstream open reading frames (uORFs) located in the 5′UTR of spliced and unspliced viral mRNAs. We further demonstrate that uORFs negatively modulate translation of the main viral ORFs, and that this effect can be alleviated by the host DEAD-box protein DDX3. uORF translation from viral transcripts also occurs in the closely related lentivirus HIV-2 and in the more phylogenetically distant retrovirus HTLV-1 (Human T Lymphotropic Virus 1), suggesting that it may be a conserved feature among retroviruses. Finally, using IFNγ ELISPOT assays in PBMCs from people living with HIV, we can detect specific T cell responses directed against uORF-derived peptides thus indicating that uORFs encode MHC-ligands with potential relevance in mediating an immune response against infected cells in vivo.

Results

HIV-1 infection induces changes in abundance and translation of specific host mRNAs

Infection with HIV-1 has been previously proposed to down-regulate host cap-dependent translation through the cleavage of translation initiation factors and the arrest of infected cells in G2-M phase[17–19,21,40]. However, this has mainly been tested under over-expression of the viral protease or transfection of plasmids coding for the infectious provirus. A multi-omics global assessment of cellular and viral transcript abundance, translation rates and protein abundance changes under productive infection is still missing. To monitor the impact of HIV-1 infection on gene expression, we infected SupT1 CD4 + T cells with HIV-1 (NL4-3 strain) at multiplicity of infection 5 (MOI 5) to obtain more than 90% of cells expressing p24 24 hours after infection, as measured by flow cytometry analysis (Supplementary Fig. S1A). Mock-infected cells were grown in parallel as a control. Cytoplasmic lysates from HIV-1 and mock-infected cells were recovered at 0, 1, 12, 24 and 36 hours post infection (hpi) to monitor changes in cytoplasmic transcript abundance and ribosome loading by RNA-seq and Ribosome Profiling (Ribo-seq), respectively (Fig. 1A and Supplementary Data 1). In parallel total protein lysates were collected and analyzed by mass spectrometry to monitor protein abundance levels (see Supplementary Data 2). For the time course ribosome profiling experiments, micrococcal nuclease was used to obtain ribosome footprints instead of RNase I, which has been shown to lead to ribosome degradation[41–43]. However, because ribosome tri-nucleotide periodicity is not as clear when working with micrococcal nuclease (Supplementary Fig. S1B, top panel), some samples were also prepared with RNase A + T1 and RNase I in order to track the reading frame of ribosomes more accurately (Supplementary Fig. S1B).

Comparison of RNA-seq and Ribo-seq reads at 1hpi indicates that infection induces a mild (less than two-fold) but significant change in the abundance of a handful of cellular transcripts (Fig. 1B, Top-left panel). However, their ribosome-loading is not affected, indicating that they have probably not engaged efficiently with the translation machinery or that the cell is buffering their expression at the translational level. At 12hpi, the overall distribution of RNA-seq and Ribo-seq reads correlates well and indicates global but mild (mostly not-significant) changes in transcript abundance (Fig. 1B, Top-right panel). At 24 and 36hpi, we observed a strong and global effect on gene expression, with thousands of cellular transcripts being differentially expressed (Fig. 1B, Bottom panels). Surprisingly, most changes are driven at the level of cytoplasmic transcript abundance (which can result from changes in transcription, nuclear export or transcript stability), with only a minor subset of genes (423 in total) being regulated at the translational level 36hpi.

To perform differential gene expression and gene-ontology analysis, we further used ribosome profiling reads. Indeed, these correspond to the combination of transcript abundance and ribosome loading and better correlate with protein output than RNA-seq reads (Supplementary Fig. S3A). Gene ontology analysis on up-regulated genes at 36hpi shows an enrichment for proteins involved in cell communication, signaling, immune system processes, and cholesterol pathway, among others (Fig. 1C). Down-regulated genes are enriched in RNA splicing, rRNA processing, ribosomal large and small subunit biogenesis, interleukin-7-mediated pathway, cell cycle, regulation of translation, IRES-dependent viral translation initiation (Fig. 1C). Gene ontology analysis performed on differentially translated transcripts (423 genes in total 36hpi, 301 up-regulated and 122 down-regulated) did not show any enrichment of specific functional category. Nevertheless, translationally down-regulated transcripts have significantly lower GC content in their CDS and longer CDS than control genes and vice versa for translationally up-regulated transcripts (Supplementary Fig. S1C). This suggests a possible role of the endogenous host

https://doi.org/10.1038/s41467-025-56772-3

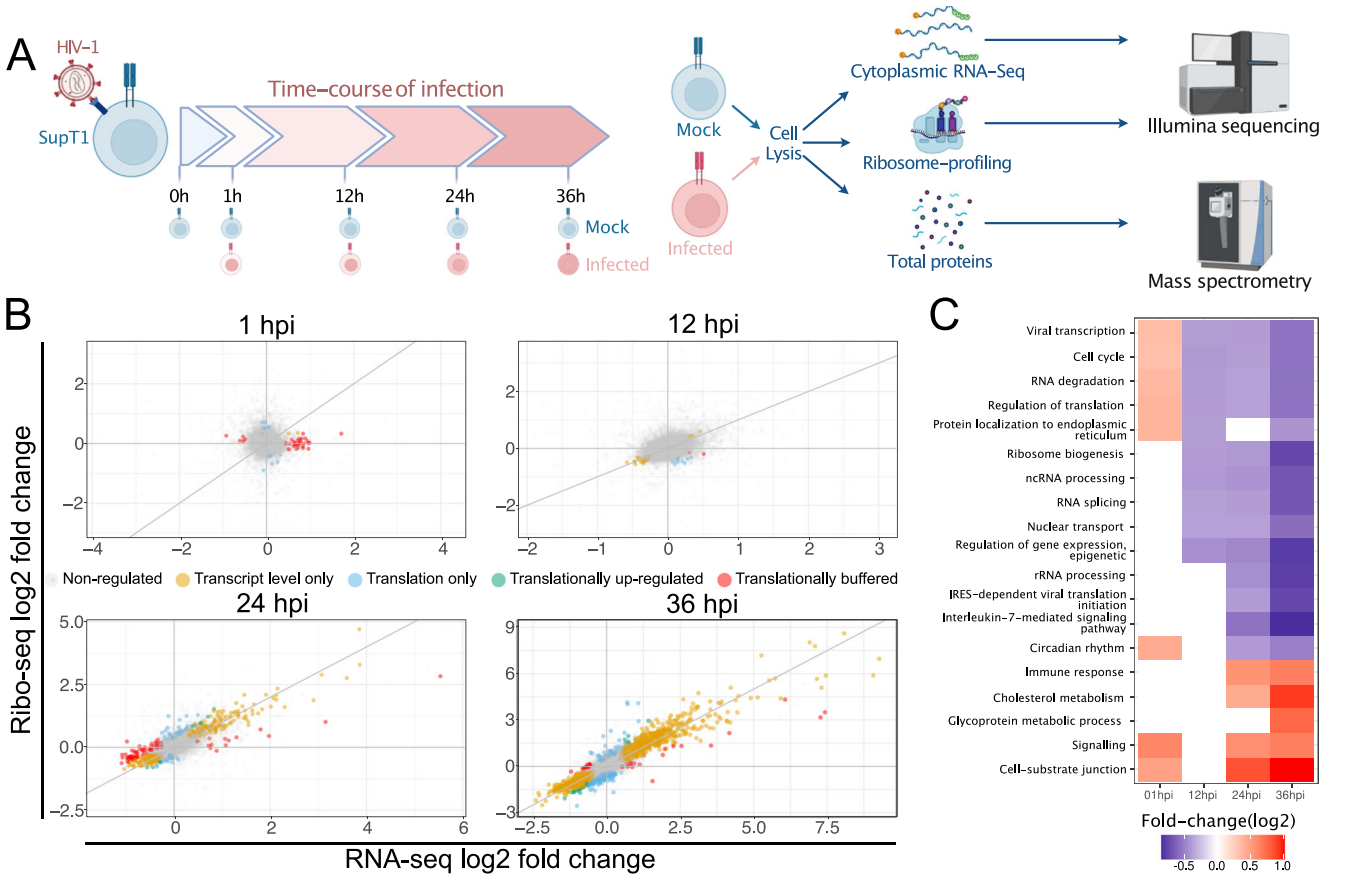


Fig. 1 | Transcriptional and translational changes in HIV-1 infected cells. **A** Schematic representation of the procedure to monitor transcript abundance and translation in HIV-1 infected cells (Created in BioRender. Ricci, E. (2022) BioRender.com/k47z556). Briefly, SupT1 cells were infected or not (Mock) with HIV-1 (NL4.3 strain) at MOI 5. At 0, 1, 12, 24 and 36 hours post infection (hpi), cells were lysed to recover the cytoplasmic fraction and prepare ribosome profiling and RNA-seq libraries subjected to high-throughput sequencing on the Illumina Hiseq platform ($n = 4$ independent experiments). Total cell lysates were recovered for mass spectrometry analysis ($n = 4$ independent experiments). **B** Scatter-plot of the fold-change (log2) in cytoplasmic RNA-seq and Ribo-Seq of the Mock-infected and HIV-1 infected cells at each time point of infection. Orange dots ("Transcript-level only") corresponds to genes exclusively regulated at the transcript abundance level. Blue dots ("Translation only") correspond to genes which display differences in ribosome occupancy while transcript abundance remains unchanged. Green dots ("Translationally regulated") correspond to genes with significant changes in transcript abundance and significantly further changes, in the same direction, in ribosome occupancy upon infection. Red dots ("Translationally buffered") correspond to transcripts displaying significant changes in transcript abundance but for which there is compensation at the translational level to maintain unchanged ribosome occupancy levels upon infection. **C** Gene ontology analysis of differentially expressed genes at each time point.

restriction factor Schlafen 11 (SLFN11), which inhibits HIV-1 replication in a GC content and codon usage dependent manner[44].

Global translation rates are modestly affected by HIV-1 infection

In the absence of synthetic exogenous RNA spike-in controls, high-throughput sequencing approaches, such as RNA-seq and ribosome profiling, can inform about relative changes in gene expression but are not able to accurately monitor bulk changes in intracellular transcript abundance or ribosome occupancy[45]. To test whether HIV-1 infection is accompanied by changes in overall translation efficiency in infected cells, we probed absolute levels of protein synthesis in infected cells using O-propargyl-puromycin (OPP), a clickable puromycin analog that incorporates into elongating polypeptide chains[46] (Fig. 2A). For this, SupT1 cells were mock-infected or infected with HIV-1 and OPP incorporation measured at 12, 24 and 36hpi (Fig. 2B and Supplementary Fig. S9). As a positive control, we infected HEK293T cells with Sindbis virus (SINV), an alphavirus that induces strong translation shutoff in host infected cells[47]. As expected, SINV infection induced a strong decrease in OPP incorporation as soon as 12hpi (Supplementary Figs. S2A and S13). However, no significant differences in OPP incorporation were detected between mock-infected and HIV-infected cells at any of the tested time points (Fig. 2B–E). A slight decrease of translation was nevertheless detected for both mock-infected and HIV-1 infected cells at 36hpi, probably due to the higher density of cells in the culture well.

To further validate these results at later time points of infection, SupT1 cells were infected at MOI 0.5 with VSVg pseudotyped, genetically modified single-round HIV-1 virions in which there is a frameshift mutation in the envelope coding sequence and the GFP coding sequence is placed upstream of Nef. This strategy allows to track infected cells and avoid the important cell mortality that is associated with the replication and spreading of wild-type HIV-1 virions at longer time-points of infection. To monitor translation and infection, the level of OPP incorporation and GFP expression was measured at 24, 36, 48, 72 and 96hpi. As observed (Fig. 2C and Supplementary Fig. S10), translation rates in GFP positive cells (indicative of HIV-1 replication) increased slightly at 24 h compared to the mock-infected control and to the GFP negative cells from the infected condition (corresponding either to non-infected cells or cells in which expression of viral proteins was still not detected). At 72 h after infection, translation efficiency increased both for GFP positive and negative cells from the infected condition compared to mock-infected cells, while at 96 h post-infection translation efficiency was similar in all tested conditions.

https://doi.org/10.1038/s41467-025-56772-3

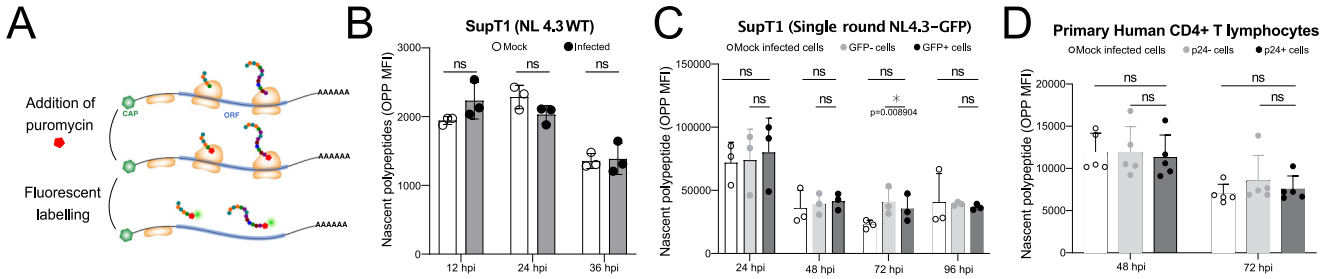


Fig. 2 | Transcriptional and translational changes in HIV-1 infected cells. A (Left panel) Scheme describing nascent protein labeling using O-propargyl-puromycin (OPP). Briefly, cells are incubated with OPP and fixed in paraformaldehyde before a fluorophore is covalently linked through a click-reaction. Cells are then analyzed by flow cytometry to monitor signal intensity at a single-cell level. **B** Flow cytometry analysis ($n = 3$ independent experiments) of OPP signal in SupT1 cells infected with HIV-1, 12, 24 and 36 hpi. **C** OPP signal in SupT1 cells infected with single-round recombinant HIV-1 virions coding for GFP at 24, 48, 72 and 96 hpi ($n = 3$ independent experiments). **D** OPP signal in primary Human CD4 + T cells infected with HIV-1, 48 and 72 hpi and either positive or negative for p24 expression (p24+ or p24-). Results from four independent donors are displayed separately for each time point tested ($n = 5$ independent experiments). **B**–**D** Multiple paired t tests were performed to compare OPP incorporation signals between control and infected samples, * corresponds to a p-value < 0.05. Barplots represent the average value of all biological replicates and error bars correspond to the standard-deviation. Source data are provided as a Source Data file.

Finally, we infected activated primary CD4 + T-cells obtained from four different donors (one of the donors was tested with two different HIV-1 virion production batches) with wild-type HIV-1 virions (NL4.3 strain) and monitored translation rates at 48 and 72hpi among p24-positive and -negative cells (Fig. 2D and Supplementary Fig. S11). In this more physiological set-up, we were not able to detect any significant change in cellular translation rates between mock infected and p24-positive cells nor between p24-negative and -positive cells within the infected condition, at both time points tested.

Taken together, while no global translational shut-off was detected upon infection, we nevertheless observed important changes in transcript abundance of host mRNAs and a milder effect on translation of specific host mRNAs in a GC-content dependent manner.

A post-translational regulatory layer modulates protein abundance of specific gene groups independently of changes in transcript-abundance or translation in infected cells

When combining all omics datasets (RNA-seq, ribo-seq and mass spectrometry) and all time-points of infection tested, we were able to obtain robust measurements for a total of 2795 genes. Using this comprehensive database, we performed a global analysis of changes in gene expression of cellular genes during the course of infection. At 12 h post infection, changes in protein abundance (as measured by mass spectrometry) are mild and not significantly correlated to changes in ribo-seq or RNA-seq expression (Fig. 3A and Supplementary Fig. S3B). At 24 h and 36 h post-infection, changes in protein abundance are overall positively correlated with changes in ribo-seq and RNA-seq (Fig. 3A and Supplementary Fig. S3B), although there is a stronger dispersion at 36 h post-infection, indicating an increased role of post-translational regulation at this time-point. Interestingly, cellular proteins known to be actively targeted by viral proteins (such as CD4 and CD28 which are targeted for degradation by Nef[48]) clearly appear down-regulated at the protein level, while their transcription/translation is on the contrary slightly stimulated at 36 h post-infection (Fig. 3A). To gain a clearer view of the different changes occurring at the transcriptional, translational and post-translational levels during infection, we performed gene clustering analysis using all datasets (RNA-seq, ribo-seq, Mass Spectrometry) and all infection time-points available (Fig. 3B and Supplementary Fig. 3C). This analysis identified 9 clusters of genes that shared similar behaviors across infection time-points. Clusters 1, 2, 3 and 8 correspond to genes displaying concordant changes in RNA-seq, ribo-seq and protein abundance indicating no major changes at the post-translational level (at least with respect to protein abundance). Among these, clusters 1 and 2 (displaying a gradual increase in all three measurements across infection time) are enriched in mitochondrial proteins involved in oxidative phosphorylation (cluster 1) and proteins involved in the lipid/cholesterol pathway as well as glycoproteins and proteins involved in T cell migration (cluster 2). Clusters 3 and 8 (displaying a gradual decrease in all three measurements across infection time) are enriched for factors involved in non-coding RNA metabolism (including tRNAs and nuclear non-coding RNAs), ribosome biogenesis and protein folding.

Surprisingly, the remaining clusters (clusters 4, 5, 6, 7 and 9) correspond to groups of genes that display discordant changes in protein abundance compared to RNA-seq/ribos-seq (Fig. 3B and Supplementary Fig. 3C). Clusters 4 and 9 contains genes for which no significant changes are observed at the transcript abundance and translation level but a strong decrease in protein abundance is observed. These clusters are enriched in genes involved in the translation process (translation initiation and elongation, tRNA processing) and the mitotic cell cycle. Clusters 5 and 7 correspond to genes in which RNA-seq and ribo-seq abundance is down-regulated but display a relatively strong protein abundance increase. These clusters are enriched in factors involved in mitochondrial translation, mitochondrial transmembrane transport and mitochondrial inner membrane proteins, as well as RNA-binding proteins and proteins involved in ribosome biogenesis, Finally, cluster 6 contains genes with increase in RNA-seq and ribo-seq abundance but a decrease in protein abundance. This cluster is enriched for factors involved in endosomal transport and proteasome ubiquitin-dependent protein degradation.

Taken together, these results indicate that although there is an overall positive correlation between changes in transcript, ribosome loading and protein abundance observed upon infection, a complex post-translational regulatory layer participates in buffering changes in transcript abundance and mRNA translation for a subset of genes with specific cellular functions. This is particularly striking for genes related to mitochondrial functions and the oxidative phosphorylation pathway for which the cell appears to compensate for a decrease in transcript abundance and translation by stabilizing the corresponding proteins. On the contrary, factors involved in the mRNA translation process and mitotic cell cycle are exclusively down-regulated at the protein level.

Dynamics of HIV-1 cytoplasmic transcript abundance

HIV-1 proviral DNA is transcribed by RNA polymerase II as a single-transcription unit that can be spliced to produce close to 100 different transcripts coding for viral proteins[2,3]. These transcripts are expressed in a temporally coordinated manner to produce all accessory and core viral proteins required to efficiently assemble new viral particles. To quantify cytoplasmic levels of HIV-1 transcripts, we collected data from

https://doi.org/10.1038/s41467-025-56772-3

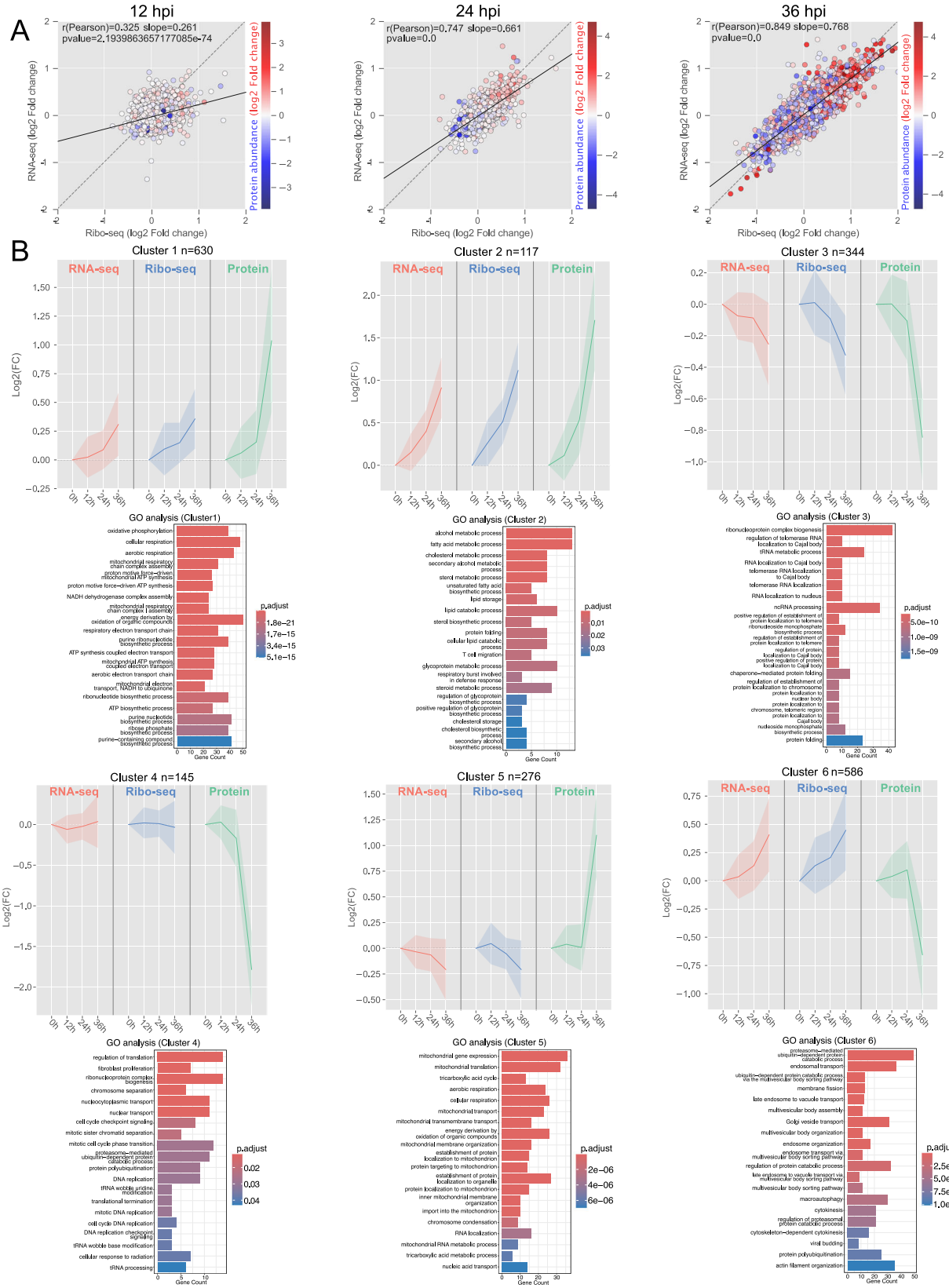


Fig. 3 | Multi-omics analysis of gene expression changes in HIV-1 infected cells. **A** Scatter plots of log2 fold changes in ribo-seq (*x* axis), RNA-seq (*y* axis) and protein (color coded) abundance at 12, 24 and 36 hours upon infection ($n = 4$ independent experiments). **B** Gene clustering analysis taking into account changes in RNA-seq, ribo-seq and protein abundance at all time points of infection. Mean trajectories (bold lines) and standard-deviation (light colored surfaces) are depicted (top panels) as well as gene ontology analysis (bottom panels) for each cluster.

https://doi.org/10.1038/s41467-025-56772-3

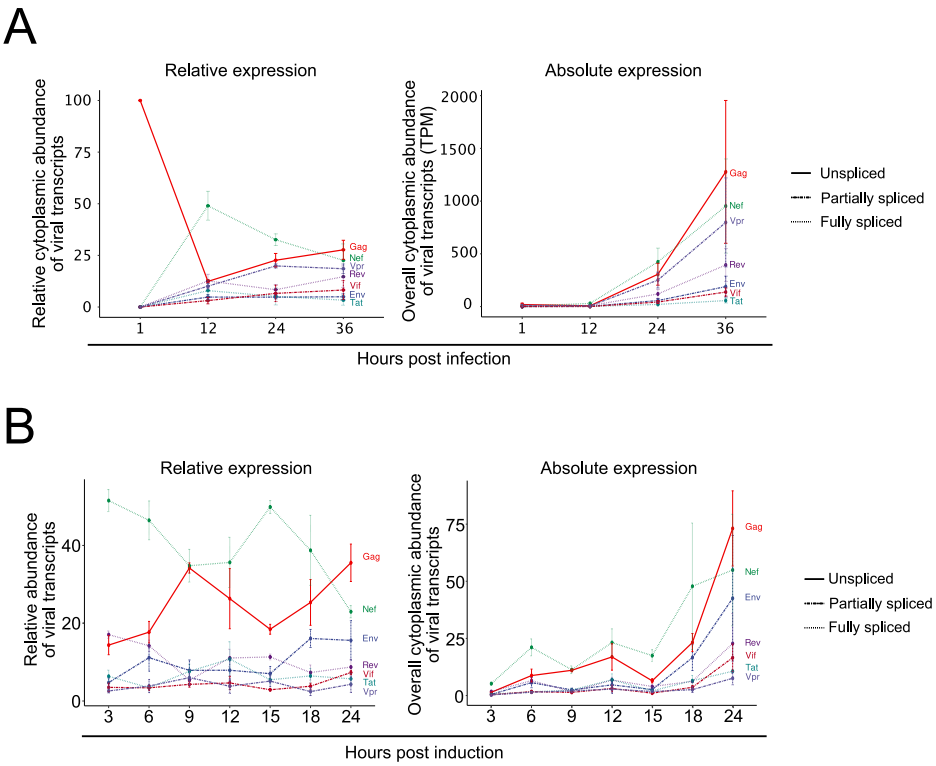


Fig. 4 | Relative and absolute expression of viral transcripts during infection. **A** (Left panel) Relative cytoplasmic amount of viral transcripts (the sum of all transcripts at any given time point corresponding to 100%) in SupT1 cells, 1, 12, 24 and 36 hpi. (Right panel) Overall abundance of viral transcripts within RNA-seq libraries displayed in transcripts per million (TPM) ($n = 4$ independent experiments). **B** Relative (left panel) and overall (right panel) abundance of viral transcripts in U937 cells bearing a latent HIV-1 provirus integrated in their genome, 3, 6, 9, 12, 15, 18 and 24 hours after induction of proviral DNA expression using PMA and ionomycin ($n = 3$ independent experiments). Error bars in figures correspond to the Standard Error of the Mean (SEM).

long read sequencing (Nanopore) of NL4-3 infected primary CD4 T cells to identify all HIV-1 transcripts expressed in infected cells[2] and used this information as a template to reconstruct isoform expression from our RNA sequencing reads. A simulated in silico dataset of RNA sequencing reads generated from the list of known HIV-1 transcripts, with a similar length distribution to our RNA sequencing libraries, validated our deconvolution approach (Supplementary Fig. S4A). At 1hpi, we could only detect unspliced viral RNAs corresponding to incoming genomic RNAs (Fig. 4A and Supplementary Fig. S4B). At 12hpi, upon reverse-transcription and integration of most of the proviral DNA, viral transcripts were dominated by the fully spliced transcript coding for Nef, although we could still detect important levels of unspliced RNAs in the cytoplasm of infected cells (Fig. 4A). The relative amount of fully spliced transcripts (coding for Nef, Tat, Rev) strongly decreased at 24 and 36hpi (Fig. 4A). On the contrary, relative abundance of single-spliced and unspliced transcripts, which require Rev for their nuclear export[49–51], increased at 24 and 36hpi (Fig. 4A). These results are in agreement with the known expression kinetics of viral transcripts during the replication cycle of HIV-1[52,53].

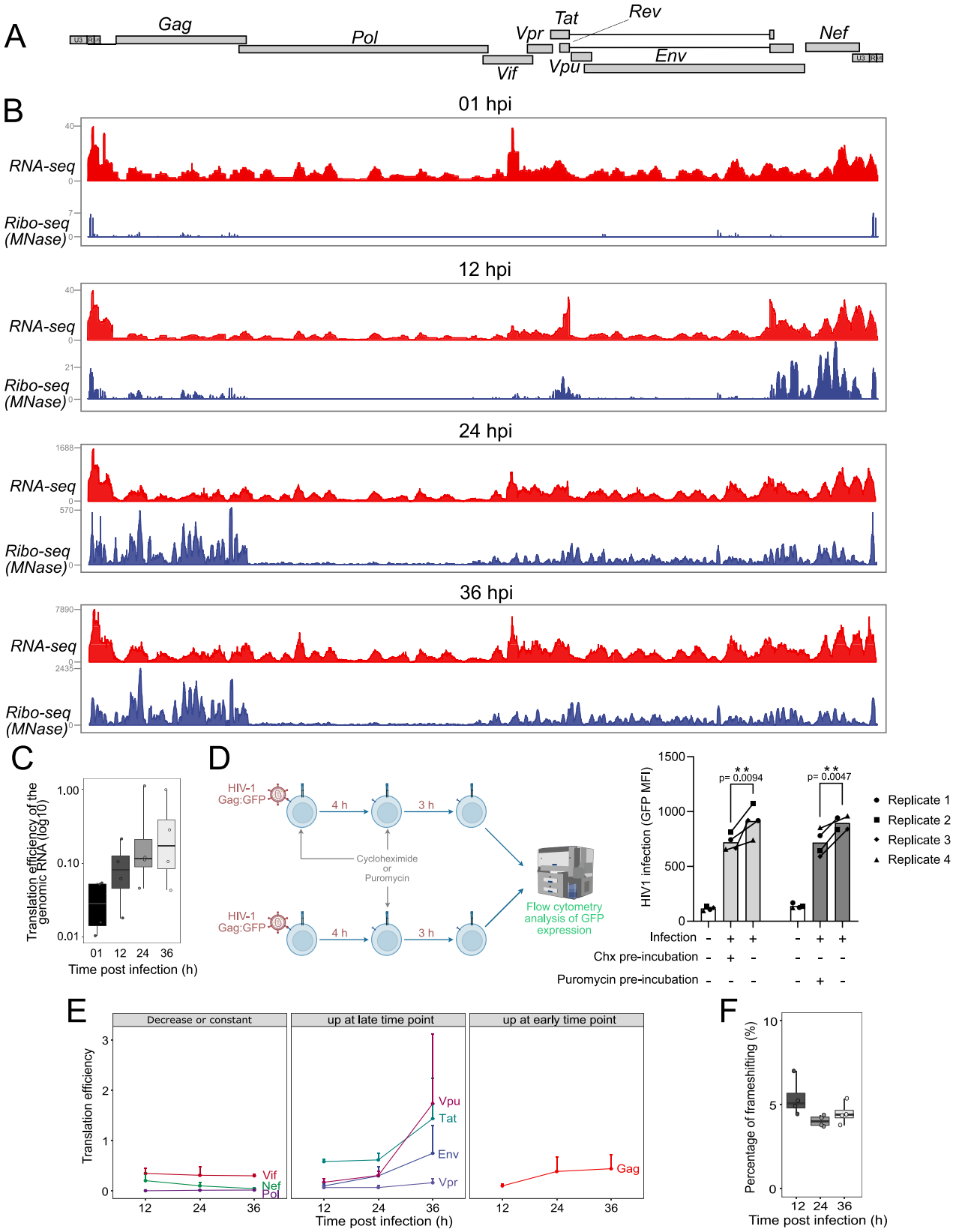
To bypass the heterogeneity introduced by variable timing in reverse-transcription of incoming genomic RNAs, integration and transcription of proviral DNAs, we performed a time-course analysis of cytoplasmic RNA expression and ribosome profiling from U937 cells in which a stably integrated HIV-1 provirus lies in a latent form[54]. In this system, expression of viral transcripts is triggered in a synchronous manner through incubation of cells with phorbol 12-myristate 13-acetate (PMA) and ionomycin. U937 with a latent integrated HIV-1 provirus were induced and cytoplasmic extracts recovered at 0, 3, 6, 9, 12, 15, 18 and 24 h to monitor RNA levels and mRNA translation. Differential gene expression analysis indicated a strong impact of PMA/ionomycin treatment on cellular gene expression and translation 24 h post-induction (Supplementary Fig. S4C). The overlap between differentially expressed transcripts obtained from PMA/ionomycin activated U937 cells and SupT1 cells infected with HIV-1 virions is nevertheless moderate (Supplementary Fig. S4D), suggesting that changes observed in U937 are mainly driven by PMA/ionomycin and not by expression of viral transcripts. Similarly to results obtained in SupT1 cells, viral transcripts were dominated by the fully spliced transcript coding for Nef at early time points (Fig. 4B). Expression of unspliced genomic RNA was also abundant at early time-points and constantly increased throughout time in parallel with the decrease in Nef expression. Surprisingly, relative expression of all other viral transcripts was much lower than that of Nef and the genomic RNA.

These results indicate that cytoplasmic levels of fully and partially spliced viral transcripts, as well as that of the unspliced genomic RNA are regulated in a temporal manner, as extensively described in the literature[52,53]. However, expression and translation of unspliced genomic RNAs is already abundant at early time points of infection.

Translational landscape of HIV-1 transcripts

Having characterized the dynamics of the cytoplasmic viral transcriptome, we set out to monitor its translational landscape. As described above, we could detect RNA-seq reads corresponding to cytoplasmic incoming unspliced HIV-1 genomic RNAs as soon as 1hpi (Fig. 4A). Surprisingly, we could also detect ribosome footprints on viral transcripts at 1 h post infection, suggesting that a fraction of incoming viral RNAs undergo translation (Fig. 5A and B). Presence of HIV-1 mapping reads at 1 h post-infection did not seem to result from cross-contamination with other samples during library preparation since the percentage of HIV-1 mapping reads in the 1 h post-infection libraries was more than 20 fold higher than in the control mock-infected cells (Supplementary Data 3). As

https://doi.org/10.1038/s41467-025-56772-3



expected, ribosome footprints on incoming genomic RNAs were mainly located on the Gag coding sequence and very few on Pol, which depends on ribosome frameshifting for translation (Fig. 5B). However, we could also detect a significant amount of ribosome profiling signal downstream of Pol, which was unexpected because unspliced mRNAs should only code for Gag and Gag-Pol (Supplementary Fig. S4B). Ribosome profiling reads downstream of the Gag-Pol coding sequence could correspond to bona-fide 80S ribosomes footprints originating from alternative translation initiation mechanisms or to footprints of RNA-binding proteins co-sedimenting with 80S ribosomes. To exclude this last option, HEK293T cells expressing Flag-tagged ribosomes from the

Fig. 5 | Translational landscape of viral transcripts. A HIV-1 genomic structure. **B** Distribution of RNA-seq (red) and ribosome profiling (blue) reads across the HIV-1 genome in SupT1 cells at 1, 12, 24 and 36 hpi ($n = 4$ independent experiments). **C** Translation efficiency of the genomic RNA (across both Gag and Pol coding sequences) at each time point of infection ($n = 4$ independent experiments). Boxplots are defined as minima, 1st quartile, median, 3th quartile and maxima. **D** Translation of incoming viral genomic RNAs as tested by infecting cells pre-incubated or not with cycloheximide or puromycin with a recombinant replication-competent HIV-1 virus bearing a GFP sequence within Gag ($n = 4$ independent experiments). Barplots correspond to the mean value of all biological replicates. A one tailed, paired t test was performed to compare samples. Source data are provided as a Source Data file (Partially created in BioRender. Ricci, E. (2022) BioRender.com/k47z556 and using an Illustration from NIAID NIH BIOART Source https://bioart.niaid.nih.gov/bioart/160). **E** Translation efficiency of canonical viral mRNAs, 12, 24 and 36 hpi ($n = 4$ independent experiments). Points correspond to the mean value of all biological replicates and error bars correspond to the Standard Error of the Mean (SEM). **F** Percentage of Gag-Pol ribosome frameshifting at each time point of infection ($n = 4$ independent experiments). Boxplots are defined as minima, 1st quartile, median, 3th quartile and maxima.

endogenous eL8 (RPL7a) locus were infected with VSV-G pseudotyped HIV-1 virions and, 1 h after infection, ribosomes were immunoprecipitated and used to prepare ribosome footprint libraries (Supplementary Fig. S4B bottom panel). As observed, ribosome profiling fragments (RPFs) obtained upon ribosome immunoprecipitation also indicate translation of incoming viral genomic RNAs and the presence of ribosomes downstream of the Pol coding sequence, with no detectable reads overlapping known splice junctions within viral transcripts (Supplementary Fig. S4B bottom panel).

To quantify the extent of incoming genomic RNAs undergoing translation, we calculated ribosome density on the Gag and Gag-Pol coding sequences at 1, 12, 24 and 36hpi (Fig. 5C). As expected, Translation efficiency (TE; calculated by normalizing ribosome footprinting reads in the Gag-Pol coding region by RNA-seq reads from the same region) was significantly lower at 1 h post infection than at all other time-points (2.46, 4.42 and 5.75 times lower at 1hpi than at 12, 24 and 36 hpi, respectively). This suggests that only a small fraction of incoming genomic RNAs are capable of uncoating from viral cores and undergo translation instead of reverse-transcription (Fig. 5C). To further confirm this finding, we produced genetically-modified replication-competent HIV-1 virions carrying a GFP coding sequence between the MA and CA domain of Gag, flanked by two functional protease cleavage sites[55]. As a control, virions were incubated for 4 hours (at MOI of 5) with SupT1 cells that were pre-incubated with cycloheximide or puromycin (impeding any translation of incoming viral genomic RNAs) in order to take into account the GFP signal corresponding to GFP proteins loaded within virions by producer cells. In parallel, virions were incubated for 4 hours with actively translating SupT1 cells (allowing for translation of incoming viral genomic RNAs). After 4 hours, the culture medium was replaced with fresh medium containing cycloheximide or puromycin and cells were further incubated for 3 hours to allow for any newly synthesized GFP proteins to fold and become competent for fluorescence (Fig. 5D, left panel and Supplementary Fig. S12). As observed (Fig. 5D, right panel), cells infected in the absence of translation inhibitors displayed significantly higher levels of GFP signal compared to cells pre-incubated with translation inhibitors before infection. These results strongly argue in favor of translation of a fraction of incoming viral genomic RNAs and indicate that a fraction of infected cells in which replication is probably aborted, nevertheless express Gag and Gag/Pol.

At later time points, translation of viral transcripts was detected on all canonical coding sequences (Fig. 5B and E). To quantify translation efficiency, we calculated ribosome occupancy from the non-overlapping part of each canonical open reading frame, with exception of Rev which overlaps with Tat and Nef. As observed, ribosome occupancy on Pol is much lower than in all other viral ORFs (Fig. 5E), owing to its specific translation mechanism which depends on the frameshifting of ribosomes translating Gag[31,56]. Using the ratio of ribosome footprints in the Pol and Gag coding sequences, we obtained the percentage of frameshifting, which is close to 5%, in agreement with previous observations[57]. The percentage of frameshifting is robust in the 4 ~ 5% range at all time points, being slightly higher at 12hpi than 24 and 36hpi (Fig. 5F), suggesting that it is not subjected to strong dynamic regulation during the replication cycle.

Ribosome occupancy on viral coding sequences is however dynamically regulated during the course of infection. Ribosome occupancy on Gag increases from 12 to 24 hpi and remains relatively steady at 36hpi. Ribosome occupancy on Env, Vpr and Vpu increases at 36hpi, while on the contrary that of Nef decreases from 12 to 36hpi (Fig. 5E). Taken together, similarly to transcription and splicing of viral transcripts, translation of viral ORFs is tightly regulated throughout the replication cycle.

Mapping translation initiation sites on viral transcripts

U937 cells in which productive replication was induced by addition of PMA/ionomycin (to obtain synchronized expression of viral transcripts) for 0, 3, 6, 9, 12, 15, 18 and 24h were incubated for 15 minutes with the translation initiation inhibitor harringtonine to allow for elongating ribosomes to run-off from translating mRNAs while blocking initiating ribosomes at their translation start sites (Fig. 6 and Supplementary Fig. S5A)[58]. As shown in Fig. 6A, this strategy led to a clear enrichment of ribosomes at the predicted start codon on host transcripts. Enrichment of RPFs at canonical AUG translation initiation sites was also detected in viral transcripts (Fig. 6B and Supplementary Fig. S5A). The HIV-1 genomic RNA has been shown to contain an IRES that is located within the Gag coding region (downstream of the canonical Gag start codon) which allows translation of a N-terminal truncated form of Gag (also known as p40) from an in-frame AUG-start codon[8,10,11,13,59]. We were able to detect a low albeit specific signal corresponding to initiating RPFs at the p40 AUG codon (Fig. 6C), suggesting that the IRES is functional *in cellula* although it is much less efficient than cap-dependent translation within the context of infection.

Multiple translation events from non-AUG start codons in the 5'UTR of viral transcripts

Although we could detect initiating RPFs on canonical AUG initiation codons, the majority of initiating ribosomes within viral transcripts originated from the 5'UTR (Fig. 6B and Supplementary Fig. S5A). This signal was also observed in regular ribosome profiling samples prepared with cycloheximide (in the absence of Harringtonine) in SupT1 and in primary human CD4 + T lymphocytes[60] infected with HIV-1 virions, as well as in U937 cells upon induction of HIV-1 transcription (Fig. 7A). Careful observation of the ribosome profiling signal indicated that it overlapped with small ORFs starting with a near-cognate AUG codon with optimal Kozak contexts (Supplementary Data 4). However, knowing the highly structured nature of the 5'UTR and the fact that multiple RNA-binding proteins (including Gag) could be tightly bound to it and co-sediment with 80S ribosomes upon RNase treatment to obtain ribosome footprints, we could not exclude that these signals may originate from footprints of RNA-binding proteins instead of ribosomes. Nevertheless, Fragment Length Organization Similarity Score (FLOSS) analysis of 5'UTR viral reads[61] from infected primary human T CD4+ lymphocytes and U937 cells, indicated that these reads had a similar length pattern than ribosome footprints mapped to the CDS of cellular coding mRNAs (Fig. 7B). This, strongly suggests that

https://doi.org/10.1038/s41467-025-56772-3

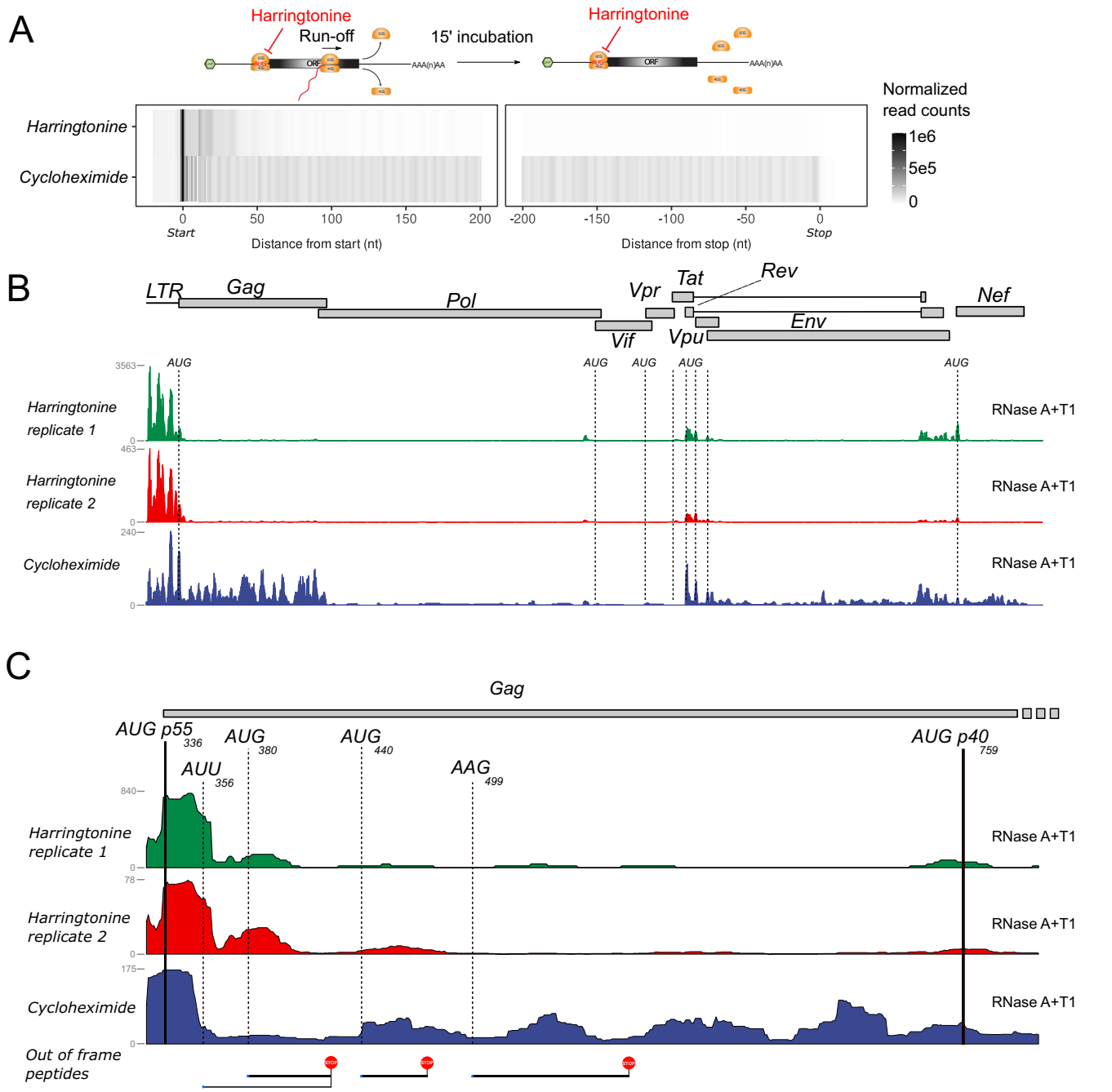


Fig. 6 | Translation initiation sites in viral transcripts. A Distribution of ribosome P-sites around annotated start and stop codons in all cellular transcripts in harringtonine (Green and Red for each biological replicate) and cycloheximide (Blue) libraries. **B** Distribution of ribosome profiling reads across the HIV-1 genome obtained from harringtonine and cycloheximide treated cells. **C** Distribution of ribosome profiling reads in the first 500 nucleotides of the Gag CDS obtained from harringtonine (Green and Red for each biological replicated) and cycloheximide (Blue) treated cells (*y* axis correspond to reads per million values - RPM). The canonical AUG start codon of Gag (p55 isoform) at position +336, as well as the position of other out-of-frame start codons predicted by Ribocode and lastly the position of the Gag (p40 isoform) start codon at position +759 are annotated in the figure.

reads originating from the 5′UTR of viral transcripts are bona-fide ribosome footprints.

To further confirm these results, we prepared ribosome profiling libraries from HIV-1 infected cells using the ribolace kit[62], which relies on a biotinilated puromycin analog to specifically pull-down 80S ribosomes competent for elongation. Ribosome profiling libraries were also prepared from cytoplasmic lysates incubated with 1 M of KCl prior to RNase treatment, in order to dissociate RNA-binding proteins from mRNAs as well as 80S ribosomes that are not engaged in active translation and lack a nascent peptide[63]. Finally, HEK293T cells expressing Flag-tagged ribosomes from the endogenous eL8 (RPL7a) locus were transfected with the pNL4.3 plasmid. Upon transfection, ribosomes were immunoprecipitated with anti-Flag beads (in the presence of nucleases to obtain ribosome mRNA footprints), followed by competitive elution with Flag peptide and preparation of ribosome profiling libraries. Ribosome immunoprecipitation bypasses the use of sucrose sedimentation to recover 80S ribosomes and avoids co-sedimenting RNP complexes that could contaminate samples. As observed in Fig. 8A, we could observe RPFs in the 5′UTR of viral transcripts in all tested protocols, strongly arguing for genuine uORF

https://doi.org/10.1038/s41467-025-56772-3

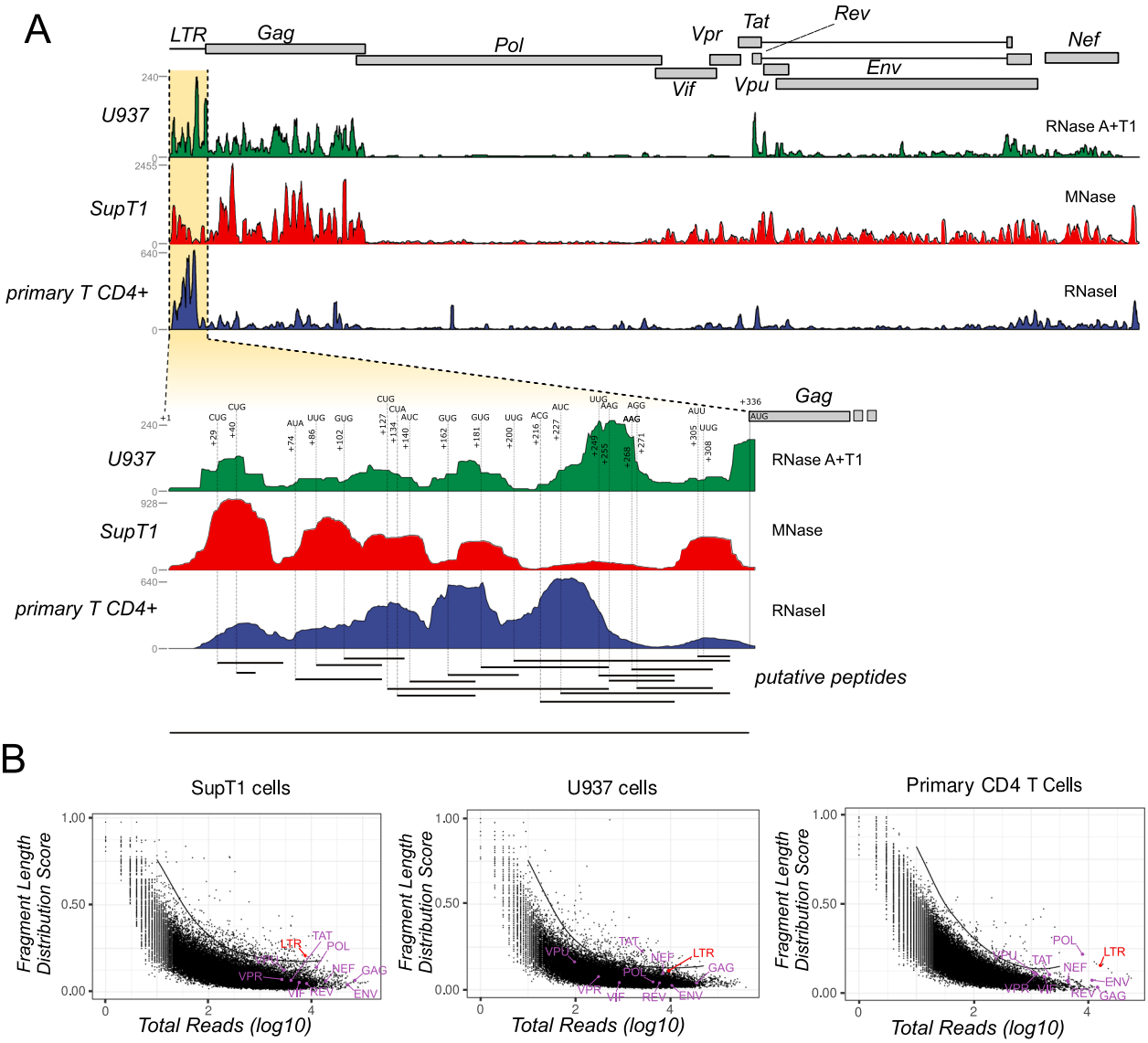


Fig. 7 | Non-AUG translation initiation sites in the 5′UTR of viral transcripts. **A** (Top panel) Distribution of ribosome profiling reads across the HIV-1 genome from infected U937 (Green, $n = 3$ independent experiments), SupT1 (Red, $n = 4$ independent experiments) and primary CD4 + T lymphocytes (Blue, $n = 2$ independent experiments). All y axis correspond to reads per million values - RPM. (Bottom panel) Close-up view of the unspliced HIV-1 5′UTR showing the distribution of ribosome profiling reads and the position of putative non-AUG start codons as well as the open-reading frames of putative peptides produced from non-AUG start codons. The y-axis corresponds to reads per million of sequenced reads (RPM). **B** Distribution of FLOSS (Fragment Length Distribution Score) values for cellular and viral transcripts computed from ribosome profiling reads from SupT1 (left panel), U937 (middle panel) or primary CD4 T cells (right panel).

translation events. As a final confirmation of these results, we created reporter constructs consisting of the 5′UTR of the HIV-1 genomic RNA, a mutant version where most non-AUG initiation codons were replaced by UAG stop codons (Fig. 8B, see No-uORFs 5′UTR) and a second mutant where 4 of the putative non-AUG initiation codons were replaced by AUG codons (Fig. 8B, see AUG-uORFs 5′UTR), followed by the renilla luciferase coding sequence. Compensatory mutations were introduced in mutant 5′UTRs in order to maintain their overall secondary structure, which was probed by SHAPE[64–66] (Supplementary Fig. S6A and B). These reporter plasmids were in vitro transcribed to generate capped and polyadenylated RNAs that were transfected into HEK293T cells. 2h30 post-transfection, cells were collected to prepare ribosome profiling samples. As shown in Fig. 8B RPFs were detected in the renilla luciferase coding sequence for all constructs tested. On the contrary, RPFs in the 5′UTR were only detected in the wild-type and AUG-uORFs mutant transcripts while almost no signal could be detected in the No-uORFs 5′UTR.

Taken together, these results strongly suggest the presence of multiple non-AUG translation initiation sites occurring in the 5′UTR of viral transcripts that might drive productive translation of small ORF upstream of the canonical viral ORF. Nevertheless, for reads located in the R region we cannot discriminate whether they originate from the 5′UTR or 3′UTR, however, their presence in Ribo-seq experiments performed upon reporter mRNA transfection (Fig. 6B) suggests that at least a fraction of these reads could originate from translation at the 5′UTR.

uORF translation is conserved among retroviruses

Having shown translation from uORFs originating from HIV-1 unspliced and spliced transcripts, we tested whether this was a conserved feature among retroviruses. For this, we prepared ribosome profiling from HIV-2 (Lentivirus) and HTLV-1 (Deltaretrovirus) infected cells. As shown in Supplementary Fig. S6C and D, non-AUG translation was detected at uORFs located in the 5′UTR of viral transcripts

https://doi.org/10.1038/s41467-025-56772-3

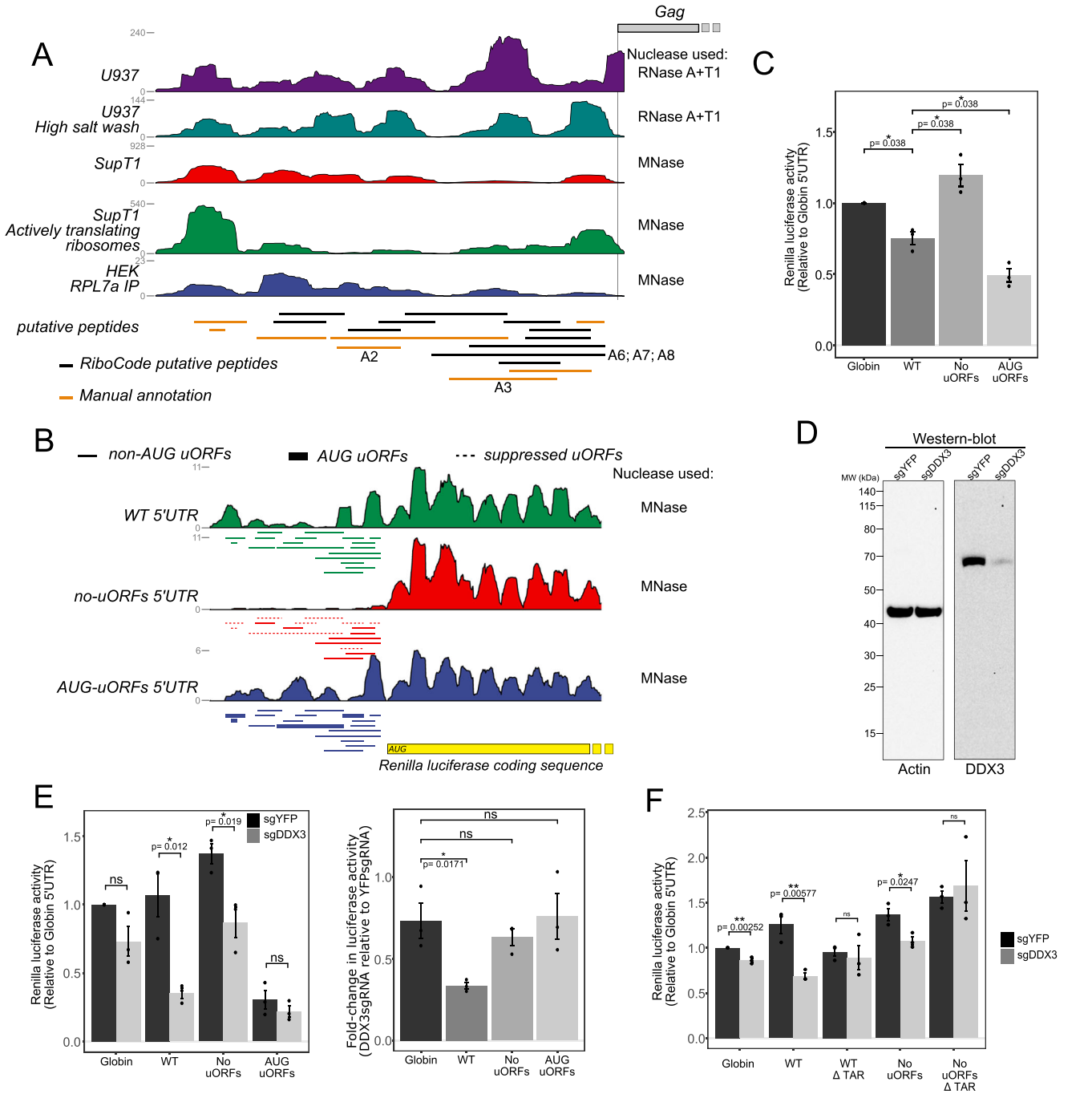


Fig. 8 | Productive translation from uORFs in viral transcripts and role of DDX3 in alleviating the negative impact of uORFs on translation from downstream main ORFs. A Ribosome profiling reads across the 5′UTR of unspliced viral mRNAs in U937 lysates incubated with or without 1 M KCl, Ribolace experiments in SupT1 cells, and RPL7a IPs from HEK293T cells ($n = 2$ independent experiments). Reads are expressed as RPM. **B** Distribution of ribosome profiling reads on luciferase coding reporter transcripts bearing the wild-type HIV-1 5′UTR (WT 5′UTR), a mutant version in which all non-AUG uORF start codons were mutated to UAA stop codons (no uORFs 5′UTR; uORFs with mutated initiation codons are labeled as discontinued lines), a mutant in which all non-AUG uORF start codons were mutated into AUG codons (AUG uORFs 5′UTR; uORFs with mutated initiation codons are labeled as bold lines) ($n = 2$ independent experiments). Reads are expressed as RPM. **C** Relative renilla luciferase activity (normalized against the Globin 5′UTR reporter mRNA) of the different 5′UTR reporter mRNAs upon in vitro translation in the rabbit reticulocyte lysate ($n = 3$ independent experiments) (**D**), Western-blot analysis of β-actin (left panel) and DDX3 (right panel) expression in HEK293T cells transfected with plasmids coding for the Cas9 and a sgRNA targeting the GFP sequence (sgGFP) or the DDX3 coding sequence (sgDDX3). ($n = 3$ independent experiments with similar results). **E** Relative luciferase activity of 5′UTR reporters in sgDDX3 or sgGFP-transfected HEK293T cells (left), and fold-change in luciferase activity upon DDX3 knockdown (right) ($n = 3$ independent experiments). **F** Relative renilla luciferase activity (normalized against the Globin 5′UTR reporter mRNA) from viral reporter mRNAs in which the TAR loop is present or absent, transfected into HEK293T cells in which DDX3 expression was knocked-down (sgDDX3) or not (sgYFP) using CRISPR-Cas9 ($n = 3$ independent experiments). For panels (**C**, **E** and **F**), a two-tailed Student t-test was performed to assess the differences between the mean values of compared conditions. Barplot values correspond to the mean value of all biological replicates and error bars correspond to the Standard Error of the Mean (SEM). Source data are provided as a Source Data file.

Table 1 | Clinical, and virological characteristics of study participants

Sample name used in the study	Age	Sex	Number of year of infection	Number of year of treatment	Treatment	Mean CD4 + T cell counts cells/mm3)	Viral load (copies)
EC-1	50-59	F	31			1106	40
EC-4	50-59	M	30			647	61
EC-2	NA	NA	NA			NA	NA
EC-3	NA	NA	NA			NA	NA
ART-2	60-69	M	22	22	edurant + kivexa	693	<40
ART-3	60-69	M	32	21	atripla	488	<40
ART-1	NA	NA	NA	NA	NA	NA	<40

for both viruses. Interestingly, RPFs originating from non-AUG putative uORFs have also been detected in the 5′UTR of Murine Leukemia Virus (Gammaretrovirus)[67]

These results indicate that non-AUG translation occurring at uORFs within the 5′UTR of viral transcripts is a conserved feature of many retroviruses.

DDX3 bypasses the negative effects of uORF on HIV-1 translation

Upstream open-reading frames generally inhibit translation of the main ORF by restricting the fraction of initiating/scanning 40S ribosomal subunits that can reach the canonical start codon[68]. To test whether uORFs located in the 5′UTR of viral transcripts restrict translation from the main ORF, we used the HIV-1 WT 5′UTR, AUG-uORFs 5′UTR and No-uORFs 5′UTR renilla luciferase reporter mRNAs described in Fig. 8B in addition to a control reporter mRNA bearing the 5′UTR of the human beta-globing transcript. To compare translation efficiencies from the different reporter RNAs, we first performed in vitro translation reactions using rabbit reticulocyte lysate (RRL). As expected, translation driven by the wild-type HIV-1 5′UTR is less efficient than that driven by the beta-globin 5′UTR in RRL (Fig. 8C). Translation of the No-uORFs mutant RNA is also more efficient than that of the wild-type RNA, therefore suggesting that uORFs are able to repress translation of the main ORF by preventing scanning 40S to efficiently reach the AUG canonical codon (Fig. 8C). Translation of the AUG-uORFs mutant RNA is however less efficient than that of the wild-type version, therefore suggesting that initiation on non-canonical codons is not as efficient as that occurring when they are replaced by AUG codons. The DEAD box helicase DDX3 has been shown, in yeast and humans, to mediate bypass of near-cognate start codon uORF initiation and stimulate translation from the main ORF in a large number of cellular transcripts with complex secondary structures similar to those present in the HIV-1 5′UTR[69–71]. Furthermore, efficient translation of the HIV-1 genomic RNA has been previously shown to rely on DDX3 in a TAR-dependent manner[29,30]. To test whether DDX3 is involved in bypassing uORF translation from the HIV-1 5′UTR, we down-regulated DDX3 in HEK293 using CRISPR-Cas9 and transfected the reporter RNAs to test their translation. As expected, translation driven by the short beta-globin 5′UTR is not significantly sensitive to DDX3 inhibition, while that driven by the wild-type HIV-1 is strongly reduced upon DDX3 down-regulation (Fig. 8D, E). Translation driven by the UAG-mutant was significantly more resistant to DDX3 depletion than that of the wild-type RNA thus indicating that removal of upstream non-AUG start codons from the HIV-1 5′UTR (“No-uORFs” reporter) partially removes DDX3-dependency for translation of the main ORF (Fig. 8D, E). Interestingly, replacing the non-AUG start codons with AUG (“AUG-uORFs” reporter) almost completely removes the dependency on DDX3 for translation of the main ORF (Fig. 8D, E). In order to validate a role of DDX3 in regulating translation of the wt HIV-1 5′UTR, we performed a rescue experiment in which cells treated with CRISPR-Cas9 to inactivate DDX3 were transduced with virus-like particles (VLPs) loaded with ready made DDX3 protein 3 hours prior to mRNA transfection (Supplementary Fig. S7A). Because the DDX3 protein is delivered 3 hours only prior to reporter mRNA transfection, this strategy allows to test whether DDX3 has a direct effect on HIV-1 translation as opposed to an indirect effect due to secondary changes induced by the knock-out of DDX3 in cells. As observed (Supplementary Fig. S7B), VLPs are able to restore DDX3 expression in DDX3 knock-out cells to levels near that of wild-type cells. Delivery of DDX3 through VLPs in control cells treated with Cas9 and sgRNAs targeting the GFP coding sequence had no effect on translation driven by the globin 5′UTR or the different HIV-1 5′UTR tested (Supplementary Fig. S7C). This indicates that endogenous DDX3 levels are not limiting with respect to translation of the HIV-1 wild-type 5′UTR. As previously observed, inactivating DDX3 expression specifically inhibits translation driven by the HIV-1 wild-type 5′UTR but not that of the mutated 5′UTRs (Supplementary Fig. S7C). Restoring DDX3 expression to near endogenous levels in these cells, allowed a partial recovery of translation driven by the HIV-1 wild-type 5′UTR (but had no effect on translation driven by the mutated versions of the HIV-1 5′UTR or the globin control) (Supplementary Fig. S7C). Overall these results strongly suggest that translation driven by the HIV-1 5′UTR is regulated by uORFs, in concert with DDX3.

Finally, requirement of DDX3 for the translation of HIV-1 viral transcripts was shown to depend on the trans-activator response region (TAR) located immediately downstream the cap-structure[29]. As expected, removing TAR from our wild-type reporter RNA led to a loss of DDX3 dependency for its translation (Fig. 8F). This was also the case when TAR was removed from the “No uORFs” reporter RNA therefore suggesting that TAR is required for DDX3 to modulate translation of the main ORF in a uORF-dependent manner.

Peptides derived from 5′ LTR uORFs elicit HIV-specific T cell responses in people living with HIV

To further demonstrate the existence of translation events from non-AUG start codons in the 5′UTR of HIV transcripts and test their potential role in the context of people living with HIV, we made used of the capacity of T cells to recognize peptide-derived from viral proteins and/or polypeptides presented by MHC molecules. HIV-specific T cells can indeed recognize peptide encoded by classical ORF but also by alternative reading frames (ARF) from HIV-1 genome[72–76]. To this end, we selected 5 peptides derived from 3 different uORFs (See Fig. 8A and Supplementary table 1 for peptide/uORF correspondence) based on different parameters including ribosomal coverage, conservation among HIV isolates and more importantly on the prediction to bind a high number of HLA alleles with a high affinity. The peptides were used to screen for T cell responses in PBMCs of people living with HIV (see Table 1 for the list and clinical characteristics of Human participants). PBMCs from treated (ART) and untreated but aviremic (so called elite controllers, EC) individuals were stimulated with a pool of peptide containing the 5 uORF-derived peptides (POOL) and cultured in the presence of cytokines (IL2 and IL7) in order to expand peptide-specific memory T cells. On day 7 (not shown) and day 12 (Fig. 9A–D and Supplementary Figs. S8A–D), T cell responses against the POOL were assessed using IFNγ-Elispot. In addition, on day 12, except for donor EC-3, T cell responses to individual peptides of the pool (A02;

https://doi.org/10.1038/s41467-025-56772-3

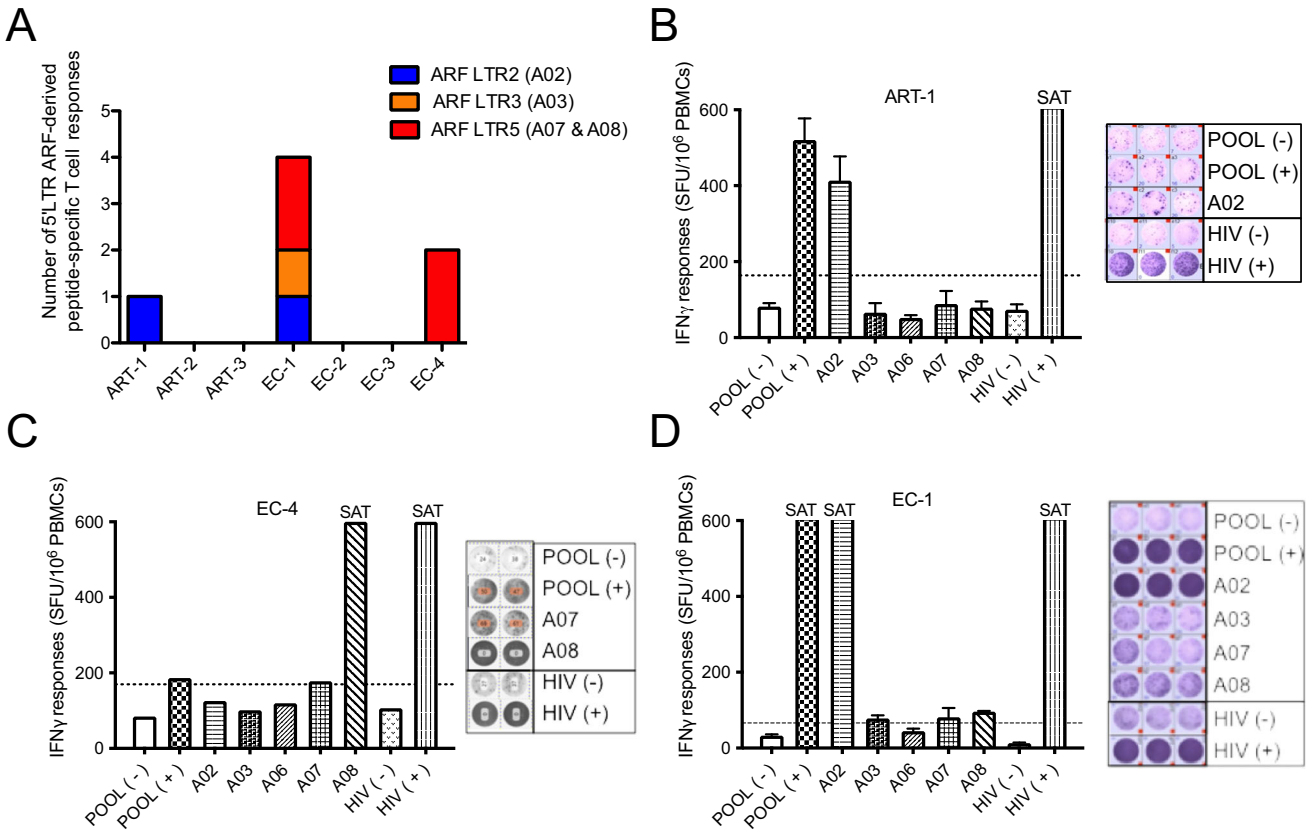


Fig. 9 | Identification of polypeptides that initiate specific T cell responses in treated and untreated HIV infected individuals. (**A to D**), Peptides potentially encoded by uORFs (see peptides "A02; A03; A06; A07; A08" in panel (A) of Fig. 8) were synthesized and used to screen for T cell responses in PBMCs of HIV-infected individuals. PBMCs from treated (ART) and untreated but aviremic (so called elite controllers, EC) individuals were stimulated with a pool of uORF-derived peptides (POOL) and cultured in the presence of IL-2 and IL-7 cytokines in order to expand peptide-specific memory T cells. On day 7 (not shown) and day 12, T cell responses against the POOL were assessed using IFNγ-Elispot. In addition, on day 12, except for donor EC-3, individual peptides of the pool (A02; A03; A06; A07; A08) were assessed using IFNγ-Elispot. As positive control for T cell expansion and activation, a pool of immunogenic peptides from HIV Env and Gag proteins was also used. **A** Number of uORF-derived peptides recognized by each HIV-infected individuals. The color code indicates from which uORF the peptides are derived (A02= Blue; A03= Orange; A07&A08= Red). **B, C** and **D** left panels, detailed IFNγ-Elispot data from the 3 individuals presenting T cell responses, expressed as spot forming units (SFU) per million PBMCs; right panels pictures of the raw data from the Elispot plates in technical triplicate ($n = 3$ technical replicates) (**B and D**), or duplicate ($n = 2$ technical replicates) (**C**), POOL (−) and POOL (+): PBMCs expanded with the pool of uORF-derived peptides but restimulated on the day of the Elispot with medium or the POOL, respectively. A02; A03; A06; A07 and A08 name of individual uORF-derived peptides used for re-stimulation. HIV (-) and HIV (+): PBMC expanded with the pool of Env- and Gag-derived peptides but re-stimulated for the Elispot assay with medium or the same pool of HIV peptides, respectively. Responses were considered positive when IFNγ production was superior to 20 spots/106 PBMCs and at least twofold higher than background production from cells re-stimulated with medium (dotted lines). SAT: saturated signal, where counts cannot be estimated due to overwhelm IFNγ secretion by activated T cells. Source data are provided as a Source Data file. Error bars correspond to the standard deviation of the mean.

A03; A06; A07; A08) were also evaluated (Fig. 9A-D and Supplementary Fig. S8A-D). As positive controls for T cell expansion and activation, PBMCs were also expanded with a pool of HIV Env- and Gag-derived peptides (HIV (+)) or a commercial pool of common virus-derived peptides (CEF (+)). Out of 7 donors tested, 6 reacted to HIV protein-derived peptides (HIV (+)). Note that the ART-3 donor, who did not show a significant T cell response to HIV protein-derived peptides, did however react to the CEF pool (Fig. 9A-D and/or Supplementary Fig. S8A-D). Remarkably, 3 individuals showed uORF-derived peptide-specific T cells responses, with the individual EC-1 reacting to peptides derived from the 3 uORF tested (Fig. 9A). Depending on the peptide, the magnitude of uORF-derived peptide-specific T cells responses (IFNγ+ spots/10^6 PBMCs) varied and could reach plateau responses equivalent to responses induced by the pool of HIV protein-derived peptides (see peptide A08 and A02, Fig. 9C and D, respectively). These results show that uORF-derived peptides elicit potent T cell responses in people living with HIV. This further implies that uORFs encode polypeptides in the course of natural infection in vivo.

Discussion

HIV-1 is a complex retrovirus capable of expressing a multitude of viral proteins from a single transcriptional unit, through a combination of alternative splicing as well as translational and post-translational mechanisms. Expression of viral proteins is tightly coordinated across time during the replication cycle[2,3,77] and it has been proposed that HIV-1 infection is associated with a translational shutoff of cellular transcripts that could favor translation of viral transcripts[17–21,40].

Our results show that HIV-1 infection induces a strong perturbation of the cellular cytoplasmic transcriptome and proteome (Figs. 1 and 3), a significant change in the translational status of a restricted group of cellular transcripts bearing a strong GC bias in their coding sequence (Supplementary Fig. S1C). However, we did not detect any significant changes in the global translation rate of the host cell, at least during the first 72 h of infection (Fig. 2). This contrasts with other RNA viruses, such as picornaviruses or alphaviruses, that induce a rapid and strong translational shut-off of cellular transcripts, which strongly favors translation of viral transcripts. A lack of cellular translational shut-off is not surprising since HIV-1 directly relies on the host

transcription machinery for synthesis of viral mRNAs which are both capped and polyadenylated and are mainly translated in a cap-dependent manner[5,8,78,79]. Furthermore, HIV-1 chronically infected cells lines have also been described in the literature that are viable in spite of strong and active viral replication[80,81].

Cellular transcripts which expression is down-regulated upon HIV-1 infection are mainly related to the regulation of gene expression and include factors involved in ribosome biogenesis, translation, tRNA biogenesis and aminoacylation, mRNA splicing and processing, which could result from a cellular response to the stress induced by infection (Fig. 1C). In addition to the above mentioned factors, our data revealed that transcripts coding for several host proteins involved in promoting HIV-1 expression are down-regulated upon infection, these include the DEAD-Box protein DDX3[30,51,82], FTSJ3[83], RNA helicase A[22,26] among others (See Supplementary Data 1). Cellular transcripts which expression is up-regulated upon infection are mostly associated with the plasma membrane and related to the immune response (including interferon signaling, cell migration and adhesion molecules), cellular stress response pathways and cholesterol metabolism. Interestingly, cholesterol is an important factor to recruit Gag at the plasma membrane[84–86] and cholesterol depletion has been shown to reduce virion release from infected cells[87]. Thus, over-expression of factors involved in cholesterol metabolism upon infection could favor Gag assembly and particle release at late time points of the replication cycle. Finally, measuring protein levels in infected cells revealed that although most changes in protein expression are concordant with changes in mRNA abundance and ribosome loading, specific functional groups of proteins appear to be strongly regulated at the post-translational level either to buffer or completely reverse changes observed in transcript abundance (Fig. 3). This finding highlights the complex and multi-layer regulatory landscape of gene expression that takes place during viral infection.

Cytoplasmic abundance and translation of viral transcripts as measured by RNA-seq and ribosome profiling at different times post-infection indicates a relatively sequential expression of fully spliced, partially spliced and unspliced viral transcripts in infected cells (Fig. 4). Nevertheless, unspliced mRNA expression and translation is still detected at early replication times. This could be partially explained by the fact that experiments were performed in a bulk population of cells with unsynchronized infection events. However, early expression of Gag/Pol could be important for the replication cycle in addition to their canonical roles at late replication steps. For instance, Integrase is known to interact with TAR RNA and Tat, and has been recently shown to regulate proviral DNA transcription at early time points of infection[88–91]. Surprisingly, we were able to detect translation of incoming viral transcripts as soon as 1 h post-infection (Fig. 5A, C and D). Translation of incoming viral genomic RNAs could correspond to non-productive infection events in which the genomic RNA fails to be reverse-transcribed and is able to release from the capsid core and recruit 40S ribosomes to engage into translation. Similar results have recently been described by an independent group[92] and could represent an important source of viral epitopes presented by infected cells during the early steps of infection as further discussed below.

Monitoring translation on viral transcripts revealed the presence of multiple uORFs in the 5′UTR (Fig. 7A). Raw ribosome profiling signal originating from these non-AUG uORFs is as important as that from canonical viral ORFs (Fig. 7A), especially in primary CD4 + T cells. This is however due to the fact that the section of the 5′UTR, which contains the uORFs is located upstream of the first splice-donor site and is therefore common to all viral transcripts. When probing translation from a single mRNA species (Fig. 8B), ribosome profiling signal from uORFs was lower than that from the main ORF. Lentiviruses have been shown to avoid the presence of upstream AUG codons in their 5′UTRs[93]. Furthermore, addition of synthetic upstream AUG codons is deleterious to lentiviral replication[78,94]. Nonetheless, ribosome profiling signals from uORFs could be readily detected under high-salt conditions, ribosome immunoprecipitation or through the Ribolace protocol indicating productive translation from uORFs (Fig. 8A). Also mutating all non-AUG start codons into UAA stop-codons led to a complete loss of ribosome profiling signal (Fig. 8B), indicating that translation from uORFs is indeed productive. Recognition of these upstream non-AUG start codons by scanning 40S are therefore not as detrimental as the presence of an upstream AUG start codon which, when being in an optimal Kozak context, impede most 40S ribosomes to reach the canonical translation initiation start site. Nevertheless, to further demonstrate that uORFs do encode polypeptides, we investigated the presence of T lymphocyte responses that are specific to uORF-derived peptides in people living with HIV-1. Note that, using the same protocol, no uORF-derived peptide-specific T cell responses could be detected in the PBMCs of HIV-negative individuals. In addition, in HIV-1 positive individuals tested, uORF-derived peptide-specific T cells responses were able to reach the magnitude induced by a pool of peptides derived from canonical HIV protein epitopes. Overall, these results strongly argue that uORFs encode polypeptides in the course of a natural infection in vivo. Furthermore, since uORF translation from incoming genomic RNA can be detected as soon as 1 h post-infection, we speculate that the short peptides derived from uORFs could represent an important source of viral antigens during the early steps of the disease.

Remarkably, removal of uORF start codons lead to a mild but significant increase in translation from the main ORF, suggesting that they exert a negative effect on viral protein expression. Nevertheless, this negative effect is partially suppressed by the DEAD-box protein DDX3 (Fig. 8D). Interestingly, translation from uORFs is conserved in other retroviruses, such as HIV-2, HTLV and MLV (Supplementary Figs. S6C, S6D and 69). This raises a question regarding the potential roles of these uORFs. Is it a consequence of the structural and sequence constraints imposed by the multiple functional elements located in the 5′UTR (TAR, poly(A), PBS, dimerization and encapsidation signals)? Do they play a *cis*-regulatory role by restraining the number of ribosomes reaching the main ORF? Cellular mRNAs with uORFs are known to be efficiently translated under stress conditions, similar to those induced by infection[95,96]. Dynamic regulation of uORF translation through cellular factors could also modulate translation from canonical ORFs in a dynamic fashion during the replication cycle. HIV-1 has been described to contain IRES sequences both in its 5′UTR and in the coding region of Gag[9,11,13,14,16,39,59]. These IRESes could bypass uORFs to maintain efficient expression of viral proteins. However, our results indicate that translation from the IRES located in the Gag coding region is inefficient in the context of infection and that addition of AUG codons in the 5′UTR strongly represses translation from the main ORF, as previously described[78], arguing for a limited contribution of IRES translation compared to cap-dependent translation, at least in the context of reporter transcripts.

Peptides originating from uORFs could have a biological role as described for other small peptides produced from cellular uORFs[97]. They might also be presented by MHC class I molecules and used by the immune system to mount potent antiviral T cell responses[98,99]. We highlight here that HIV-1 uORFs can regulate the expression of canonical HIV proteins and that this can be further modulated by host factors such as DDX3. In addition, we show that uORFs induce T cell responses in treated and untreated people living with HIV. Remarkably, from one individual to the other, but also within the same individual, the magnitude of T cell responses targeting uORF-derived peptide varies depending on the peptide tested. Several parameters might explain these variations such as the difference in peptide sequences, which determine the capacity to bind to restricted HLA molecules, but also the frequency of uORF-derived peptide-specific T cells in the blood of people living with HIV. In vivo, the initial viral load and the efficiency of uORF translation might indeed influence the

quantity of peptide-MHC complexes exposed at the surface of infected cells and thus the likelihood of activating uORF-derived peptide-specific T cells[100]. In the future it will be important to characterize the contribution of HIV uORF-specific T cell responses in the control of viral replication and disease progression and to test whether the initiation of uORF-specific T cell might be counteracted by cellular factors such as DDX3 or cellular stress conditions that are known to regulate uORF translation.

Methods

Plasmids

LentiCRISPRv2 was a gift from Feng Zhang (Addgene plasmid #52961[101]). pMD2.G was a gift from Didier Trono (Addgene plasmid # 12259; http://n2t.net/addgene:12259; RRID:Addgene_12259). pBlades plasmids coding for single-guide RNAs targeting the genes coding for eL8 (RPL7a), DDX3 and GFP and were constructed as described in refs. 102,103. The plasmid coding for the murine leukemia virus Gag:DDX3 fusion protein was cloned as described in ref. 102.

Primers and DNA sequences

Sequence of the sgRNA targeting eL8 used to introduce the Flag-tag is 5′ GTACAGCTCACTCACCATCTT 3′. Sequence of the single-stranded DNA oligo used for insertion of a Flag-tag sequence in the eL8 locus by homology directed repair is 5′TTACCCACAATTCCCTTTCCTTTCTC TCTCCTCCCGCCGCCCAAGATGGACTACAAAGACGATGACGACAAGG TGAGTGAGCTGTAGTTCCGTGGCACTATAGCCAGGTTCCGGCTGTAT 3′. The sequence of the sgRNA targeting the GFP coding sequence is 5′CGAGGAGCTGTTCACCGGGG3′. the sequence of the sgRNA targeting DDX3 is 5′TGGCAGTGGAAAATGCGCTC3′.

Cell lines and cell culture

HIV-1 latently infected U937 Cells (U1) were obtained from the AIDS reagent program (Reagent number 165) and cultured in RPMI medium supplemented with 10% fetal calf serum (FCS). Induction of productive HIV-1 replication in latently infected U937 cells was obtained by complementing the medium with 100 ng/mL of PMA and 500 ng/mL of Ionomycin. SupT1 cells were obtained from ATCC and cultured in RPMI medium supplemented with 10% fetal calf serum (FCS).

HEK293T Flag-eL8 (Rpl7a) cells cultured in DMEM medium supplemented with 10% fetal cal serum (FCS). They were produced by transfection of lentiCRISPRv1 together with pBLADES-eL8, that expresses a sgRNA targeting the eL8 locus near the translation start codon, and a single-stranded DNA donor template for homology-directed repair containing homology arms to eL8 and a Flag-tag downstream the AUG start-codon. Upon transfection, a clonal population expressing Flag-eL8 was isolated by limit dilution and screening of isolated clones.

Primary cells

Blood from healthy donors was obtained from the Etablissement Français du Sang, under agreement n°21PLER2019-0106. Peripheral blood mononuclear cells (PBMCs) were isolated by centrifugation through a Ficoll® Paque Plus cushion (Sigma-Aldrich). Primary human T cells were purified by positive selection using CD3 MicroBeads (Miltenyi Biotec), as previously described[104]. Primary T cells were cultured in RPMI supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin, and stimulated for 48 h with 10 µg/ml phytohemagglutinin (PHA) (Fisher Scientific) and 50 U/mL interleukin-2 (IL-2, Miltenyi Biotec) prior to infection.

T cell lines derived from humanized mice infected with HTLV were obtained as described in ref. 105. Briefly, single cell suspensions of cells collected from infected humanized mice were cultured in complete RPMI 1640 medium supplemented with 10% FCS (Sigma-Aldrich, France). Recombinant IL2 (20 U/ml) (Peprotech, France) was added to the cultures every 3 days. After one month of culture, selected cell lines were tested for the relevant proviral sequence and weekly monitored for their proliferation.

Viral production and infection

Wild-type HIV-1 NL4-3 was produced by standard polyethylenimine transfection of HEK293T. The culture medium was changed 6 h later, and virus-containing supernatants were harvested 42 h later. Viral particles were filtered, purified by ultracentrifugation through a sucrose cushion (25% weight/volume in Tris-NaCl-EDTA buffer) for 75 min at 4 °C and 28,000 rpm (100,000 × g at r_{AV} of 109.7 mm) using a SW 32 TI rotor (Beckman Colter), resuspended in serum-free DMEM medium and stored in small aliquots at −80 °C. Viral particles were titrated using an HIV-1 p24Gag Alpha-Lisa kit and an Envision plate reader (Perkin Elmer). VSV-G pseudotyped HIV-1 (NL4.3) and non replicating VSV-G pseudotyped HIV-1-GFP virions were produced by co-transfecting the wild-type HIV-1 NL4-3 or HIV-1 NL4-3 Δ-Env GFP plasmids, with the pMD2.G plasmid coding for the VSV-G envelope glycoprotein.

Infection of SupT1 cells was carried out as follows. Cells were concentrated at 10 million cells per mL and incubated for 1 h at 37 °C in the presence of VSV-G pseudotyped HIV-1 virions (NL4.3 strain) at MOI 5. Following this incubation, cells were diluted in RPMI medium (supplemented with 10 % FCS) to a final concentration of 1 million cells per mL and collected at the indicated time points. Cells were then analyzed with the MACSQuant VYB cytometer.

Infection of primary T cells was carried out as follows. Activated cells (see above in the "**primary cells**" section) were plated at 250,000 cell per well in a 96-well plate and infected with 10 or 100 ng p24Gag for 4 h prior to media replacement. At 48 h and 72 h post-infection, supernatants were harvested for viral production quantification by Alpha-Lisa.

Production of murine leukemia virus-like particles loaded with DDX3 was performed as described in ref. 103.

Flow cytometry analysis of p24 expression in infected cells

Cells infected with HIV-1 virions (NL4.3 strain) were fixed in PBS-paraformaldehyde 4% for 20 min at room temperature and permeabilized with PBS-0.5% Triton X-100 for 15 min. Cells were then incubated with the fluorescently labeled anti-p24 antibody (Clone: KC57, Beckman Colter, 6604667; 1/200 dilution).

Measurement of protein synthesis using O-propargyl-puromycin (OPP)

After infection with VSVg-pseudotyped HIV-1-GFP or wild-type HIV1 virions (NL4.3 strain), cells were treated with 10 µM OPP (Immagina Biotechnology, OP001-26) for 30 min at 37 °C. Then, cells were fixed in PBS-paraformaldehyde 4% for 20 min at room temperature and permeabilized with PBS-0.5% Triton X-100 for 15 min. For cells infected with HIV1 virus NL4.3 strain, a fluorescent labeling was first made with anti-p24 antibody (Clone: KC57, Beckman Colter, 6604667; 1/200 dilution). Then, fluorescent labeling of the OPP was made with the Click-IT™ Plus Alexa Fluor™ 488 Picolyl Azide (Thermo Fisher Scientific, C10641), according to the manufacturer's instructions.

For cells infected with VSVg-pseudotyped HIV-1-GFP, fluorescent labeling of the OPP was made with the Click-IT™ Plus Alexa Fluor™ 555 Picolyl Azide (Thermo Fisher Scientific, C10642), according to the manufacturer's instructions. Finally, the cells were analyzed with the MACSQuant VYB cytometer.

Measurement of incoming viral genomic RNA translation

SupT1 cells (1×10^6 cells), pre-incubated or not with cycloheximide (100 µg/mL final) or Puromycin (200 µg/mL final) were infected at MOI 5 with a genetically modified replication-competent HIV-1 virus bearing the GFP coding sequence embedded withing the Gag coding sequence (previously described in ref. 55). Four hours post-infection, the cell

medium is replace with fresh medium containing cycloheximide (100 μg/mL final) or Puromycin (200 μg/mL final) for 3 additional hours. Cells were then analyzed with the MACSQuant VYB cytometer to measure GFP fluorescence levels.

mRNA transfection, transduction with Virus-like particles and western-blotting

For RNA transfection experiments in cells depleted of DDX3, 3 million HEK293T cells were plated in a 10 cm petri-dish. 24 hours after plating, cells were transfected with 1.7 μg of the LentiCRISPRv2 plasmid and 4.4 μg of the pBLADE plasmid (coding either for the sgRNA targeting the GFP 5′CGAGGAGCTGTTCACCGGGG3′ or DDX3 5′TGGCAGTGGAAAATGCGCTC3′) using lipofectamine 3000. 48 h after transfection, cells were collected and plated at 300,000 cells per well in a 6-well plate. 24 hours later, cells were transduced with DDX3-loaded virus like particles or with control empty virus-like particles. 3 hours after transduction, cells were transfected with 1 μg of renilla luciferase coding mRNAs and 500 ng of a firefly coding mRNA (internal control for normalization). 3 hours after transfection, cells were collected, lysed and renilla and firefly luciferase activity was measured using the Dual-Glo luciferase assay system (Promega). To monitor DDX3 levels in cells transfected with the Cas9 and sgRNA expressing plasmids, cellular lysates used for luciferase assays were also loaded on a 4–12% acrylamide gel, proteins transferred to a PVDF membrane and DDX3 protein levels detected using an anti-DDX3 rabbit antibody (Cell signaling (D19B4), 8192 T; 1/1000 dilution in 5% milk). Beta-actin protein levels were detected using an anti-Beta-actin mouse antibody (Sigma Aldrich, Catalog number: A1978-100UL; 1/1000 dilution in 5% milk).

Mass Spectrometry analysis

SupT1 cells were infected with VSV-G pseudotyped HIV-1 (NL4.3) virions at MOI 5 as described above in the "Viral production and infection" section. At 12 h, 24 h and 36 h post Mock and HIV infection, cells were washed twice in cold PBS, then, lysed in 50 mM Tris-HCl, 150 mM NaCl, 1% NP40, 0,5 % Na-deoxycholate, 0,1% SDS and Phos-stopTM phosphatase inhibitors (Roche) during 15 minutes at 4 °C. Finally, the lysate was centrifuged and the supernatant was collected. For in-solution enzymatic digestion, 50 μg of protein samples were mixed with Lysis buffer provided in the easyPep mini Kit (Thermo Scientific). The samples were reduced and alkylated by incubation at 95 °C for 10 min and continuously shaken at 1000 rpm. Then samples were digested for 3 h at 37 °C and processed following the manufacturer's protocol.

For nanoLC-MS/MS analysis, samples were analyzed using an Ultimate 3000 nano-RSLC (Thermo Scientific, San Jose California) coupled on line with a Q Exactive HF mass spectrometer via a nano-electrospray ionization source (Thermo Scientific, San Jose California).

400 ng of each peptide mixtures were loaded on a PepMap NEO C18 trap-column (300 μm ID × 5 mm, 5 μm, Thermo Fisher Scientific) for 3.0 minutes at 20 μL/min with 2% ACN, 0.05% TFA in H2O and then separated on a C18 Acclaim Pepmap100 nano-column, 50 cm × 75 μm i.d, 2 μm, 100 Å (Thermo Scientific) with a 100 minutes linear gradient from 3.2% to 20% buffer B (A: 0.1% FA in H2O, B: 0.1% FA in ACN), from 20% to 32% of B in 20 min and then from 32% to 90% of B in 2 min, hold for 10 min and returned to the initial conditions in 2 min for 13 min. The total duration was set to 150 minutes at a flow rate of 300 nL/min. The oven temperature was kept constant at 40 °C.

Samples were analyzed with a TOP15 HCD method: MS data were acquired in a data dependent strategy selecting the fragmentation events based on the 15 most abundant precursor ions in the survey scan (350-1650 Th). The resolution of the survey scan was 120,000 at m/z 200 Th. The Ion Target Value for the survey scans in the Orbitrap and the MS2 mode were set to 3E6 and 1E5 respectively and the maximum injection time was set to 60 ms for both scan modes. HCD MS/MS spectra acquisition parameters were as follows: collision energy = 27; isolation width of 1.4 m/z; precursors with unknown charge state or a charge state of 1 were excluded. Peptides selected for MS/MS acquisition were then placed in an exclusion list for 20 s using the dynamic exclusion mode to limit duplicate spectra. The spray voltage was set in positive ion at 1800 V and ion transfer tube maintained at a temperature of 250 °C.

For data analysis, proteins were identified by database searching using Sequest HT with Proteome Discoverer 2.5 software (Thermo Scientific) against homo sapiens swissprot database (2023-06 release, 20348 sequences), HIV-1 NL4-3 protein sequences (9 sequences) and a contaminant database. Precursor and fragment mass tolerance were set at 10 ppm and 0.02 Da respectively, and up to 2 missed cleavages were allowed. Oxidation (M), Acetylation (Protein N-terminus), were set as variable modification, and Carbamidomethylation (C) as fixed modification. Full Trypsin was selected as digestion enzyme parameter. Peptides and proteins validation were performed by Percolator and a false discovery rate of 1% for peptides and proteins was set. Protein quantification was done by Label Free Quantification (LFQ) approach, and LFQ abundance values were obtained and normalized to the total peptide amount. Protein quantitation was performed with precursor ions quantifier node in Proteome Discoverer 2.5 software with protein quantitation based on pairwise ratios and hypothesis t-test. Proteins were considered as differentially expressed between the two conditions when FC > 2 or FC < 0.5 and pv < 0.05.

Ribosome profiling

Ribosome profiling samples were prepared as described in ref. 106 with the following modifications. At each time point of infection, 10 mL of the cell culture suspension were collected and immediately mixed with 40 mL of ice-cold 1X PBS supplemented with 120 μg/mL of cycloheximide. Cells were then pelleted at 500 g for 5 min at 4 °C, the supernatant was discarded and the cell pellet resuspended in 10 mL of ice-cold 1X PBS + cycloheximide (100 μg/mL). Cells were pelleted again at 500 g for 5 min, the supernatant was discarded and the cell pellet lysed in 1 ml of lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 μg/ml cycloheximide and 1× Protease-Inhibitor Cocktail EDTA-free (Roche)). Lysate was homogenized with a P1000 pipettor by gentle pipetting up and down for a total of eight strokes and incubated at 4 °C for 10 min. The lysate was centrifuged at 1300 g for 10 min at 4 °C, the supernatant recovered and the absorbance at 260 nm measured. For ribosome footprinting, 5 A260 units of the cleared cell lysates were incubated either for 30 min at RT either with 3 μg of micrococcal nuclease (Sigma), or 300 units of RNase T1 (Fermentas) and 500 ng of RNase A (Ambion) or for 1 h at 4 °C with 350 units of RNase I (Ambion). After nuclease treatment, samples were loaded on top of a 10–50% (w/v) linear sucrose gradient (20 mM HEPES-KOH, pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 2 mM DTT and 100 μg/mL of cycloheximide) and centrifuged in a SW-40Ti rotor at 35,000 rpm (217,290 × g) for 2 h 40 min at 4 °C. The collected 80S fraction was complemented with SDS to 1% final and proteinase K (200 μg/mL) and then incubated at 42 °C for 45 min. After proteinase K treatment, RNA was extracted with one volume of phenol (pH 4.5)/chloroform/isoamyl alcohol (25:24:1). The recovered aqueous phase was supplemented with 20 μg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM $MgCl_2$. RNA was precipitated with three volumes of 100% ethanol at −20 °C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 μl of water and the 3′ ends dephosphorylated with PNK (New England BioLabs) in MES buffer (100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 300 mM NaCl) at 37 °C for 3 h. Dephosphorylated RNA footprints were then resolved on a 15% acrylamide (19:1), 8 M urea denaturing gel for 1 h 30 min at 35 W and fragments ranging from 26 nt to 32 nt size-selected from the gel. Size-selected RNAs were extracted from the gel slice by overnight nutation at RT

in RNA elution buffer (300 mM NaCl, and 10 mM EDTA). The recovered aqueous phase was supplemented with 20 µg of glycogen, 300 mM sodium acetate, pH 5.2, and 10 mM $MgCl_2$. RNA was precipitated with three volumes of 100% ethanol at −20 °C overnight. After a wash with 70% ethanol, RNA was resuspended in 5 µl of water and subjected to cDNA library construction as described below.

For ribosome profiling samples exposed to high-salt in order to induce dissociation of non-elongating 80S ribosomes, the protocol was similar to that described above with the following modification. 1 M KCl (final concentration) was added to 5 A260 units of the cleared lysates and incubated for 30 min at 4 °C. After this, samples were passed through Zeba Spin desalting columns 7 K MWCO (Thermo Scientific) pre-equilibrated with cell lysis buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl2, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100 µg/mL cycloheximide and 1× Protease-Inhibitor Cocktail EDTA-free (Roche)). Then, samples were treated with nucleases as described above to obtain ribosome footprints.

Ribolace ribosome footprinting experiments were performed following the manufacturer's protocol (Immagina BioTechnology) with the following change, RNase I was replaced with 7 µg of micrococcal nuclease (Sigma) as described above.

Ribosome immunoprecipitation-based profiling was performed as follows. Two 10 cm plates of HEK293T cells expressing FLAG-eL8 (RPL7a) from the endogenous locus were transfected with the pNL4.3-HIV-1 plasmid. 24 h after transfection, medium was removed, cells were washed with ice-cold 1X PBS supplemented with cycloheximide (final concentration of 100 µg/mL) and scraped from the plates. Cells were pelleted at 500 g for 5 min and lysed in 1 mL of ice-cold lysis buffer (25 mM Tris pH 7.4, 150 mM NaCl, 15 mM $MgCl_2$, 1 mM DTT, 1X cOmplete EDTA-free protease inhibitor (Roche), 100 µg/mL CHX and 1% Triton X-100) for 10 min on ice. Cell debris and nuclei were removed by centrifugation at 1,300 g for 10 min at 4 °C. Cytoplasmic lysate 254 nm UV absorption was measured and adjusted to 25 A.U./mL and lysate was treated with 60U RNase T1 and 10 ng RNase A per A.U. for 30 min at 25 °C to generate ribosome footprints. Lysate was then diluted-up to 10 mL with ice-cold lysis buffer. 150 µL of anti-FLAG agarose beads suspension (Sigma-Aldrich, A2220) were washed 3 times with 10 mL lysis buffer and added to the diluted sample. Ribosome binding was allowed to proceed for 1 h at 4 °C on rotating wheel. Beads were then washed for a total of 5 times in 15 mL tubes using 10 mL of lysis buffer for 5 min at 4 °C, and tubes were changed after the 1st and 4th wash to minimize contaminants recovery. Beads were washed one final time in a 2 mL low-protein binding tube with 1 mL and ribosomes were eluted using 150 µL of lysis buffer supplemented with 500 µg/mL FLAG peptide (DYKDDDDK peptide - GenScript) for 1 h at 4 °C. Eluates were finally digested with proteinase K (200 µg/mL final) for 45 min at 42 °C and RNAs were purified using acid Phenol:Chloroform purification. cDNA libraries were generated as described below.

Ribosomal RNA depletion

For RNA-seq samples, cytoplasmic RNAs were depleted from ribosomal RNAs using antisense DNA oligos complementary to rRNA and RNaseH as previously described[107]. Briefly, 50 ng of total RNA was incubated with 15pmol of a mixture of 50 bp ssDNA depletion oligos complementary to rRNA (in 35 µL of a 20 mM Tris-HCl pH = 7.4, 40 mM NaCl solution). Samples were incubated at 95 °C for 2 minutes and then cooled down to 22 °C with a 0.1 °C/sec ramp. Samples are then complemented with 4 µL of 10X RNaseH buffer (500 mM Tris-HCl ph=7.4; 1 M NaCl; 200 mM MgCl2) and 1 µL of RNaseH (Thermo Scientific EN0201) and incubated for 1 h at 45 °C. Following DNase treatment to remove DNA oligos, rRNA-depleted RNAs were fragmented using RNA fragmentation reagent (Ambion, Cat: AM8740) for 3 min at 70 °C followed by inactivation with the provided "Stop" buffer.

cDNA library preparation for Ribosome profiling and RNA-sequencing

High-throughput sequencing libraries were prepared as described in ref. 106. Briefly, RNAs (either from ribosome footprints or fragmented RNAs from total cytoplasmic RNAs) were dephosphorylated at their 3′ end using PNK (New England Biolabs, Cat: M0201) in the following buffer: 100 mM Tris-HCl pH 6.5, 100 mM magnesium acetate, 5 mM β-mercaptoethanol and incubate at 37 °C during 3 h. RNA fragments with a 3′-OH were ligated to a preadenylated DNA adapter. Then, ligated RNAs were reverse transcribed with Superscript III (Invitrogen) with a barcoded reverse-transcription primer that anneals to the pre-adenylated adapter. After reverse transcription, cDNAs were resolved in a denaturing gel (10% acrylamide and 8 M urea) for 1 h and 45 min at 35 W. Gel-purified cDNAs were then circularized with CircLigase II (Lucigen, Cat: CL4115K) and PCR-amplified with Illumina's paired-end primers 1.0 and 2.0. PCR amplicons (12-14 cycles for RNA-seq and 4-6 cycles for ribosome profiling) were gel-purified and submitted for sequencing on the Illumina HiSeq 2000 platform.

Bioinformatics analysis

First, fastq files were demultiplexed using Flexi-splitter software. Next, adapter sequences were trimmed from RNAseq and Riboseq reads and bad quality reads were filtered with fastp software (v0.20.1)[108]. Filtered reads were then mapped on rRNA and tRNA sequenced with bowtie2 v2.3.4.1[109] to remove contaminants. Non-aligned reads are then mapped against the human genome with hisat2 v2.1.0[110]. Reads that are not mapped against the human genome are mapped against HIV-1 genome with Hisat2 v2.1.0. Duplicated reads were removed using a home-made script. Deduplicated reads were used to quantify gene expression with htseq-count v0.11.2[111]. Finally, bam files were used to generate normalized bigwig files with deeptools v3.5.1[112].

The periodicity of Ribo-seq experiments was checked with plastid v0.5.1[113] and Ribowaltz[114]. To identify new ORFs from our datasets, we first mapped filtered reads against hybrid human/HIV-1 genome with STAR v2.7.3a[115] to generate bam files with transcript coordinates. These files are then used by RiboCode v1.2.11[116], RiboTish v0.2.5[117] and RibоTricer v1.3.2[118] to identify new ORFs in the HIV-1 genome. FLOSS scores were calculated with scripts from[61].

For gene clustering analyzes (shown in Figs. 3 and Supplementary Fig. S3), the log2(Fold Change) metric of each gene was computed for each time point, omic type, and replicate. Then the mean log2(Fold Change) value was computed using all available log2(FC) replicates values. The mean log2(Fold Change) metric allowed to distingish for each gene, and at each time point, increase or decrease behavior in the HIV1 infected condition compared to the Mock one. Moreover the value dynamic is pretty homogeneous among the omics (−2 to 2 for RNA-seq and Ribo-seq, against -4 to 4 for the Protein abundance). The same process was applied to obtain the mean raw values. But, as their dynamic range are not comparable, an extra normalization step was performed. The raw and mean raw values are transformed using the scikit-learn (1,4,1)[119] QuantileTransformer function. This method converts a set of values to follow a normal distribution using 100 quantiles. As consequences the most frequent values are spread out, and the impact of outliers is reduced. Thus all raw and mean raw values are comparable as they are set between −5 and 5 after quantile transformation.

Finally a set of 2795 genes for which a mean log2(Fold Change) value was available for each omic and at each time point (0, 12, 24 and 36 h) was used for further analyses. These genes were used to perform clustering using the scikit-learn Kmeans function, powered by kmeans + +, which determines the contribution to the overall inertia of each points as an empirical distribution to sample the initial cluster centroids. We set the number of wanted clusters at 12, as it gave 11 homogeneous functional clusters after validation with string-db, and 1

cluster with a single gene (indicating that the maximal number of centroids was already reached). The enrichment of GOs in each cluster was then automatically determined using the R clusterProfiler 4.12.2 library[120]. Subsequently clusters with similar trends and GO enrichments were merged to finally obtain 9 clusters. More details are available here (https://gitbio.ens-lyon.fr/LBMC/RMI2/hiv_project).

Human participants and samples

Elite controllers (EC, $n = 4$) were recruited from the CO21 CODEX cohort implemented by the ANRS (Agence nationale de recherches sur le SIDA et les hépatites virales). PBMCs were cryopreserved at enrollment. ECs were defined as people living with HIV (PLWH) maintaining viral loads (VL) under 400 copies of HIV RNA/mL without cART for more than 5 years. HIV-infected efficiently treated participants (ART) ($n = 3$) were recruited at Kremlin Bicêtre Hospital. They were treated according to standard of care for at least 1 year (mean of 10 years) and had an undetectable viral load using standard assays. PBMCs from anonymous HIV-negative blood donors ($n = 3$), for which we have no baseline information, were purchased from EFS (Établissement Français du sang) under the agreement number 15/EFS/022. Age and sex2 were not considered in the study design.

Ethic statement

All the human participants provided their written informed consent to participate in the study. The CO21 CODEX cohort and this sub-study were funded and sponsored by ANRS and approved by the Ile de France VII Ethics Committee. The study was conducted according to the principles expressed in the Declaration of Helsinki.

12-day in vitro T cell amplification prior Elipsot-assay

PBMCs were thawed according to standard procedure and rested 2-3 h in IMDM (Gibco/ThermoFisher Scientific) containing 5% human AB serum (SAB, Institut Jacques Boy), recombinant human IL-2 (rhIL2, 10 U/ml, Miltenyi) and DNase (1 U/ml, New England Biolabs), washed and enumerated. 5-9 × 10^6 PBMCs were then seeded/well of a 24-well plate in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids and sodium pyruvate (all from Gibco/ThermoFisher Scientific), Nevirapine (NVP, 1,2 μM, AIDS reagent program) to inhibit potential viral replication, and poly I:C (2 μg/ml, InvivoGen) to facilitate the presentation of long peptides by antigen presenting cells[121]. PBMCs were immediately loaded with peptides (Vivitide) all at 10 μg/ml except for the CEF pool (CEF extended peptide pool for human CD8 T cells, Mabtech) used at 5 μg/ml and cultured overnight at 37 °C, 5% CO_2. HIV and/or CEF peptide pools were used as positive control for the expansion of HIV Env- and Gag-specific and common anti-virus-specific T cells, respectively. On day 1 and 3, T cell media were then complemented with rhIL2 (10 U/ml) and rhIL-7 (20 ng/ml, Miltenyi), respectively, and maintained though out the culture. On days 3, 5, 7, and 9, when the cell layer was > 70% confluent, cells were splitted 1:2 before the addition of rhIL-2, rhIL-7 and NVP. On day 7, cells were harvested, counted and a fraction submitted to IFN-γ Elispot assay using the uORF-derived peptide pool and positive controls: HIV and/or CEF peptide pools. The remaining cells were maintained in culture and submitted, on day 12, to IFN-γ Elispot assay using individual uORF peptides and the controls.

Enzyme-linked immunospot (IFN-γ Elispot) assay

From 10-25 × 10^5 cells/well were seeded Elispot plates (MSIPN4550, Millipore) in IMDM supplemented with 10 % SAB, Pen/Strep, nonessential amino acids and sodium pyruvate, loaded with either 50, 10 or 2 μg/ml of uORF-derived peptides, HIV and CEF peptides, respectively and incubated at 37 °C for 16 h. Elispot plates were pre-coated with anti-IFNγ primary antibody and after cell-incubation IFNγ revealed using anti-IFNγ secondary antibody conjugated with biotin (both from Mabtech) as described[122]. The Elispot analysis was performed in technical triplicates or duplicates. Spots were counted with the AID ELISPOT Reader according to standard protocols. Responses were considered positive when IFNγ production was superior to 20 spots/10^6 PBMCs and at least twofold higher than background (cells loaded with the DMSO containing medium used to solubilized the peptides).

hSHAPE analysis

Benzoyl cyanide (BzCN) was used to acylate the 2′-hydroxyl group of the unconstrained nucleotides in the RNA structure, followed by interrogation of each nucleotide using two sets of identical but differentially labeled primers: one set (AS1 and AS2) corresponds to the sequence 5′-TCGCTTTCAAGTCCCTGTTCG-3′ (complementary to HIV-1 nt 189-209) and the second set (AS3 and AS4) correspond to 5′-TTCTTTCCCCCTGGCCTTAAC-3′ (complementary to luciferase sequence nt 395-415). One primer within each set was labeled with either VIC (AS1 & AS3) or NED (AS2 & AS4). The NED-labeled primers from each set were used to prepare a ddG sequencing ladder from the untreated RNA samples. The VIC-labeled primers were used for reverse transcription of the modified or DMSO-treated control RNAs.

Briefly, 1 pmol of in vitro transcribed RNA was denatured at 90 °C for 2 min and then cooled on ice for 2 min, followed by the addition of excess yeast tRNA (2 μg) and RNasin (5U) in 10 μl HEPES Buffer (30 mM HEPES pH 8, 300 mM KCl, 5 mM $MgCl_2$). The RNA was then folded at 37 °C for 20 min and modified by 3 μl of 300 mM BzCN in DMSO one min at room temperature. After adding 82 μl of water, the chemically modified RNA was extracted (Roti-Phenol/Chloroform/Isoamyl alcohol), ethanol precipitated and resuspended in 7 μl of water. Similarly, for the control (unmodified RNA sample), 3 μl of anhydrous DMSO was added instead of BzCN and treated in the same manner.

For elongation of both the modified and control samples, 1 μl of AS1 or AS3 (1 μM) were added to the resuspended RNA and incubated at 90 °C for 2 min, then cooled on ice for 2 min. Two μl of 5X RT buffer was added to each of the samples and incubated at room temperature for 10 min. Reverse transcription was initiated by addition of 10 μl of the elongation mix (2 μl of 5X RT Buffer, 0.6 μl of 25 mM dNTP and 2 U of AMV RT (Life Science)) and incubation at 42 °C for 20 min, 50 °C for 30 min and 60 °C for 10 min. For the ddG sequencing ladder, 2 pmol of untreated RNA and 1 μl of the A2 or AS4 (2 μM) were used and treated as above except for the composition of the sequencing mix (2 μl of 5X RT Buffer, 2 μl of 100 μM ddGTP, 6 μl of G10 (0.25 mM dGTP, 1 mM dATP, 1 mM dCTP, 1 mM dTTP) and 2U of AMV RT (Life Science)). For each experiment, 80 μl of water were added and cDNA was extracted using Roti Aqua-Phenol/Chloroform/Isoamyl alcohol (Carl Roth). The aqueous phase of modified or unmodified samples were pooled with the aqueous phase of the ddG sequencing ladder. The samples were then ethanol precipitated and resuspended in 10 μl of HiDi Formamide (ABI). The samples were then denaturated 5 min at 90 °C, cooled on ice for 5 min, centrifuged, and loaded onto a 96-well plate for sequencing (Applied Biosystems 3130xl genetic analyzer).

The electropherograms obtained were analyzed with QuShape algorithm[123] to extract reactivity data for each sample. The mean reactivity data from three independent experiments were applied as constraints to the RNA sequence in RNAstructure (version 6.1;[124]). The dot bracket file obtained was then used to draw the RNA 2D structure into Structure Editor graphical tool, a module of the RNAstructure software.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data can be retrieved as a BioProject from NCBI using the accession number PRJNA783479. Ribosome profiling reads from

primary CD4 + T lymphocytes infected with HIV-1 virions were obtained from[60] and can be retrieved in the GEO database under GSE158930. The mass spectrometry proteomics data have been deposited to the Center for Computational Mass Spectrometry repository (University of California, San Diego) via the MassIVE tool with the dataset identifier MassIVE MSV000095969 [https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=906c735a58be4bc5845f1a941eb4d07b]. Source data are provided with this paper. The following plasmid, required to generate DDX3 loaded virus-like particles, is available from Addgene: BIC-Gag-DDX3, Plasmid ID: 233687.

Code availability

Code for Flexi-splitter program can be found here: https://gitbio.ens-lyon.fr/LBMC/RMI2/flexi_splitter Processed datasets and code to reproduce the main figures can be found here: http://gitbio.ens-lyon.fr/LBMC/RMI2/translational_landscape_in_HIV_infected_cells and https://gitbio.ens-lyon.fr/LBMC/RMI2/hiv_project.

References

1. Emery, A., Zhou, S., Pollom, E. & Swanstrom, R. Characterizing HIV-1 Splicing by Using Next-Generation Sequencing. *J. Virol.* **91**, e02515–e02516 (2017).
2. Nguyen Quang, N. et al. Dynamic nanopore long-read sequencing analysis of HIV-1 splicing events during the early steps of infection. *Retrovirology* **17**, 25 (2020).
3. Ocwieja, K. E. et al. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. *Nucleic Acids Res* **40**, 10345–10355 (2012).
4. Felber, B. K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T. & Pavlakis, G. N. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. *Proc. Natl Acad. Sci. USA.* **86**, 1495–1499 (1989).
5. de Breyne, S. & Ohlmann, T. Focus on Translation Initiation of the HIV-1 mRNAs. *IJMS* **20**, 101 (2018).
6. Guerrero, S. et al. HIV-1 replication and the cellular eukaryotic translation apparatus. *Viruses* **7**, 199–218 (2015).
7. Jaafar, Z. A. & Kieft, J. S. Viral RNA structure-based strategies to manipulate translation. *Nat. Rev. Microbiol.* **17**, 110 (2019).
8. Ricci, E. P., Soto Rifo, R., Herbreteau, C. H., Decimo, D. & Ohlmann, T. Lentiviral RNAs can use different mechanisms for translation initiation. *Biochem Soc. Trans.* **36**, 690–693 (2008).
9. Brasey, A. et al. The leader of human immunodeficiency virus type 1 genomic RNA harbors an internal ribosome entry segment that is active during the G2/M phase of the cell cycle. *J. Virol.* **77**, 3939–3949 (2003).
10. Daudé, C. et al. HIV-1 sequences isolated from patients promote expression of shorter isoforms of the Gag polyprotein. *Arch. Virol.* **161**, 3495–3507 (2016).
11. Deforges, J. et al. Two ribosome recruitment sites direct multiple translation events within HIV1 Gag open reading frame. *Nucleic Acids Res* **45**, 7382–7400 (2017).
12. Herbreteau, C. H. et al. HIV-2 genomic RNA contains a novel type of IRES located downstream of its initiation codon. *Nat. Struct. Mol. Biol.* **12**, 1001–1007 (2005).
13. Locker, N., Chamond, N. & Sargueil, B. A conserved structure within the HIV gag open reading frame that controls translation initiation directly recruits the 40S subunit and eIF3. *Nucleic Acids Res.* **39**, 2367–2377 (2011).
14. Plank, T.-D. M., Whitehurst, J. T. & Kieft, J. S. Cell type specificity and structural determinants of IRES activity from the 5′ leaders of different HIV-1 transcripts. *Nucleic Acids Res* **41**, 6698–6714 (2013).
15. Ricci, E. P. et al. In vitro expression of the HIV-2 genomic RNA is controlled by three distinct internal ribosome entry segments that are regulated by the HIV protease and the Gag polyprotein. *RNA* **14**, 1443–1455 (2008).
16. Vallejos, M. et al. Functional and structural analysis of the internal ribosome entry site present in the mRNA of natural variants of the HIV-1. *PLoS ONE* **7**, e35031 (2012).
17. Alvarez, E., Menéndez-Arias, L. & Carrasco, L. The eukaryotic translation initiation factor 4GI is cleaved by different retroviral proteases. *J. Virol.* **77**, 12392–12400 (2003).
18. Castelló, A. et al. HIV- 1 protease inhibits Cap- and poly(A)-dependent translation upon eIF4GI and PABP cleavage. *PLoS One* **4**, e7997 (2009).
19. Ohlmann, T. et al. In vitro cleavage of eIF4GI but not eIF4GII by HIV-1 protease and its effects on translation in the rabbit reticulocyte lysate system. *J. Mol. Biol.* **318**, 9–20 (2002).
20. Prévôt, D. et al. Characterization of a novel RNA-binding region of eIF4GI critical for ribosomal scanning. *EMBO J.* **22**, 1909–1921 (2003).
21. Ventoso, I., Blanco, R., Perales, C. & Carrasco, L. HIV-1 protease cleaves eukaryotic initiation factor 4G and inhibits cap-dependent translation. *PNAS* **98**, 12966–12971 (2001).
22. Bolinger, C., Sharma, A., Singh, D., Yu, L. & Boris-Lawrie, K. RNA helicase A modulates translation of HIV-1 and infectivity of progeny virions. *Nucleic Acids Res* **38**, 1686–1696 (2010).
23. Chatel-Chaix, L. et al. Identification of Staufen in the Human Immunodeficiency Virus Type 1 Gag Ribonucleoprotein Complex and a Role in Generating Infectious Viral Particles. *Mol. Cell Biol.* **24**, 2637–2648 (2004).
24. Dugré-Brisson, S. et al. Interaction of Staufen1 with the 5′ end of mRNA facilitates translation of these RNAs. **33**, 4797–4812 (2004).
25. García-de-Gracia, F. et al. CBP80/20-dependent translation initiation factor (CTIF) inhibits HIV-1 Gag synthesis by targeting the function of the viral protein Rev. *RNA Biol.* **18**, 745–758 (2021).
26. Hartman, T. R. et al. RNA helicase A is necessary for translation of selected messenger RNAs. *Nat. Struct. Mol. Biol.* **13**, 509–516 (2006).
27. Ramos, H. *et al*. The double-stranded RNA-binding protein, Staufen1, is an IRES-transacting factor regulating HIV-1 cap-independent translation initiation. *Nucleic Acids Research* gkab1188 https://doi.org/10.1093/nar/gkab1188 (2021).
28. Singh, G. et al. HIV-1 hypermethylated guanosine cap licenses specialized translation unaffected by mTOR. *Proc. Natl Acad. Sci. Usa.* **119**, e2105153118 (2022).
29. Soto-Rifo, R. et al. DEAD-box protein DDX3 associates with eIF4F to promote translation of selected mRNAs: Translation initiation mediated by DDX3. *EMBO J.* **31**, 3745–3756 (2012).
30. Soto-Rifo, R., Rubilar, P. S. & Ohlmann, T. The DEAD-box helicase DDX3 substitutes for the cap-binding protein eIF4E to promote compartmentalized translation initiation of the HIV-1 genomic RNA. *Nucleic Acids Res.* **41**, 6286–6299 (2013).
31. Jacks, T. *et al*. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. 4 (1988).
32. Wei, J. et al. Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance. *Mol. Cell* **73**, 1162–1173 (2019).
33. Yewdell, J. W. DRiPs get molecular. *Current Opinion in Immunology* 7 (2020).
34. Erhard, F., Dölken, L., Schilling, B. & Schlosser, A. Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol. Res* **8**, 1018–1026 (2020).
35. Pierre, P. Immunity and the regulation of protein synthesis: surprising connections. *Current Opinion in Immunology* 8 (2009).
36. Casartelli, N. et al. The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells. *J. Exp. Med.* **207**, 39–49 (2010).

37. Schubert, U., Norbury, C. C., Yewdell, J. W. & Bennink, J. R. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. **404**, 5 (2000).
38. Coulon, P.-G. et al. HIV-infected dendritic cells present endogenous MHC class II–restricted antigens to HIV-specific CD4+ T Cells. *J. Immunol.* **197**, 517–532 (2016).
39. de Breyne, S. et al. In vitro studies reveal that different modes of initiation on HIV-1 mRNA have different levels of requirement for eukaryotic initiation factor 4F: eIF4F requirement for HIV-1 mRNA translation. *FEBS J.* **279**, 3098–3111 (2012).
40. Sharma, A., Yilmaz, A., Marsh, K., Cochrane, A. & Boris-Lawrie, K. Thriving under stress: selective translation of HIV-1 structural protein mRNA during Vpr-mediated impairment of eIF4E translation activity. *PLoS Pathog.* **8**, e1002612 (2012).
41. Cenik, C. et al. Integrative analysis of RNA, translation, and protein levels reveals distinct regulatory variation across humans. *Genome Res* **25**, 1610–1621 (2015).
42. Gerashchenko, M. V. & Gladyshev, V. N. Ribonuclease selection for ribosome profiling. *Nucleic Acids Res* **45**, e6–e6 (2017).
43. Ricci, E. P. et al. Staufen1 senses overall transcript secondary structure to regulate translation. *Nat. Struct. Mol. Biol.* **21**, 26–35 (2014).
44. Li, M. et al. Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* **491**, 125–128 (2012).
45. Jiang, L. et al. Synthetic spike-in standards for RNA-seq experiments. *Genome Res* **21**, 1543–1551 (2011).
46. Liu, J., Xu, Y., Stoleru, D. & Salic, A. Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. *PNAS* **109**, 413–418 (2012).
47. Garcia-Moreno, M. et al. System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell* **74**, 196–211.e11 (2019).
48. Swigut, T. Mechanism for down-regulation of CD28 by Nef. *EMBO J.* **20**, 1593–1604 (2001).
49. Cullen, B. R. Nuclear RNA export. *J. Cell Sci.* **116**, 587–597 (2003).
50. Harris, M. E. & Hope, T. J. RNA export: insights from viral models. *Essays Biochem* **36**, 115–127 (2000).
51. Yedavalli, V. S. R. K., Neuveut, C., Chi, Y.-H., Kleiman, L. & Jeang, K.-T. Requirement of DDX3 DEAD box RNA helicase for HIV-1 Rev-RRE export function. *Cell* **119**, 381–392 (2004).
52. Davis, A. J., Li, P. & Burrell, C. J. Kinetics of viral RNA synthesis following cell-to-cell transmission of human immunodeficiency virus type 1. *J. Gen. Virol.* **78**, 1897–1906 (1997).
53. Kim, S. Y., Byrn, R., Groopman, J. & Baltimore, D. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. *J. Virol.* **63**, 3708–3713 (1989).
54. Folks, T. M., Justement, J., Kinter, A., Dinarello, C. A. & Fauci, A. S. Cytokine-induced expression of HIV-1 in a chronically infected promonocyte cell line. *Science* **238**, 800–802 (1987).
55. Hübner, W. et al. Sequence of human immunodeficiency virus type 1 (HIV-1) gag localization and oligomerization monitored with live confocal imaging of a replication-competent, fluorescently tagged HIV-1. *J. Virol.* **81**, 12596–12607 (2007).
56. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. **441**, 4 (2006).
57. Mouzakis, K. D., Lang, A. L., Vander Meulen, K. A., Easterday, P. D. & Butcher, S. E. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res* **41**, 1901–1913 (2013).
58. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell* **147**, 789–802 (2011).
59. Weill, L. et al. A new type of IRES within gag coding region recruits three initiation complexes on HIV-2 genomic RNA. *Nucleic Acids Res.* **38**, 1367–1381 (2010).
60. Puray-Chavez, M. et al. The translational landscape of SARS-CoV-2 and infected cells. *mBio.* **13**, 0081522 (2022).
61. Ingolia, N. T. et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* **8**, 1365–1379 (2014).
62. Clamer, M. et al. Active ribosome profiling with riboLace. *Cell Rep.* **25**, 1097–1108.e5 (2018).
63. Mills, E. W., Wangen, J., Green, R. & Ingolia, N. T. Dynamic regulation of a ribosome rescue pathway in erythroid cells and platelets. *Cell Rep.* **17**, 1–10 (2016).
64. Gilmer, O. et al. Structural maturation of the HIV-1 RNA 5′ untranslated region by Pr55 ^{Gag} and its maturation products. *RNA Biol.* **19**, 191–205 (2022).
65. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J. Am. Chem. Soc.* **127**, 4223–4231 (2005).
66. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1**, 1610–1616 (2006).
67. Irigoyen, N., Dinan, A. M., Brierley, I. & Firth, A. E. Ribosome profiling of the retrovirus murine leukemia virus. *Retrovirology* **15**, 10 (2018).
68. Zhang, H., Wang, Y. & Lu, J. Function and Evolution of Upstream ORFs in Eukaryotes. *Trends in Biochemical Sciences* https://doi.org/10.1016/j.tibs.2019.03.002 (2019).
69. Calviello, L. et al. DDX3 depletion represses translation of mRNAs with complex 5′ UTRs. *Nucleic Acids Res* **49**, 5336–5350 (2021).
70. Chen, H.-H., Yu, H.-I., Yang, M.-H. & Tarn, W.-Y. DDX3 Activates CBC-eIF3-Mediated Translation of uORF-Containing Oncogenic mRNAs to Promote Metastasis in HNSCC. *Cancer Res* **78**, 4512–4523 (2018).
71. Guenther, U.-P. et al. The helicase Ded1p controls use of near-cognate translation initiation codons in 5′ UTRs. *Nature* **559**, 130–134 (2018).
72. Bansal, A. et al. Enhanced recognition of HIV-1 cryptic epitopes restricted by HLA class i alleles associated with a favorable clinical outcome. *JAIDS J. Acquired Immune Defic. Syndromes* **70**, 1–8 (2015).
73. Berger, C. T. et al. Viral adaptation to immune selection pressure by HLA class I–restricted CTL responses targeting epitopes in HIV frameshift sequences. *J. Exp. Med.* **207**, 61–75 (2010).
74. Bet, A. et al. The HIV-1 Antisense Protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* **12**, 15 (2015).
75. Cardinaud, S. et al. Identification of cryptic MHC I–restricted epitopes encoded by HIV-1 alternative reading frames. *J. Exp. Med.* **199**, 1053–1063 (2004).
76. Cardinaud, S. et al. CTL escape mediated by proteasomal destruction of an HIV-1 Cryptic Epitope. *PLoS Pathog.* **7**, 15 (2011).
77. Lefebvre, G. *et al*. Analysis of HIV-1 Expression Level and Sense of Transcription by High-Throughput Sequencing of the Infected Cell. *jvi.asm.org*.
78. Berkhout, B., Arts, K. & Abbink, T. E. M. Ribosomal scanning on the 5′-untranslated region of the human immunodeficiency virus RNA genome. *Nucleic Acids Res.* **39**, 5232–5244 (2011).
79. Miele, G., Mouland, A., Harrison, G. P., Cohen, E. & Lever, A. M. The human immunodeficiency virus type 1 5′ packaging signal structure affects translation but does not function as an internal ribosome entry site structure. *J. Virol.* **70**, 944–951 (1996).
80. Grigorov, B., Arcanger, F., Roingeard, P., Darlix, J.-L. & Muriaux, D. Assembly of Infectious HIV-1 in Human Epithelial and T-Lymphoblastic Cell Lines. *J. Mol. Biol.* **359**, 848–862 (2006).

81. Grigorov, B. et al. A role for CD81 on the late steps of HIV-1 replication in a chronically infected T cell line. *Retrovirology* **6**, 28 (2009).
82. Rao, S. et al. Selective cell death in HIV-1-infected cells by DDX3 inhibitors leads to depletion of the inducible reservoir. *Nat. Commun.* **12**, 2475 (2021).
83. Ringeard, M., Marchand, V., Decroly, E., Motorin, Y. & Bennasser, Y. FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature* **565**, 500–504 (2019).
84. Dick, R. A., Goh, S. L., Feigenson, G. W. & Vogt, V. M. HIV-1 Gag protein can sense the cholesterol and acyl chain environment in model membranes. *Proc. Natl Acad. Sci. USA* **109**, 18761–18766 (2012).
85. Favard, C. *et al*. HIV-1 Gag specifically restricts PI(4,5)P2 and cholesterol mobility in living cells creating a nanodomain platform for virus assembly. *Science Advances* https://doi.org/10.1126/sciadv.aaw8651 (2019).
86. Nieto-Garai, J. A. et al. Cholesterol in the Viral Membrane is a Molecular Switch Governing HIV-1 Env Clustering. *Adv. Sci. (Weinh.)* **8**, 2003468 (2021).
87. Ono, A., Waheed, A. A. & Freed, E. O. Depletion of cellular cholesterol inhibits membrane binding and higher-order multimerization of human immunodeficiency virus type 1 Gag. *Virology* **360**, 27–35 (2007).
88. Elliott, J. L. et al. Integrase-RNA interactions underscore the critical role of integrase in HIV-1 virion morphogenesis. *eLife* **9**, e54311 (2020).
89. Liu, S. et al. HIV-1 integrase binding to genomic RNA 5′-UTR induces local structural changes in vitro and in virio. *Retrovirology* **18**, 37 (2021).
90. Rocchi, C. et al. The HIV-1 Integrase C-Terminal Domain Induces TAR RNA Structural Changes Promoting Tat Binding. *Int J Mol Sci.* **23**, 13742 (2022).
91. Winans, S. & Goff, S. P. Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. *PLOS Pathog.* **16**, e1009147 (2020).
92. Köppke, J. et al. Direct translation of incoming retroviral genomes. *Nat. Commun.* **15**, 299 (2024).
93. Berkhout, B. Function of the human immunodeficiency virus leader RNA. *Prog. Nucleic Acid Res Mol. Biol.* **54**, 1–34 (1996).
94. Haché, G., Abbink, T. E. M., Berkhout, B. & Harris, R. S. Optimal Translation Initiation Enables Vif-Deficient Human Immunodeficiency Virus Type 1 To Escape Restriction by APOBEC3G. *J. Virol.* **83**, 5956–5960 (2009).
95. Meijer, H. A. & Thomas, A. A. M. Control of eukaryotic protein synthesis by upstream open reading frames in the 5′-untranslated region of an mRNA. *Biochemical J.* **367**, 1–11 (2002).
96. Morris, D. R. & Geballe, A. P. Upstream open reading frames as regulators of mRNA translation. *Mol. Cell. Biol.* **20**, 8635–8642 (2000).
97. Jayaram, D. R. et al. Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc. Natl Acad. Sci. Usa.* **118**, e2018899118 (2021).
98. Antón, L. C. & Yewdell, J. W. Translating DRiPs: MHC class I immunosurveillance of pathogens and tumors. *J. Leukoc. Biol.* **95**, 551–562 (2014).
99. Ouspenskaia, T. *et al*. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat Biotechnol* 1–9 https://doi.org/10.1038/s41587-021-01021-3 (2021).
100. Akondy, R. S. et al. Initial viral load determines the magnitude of the human CD8 T cell response to yellow fever vaccination. *Proc. Natl Acad. Sci. USA.* **112**, 3050–3055 (2015).
101. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
102. Mangeot, P. E. et al. Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
103. Mangeot, P. E., Guiguettaz, L., Sohier, T. J. M. & Ricci, E. P. Delivery of the Cas9/sgRNA ribonucleoprotein complex in immortalized and primary cells via virus-like particles ('Nanoblades'). *J. Vis. Exp.* https://doi.org/10.3791/62245 (2021).
104. Goujon, C. et al. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559–562 (2013).
105. Pérès, E. et al. PDZ domain-binding motif of Tax sustains T-cell proliferation in HTLV-1-infected humanized mice. *PLoS Pathog.* **14**, e1006933 (2018).
106. Heyer, E. E., Ozadam, H., Ricci, E. P., Cenik, C. & Moore, M. J. An optimized kit-free method for making strand-specific deep sequencing libraries from RNA fragments. *Nucleic Acids Res.* **43**, e2–e2 (2015).
107. Adiconis, X. et al. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013).
108. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
109. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
110. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
111. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
112. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
113. Dunn, J. G. & Weissman, J. S. Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* **17**, 958 (2016).
114. Lauria, F. et al. riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput Biol.* **14**, e1006169 (2018).
115. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
116. Xiao, Z. et al. De novo annotation and characterization of the translatome with ribosome profiling data. *Nucleic Acids Res* **46**, e61 (2018).
117. Zhang, P. et al. Genome-wide identification and differential analysis of translational initiation. *Nat. Commun.* **8**, 1749 (2017).
118. Choudhary, S., Li, W. & Smith, A. D. Accurate detection of short and long active ORFs using Ribo-seq data. *Bioinformatics* **36**, 2053–2059 (2020).
119. Pedregosa, F. Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.
120. Wu, T. et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
121. Schuhmacher, J. et al. Vaccination against RhoC induces long-lasting immune responses in patients with prostate cancer: results from a phase I/II clinical trial. *J. Immunother. Cancer* **8**, e001157 (2020).
122. Cardinaud, S. et al. Triggering of TLR-3, -4, NOD2, and DC-SIGN reduces viral replication and increases T-cell activation capacity of HIV-infected human dendritic cells. *Eur. J. Immunol.* **47**, 818–829 (2017).

123. Karabiber, F., McGinnis, J. L., Favorov, O. V. & Weeks, K. M. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* **19**, 63–73 (2013).
124. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinforma.* **11**, 129 (2010).

Acknowledgements

We gratefully acknowledge support from the PSMN (Pôle Scientifique de Modélisation Numérique) of the ENS de Lyon for all computing resources. We acknowledge Adeline Page from Protein Science Facility of SFR Biosciences (Université Claude Bernard Lyon 1, CNRS UAR3444, Inserm US8, ENS de Lyon) for mass spectrometry analysis. We acknowledge the contributions of the CELPHEDIA Infrastructure (http://www.celphedia.eu/), especially the center AniRA in Lyon for flow cytometry and the BSL3 facility. We thank all participants of the ANRS CODEX cohort and the NIH AIDS Research and Reference Reagent Program for providing drugs. We thank, Eléonore Pérès, Madeleine Duc-Dodon, Louis Gazzolo, Pierre Jalinot and Vincent Mocquet for kindly sharing immortalized T cells infected with HTLV-1. We thank Franck Picard from LBMC for technical advice regarding the multi-omics data analysis. We thank all members of the RMI2 team and Antoine Corbin for manuscript proofreading. We thank Andrea Cimarelli, head of the LP2L team at CIRI. We also thank the Dormeur Foundation, Vaduz, for providing AID ELISPOT Reader at CIMI and Bernard Maillere (CEA) and his team for granting access to their ELISPOT reader. We thank Delphine Muriaux for sharing the replication-competent HIV-1 Gag-GFP recombinant virus. We thank Fabien Blanchet and Aude Boulay (IRIM) who provided purified NL4-3 viral stocks for metabolic labeling of primary CD4 + T cells infection experiments. This work was funded by Agence Nationale des Recherches sur le SIDA et les Hépatites Virales (ANRS – ECTZ3306), Fondation FINOVI, Sidaction, the European Research Council (ERC-StG-LS6-805500) under the European Union's Horizon 2020 research and innovation programs and the ATIP-Avenir program to E.P.R. Emmanuel Labaronne was supported by a postdoctoral fellowship from Agence Nationale des Recherches sur le SIDA et les Hépatites Virales (ANRS – ECTZ41715). L.B is a former Sidaction fellow (Aide aux Equipes) and her current PhD-salary is supported by ANRS. L.E is supported by the CNRS and funding from ANRS/MI ECTZ118944. T.J.M.S benefited from a fourth year PhD grant by Fondation pour la recherche médicale (FRM) The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

E.P.R. conceived the study with input from T.O. E.P.R., E.L., and D.D. designed and performed most experiments in cell lines. L.B., B.C.R., O.N., and A.M. designed and performed experiments with samples from healthy donors and people living with HIV-1, with help and input from P.F., I.H., O.L., A.S., and B.A. to collect and process samples. L.G performed most flow cytometry assays, prepared samples for proteomics experiments and performed DDX3-rescue experiments. T.J.M.S. generated Rpl7a-Flag tagged cell lines, performed ribosome immunoprecipitation from infected cells and prepared the corresponding ribosome footprinting cDNA libraries. D.C. analyzed the proteomics datasets and performed the multi-omics analysis that integrates the proteomics, RNA-seq and Ribo-seq datasets. V.V.B. and J.C.P. performed all SHAPE experiments to monitor the RNA secondary structure of wild-type and mutant HIV-1 5′ untranslated-regions. A.L.C.V., C.D., L.E., and C.G. performed infection experiments with primary human CD4 + T lymphocytes to monitor translation rates. E.P.R. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-56772-3.

Correspondence and requests for materials should be addressed to Emiliano P. Ricci.

Peer review information *Nature Communications* thanks Sandra Pankow, and the other, anonymous, reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Laboratoire de Biologie et Modelisation de la Cellule, Ecole Normale Superieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Universite Claude Bernard Lyon 1, 46 allee d'Italie F-69364, Lyon, France. [2]ADLIN Science, Evry-Courcouronnes, France. [3]CIRI, Centre International de Recherche en Infectiologie, Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS, UMR5308, ENS de Lyon, F-69007, Lyon, France. [4]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), Gif-sur-Yvette, France. [5]Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, Strasbourg, France. [6]IRIM, CNRS, Montpellier University, Montpellier, France. [7]Université Paris Saclay, Inserm, CEA, AP-HP, UMR1184 IDMIT, Department of Internal Medicine & Clinical Immunology, Bicêtre Hospital, Le Kremlin Bicêtre, Paris, France. [8]CEA, DSV/iMETI, Division of Immuno-Virology, IDMIT, Fontenay-aux-Roses, France. [9]Sorbonne Université, Inserm U1135, CNRS ERL 8255, Centre d'Immunologie et des Maladies Infectieuses (CIMI-Paris), Paris, France. [10]Mucosal entry, persistance and neuro-immune control of HIV and other viruses. Institut Cochin, INSERM U1016, CNRS UMR8104, Université de Paris Cité, Paris, France. [11]These authors contributed equally: Didier Décimo, Lisa Bertrand. ✉e-mail: emiliano.ricci@ens-lyon.org