nature communications

Article

https://doi.org/10.1038/s41467-025-56773-2

Unveiling conserved HIV-1 open reading frames encoding T cell antigens using ribosome profiling

Received: 11 August 2023

Accepted: 29 January 2025

Published online: 18 February 2025

Check for updates

Lisa Bertrand[1,2], Annika Nelde[3,4,5,17], Bertha Cecilia Ramirez[1,17], Isabelle Hatin[1,17], Hugo Arbes[1], Pauline François[1], Stéphane Demais[1], Emmanuel Labaronne[6,7], Didier Decimo[8], Laura Guiguettaz[6], Sylvie Grégoire[1,2], Anne Bet[2], Guillaume Beauclair[1], Antoine Gross[9], Maja C. Ziegler[10], Mathias Pereira[1,2], Raphaël Jeger-Madiot[2], Yann Verdier[11], Joelle Vinh[11], Sylvain Cardinaud[2,12], Stéphanie Graff-Dubois[2], Audrey Esclatine[1], Cécile Gouttefangeas[4,5,13], Marcus Altfeld[10], Laurent Hocqueloux[14], Assia Samri[2], Brigitte Autran[2], Olivier Lambotte[15], Hans-Georg Rammensee[4,5,13], Emiliano P. Ricci[6], Juliane Walz[3,4,5,13,16], Olivier Namy[1] ✉ & Arnaud Moris[1,2] ✉

The development of ribosomal profiling (Riboseq) revealed the immense coding capacity of human and viral genomes. Here, we used Riboseq to delineate the translatome of HIV-1 in infected CD4$^+$ T cells. In addition to canonical viral protein coding sequences (CDSs), we identify 98 alternative open reading frames (ARFs), corresponding to small Open Reading Frames (sORFs) that are distributed across the HIV genome including the UTR regions. Using a database of HIV genomes, we observe that most ARF amino-acid sequences are likely conserved among clade B and C of HIV-1, with 8 ARF-encoded amino-acid sequences being more conserved than the overlapping CDSs. Using T cell-based assays and mass spectrometry-based immunopeptidomics, we demonstrate that ARFs encode viral polypeptides. In the blood of people living with HIV, ARF-derived peptides elicit potent poly-functional T cell responses mediated by both CD4$^+$ and CD8$^+$ T cells. Our discovery expands the list of conserved viral polypeptides that are targets for vaccination strategies and might reveal the existence of viral microproteins or pseudogenes.

Open reading frames (ORFs) of the human genome have been long annotated using restricted criteria, including the presence of canonical start (AUG) and stop codons, and the potential to encode protein longer than 100 amino-acids (aa)[1]. Recent advances in detection methods challenged this definition revealing that thousands of small ORFs (sORFs) encode polypeptides or microproteins shorter than 100 aa[2]. These sORFs are widely spread along the human genome, some overlap classical ORF, on a different frame (so-called alternative reading frames (ARFs)), but the majority is found within 5′ untranslated regions (UTR) of known genes[3]. Although some microproteins encoded by sORFs have been shown to play fundamental roles, for instance, in DNA repair, mitochondrial functions, RNA regulation, etc[4]., the biological functions of most remain enigmatic. Nonetheless, the thorough characterization of sORFs greatly diversified the translation landscape in healthy and malignant cells expanding, in particular, the sources of cancer antigens.

To date the characterization of sORFs mostly relies on ribosomal profiling (Riboseq) that allows unbiased assessment of actively

A full list of affiliations appears at the end of the paper. ✉e-mail: olivier.namy@i2bc.paris-saclay.fr; arnaud.moris@i2bc.paris-saclay.fr

translated mRNA sequences. However, only a limited number of studies correlated Riboseq data with the generation of peptides[5]. Combining Riboseq and mass spectrometry (MS)-based immunopeptidomics several studies revealed that a fraction of peptides presented by major histocompatibility complex (MHC) molecules (known as ligandome) is derived from sORFs encoded polypeptides that are specific or overrepresented in tumor cells[6–9]. Other studies demonstrated that the tumor immune microenvironment is likely to favor the expression of human endogenous retroviruses (hERVs) that are normally silent in healthy tissues and of out-of-frame polypeptides, both constituting antigen sources for T cell immunity[10,11]. Remarkably, these findings and others[12] highlighted that noncanonical translation events of sORF located in 5′UTR or out-of-frame are likely favored in the context of stress.

Viral infections represent a major stress to the cell whose translation machinery is mobilized for the translation of viral RNAs. Viruses acquired specific features to control viral mRNA translation such as sequences favouring frameshifting or internal entry sites of ribosomes[13]. Thereafter, not surprisingly during the last decades, peptides derived from alternative translation events were described in cells infected with influenza virus[14–16], murine leukemia virus (MLV)[17], and human immunodeficiency virus-1 (here referred to as HIV)[18–20].

The expression of these noncanonical viral peptides often referred to as cryptic epitopes (CE), was mainly studied in the context of peptide presentation by MHC class I (MHC-I) molecules to cytotoxic $CD8^+$ T cells (CTLs). To date, in the context of viral infections, only a limited number of studies have combined Riboseq and immunopeptidomic approaches to characterize the landscape and origin of viral peptides presented by MHC-I molecules. These studies confirmed that MHC-I molecules present peptides from unannotated ORFs within ARFs of the human cytomegalovirus (HCMV)[21] and of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[22].

In the context of HIV infection, the existence of peptides derived from ARFs (ARFP) was, so far, revealed using T cell-based assays[20,23–26]. CTLs specific to peptides derived from ARFs were indeed detected during the acute and chronic phases of infection[18–20,24]. ARFP-specific T cells seem preferentially abundant in people living with HIV (PLWH) with favorable clinical outcomes expressing protective human leukocyte antigen (HLA) alleles[23]. Further underlining the contribution of ARFP-specific CTLs in the control of viral replication, several labs have shown that HIV adapts to and escapes from ARFP-specific T cell responses by introducing mutations within ARF sequences[18–20,24]. Remarkably, in the macaque model of simian immunodeficiency virus (SIV) infection, a single mutation within an ARF-derived epitope was strongly associated with viral rebound[27]. Bioinformatics approaches analysing the association between HLA polymorphisms and HIV sequence variations (HLA footprint) revealed that the virus might produce peptides from ARFs buried within sense (5′ to 3′) or antisense (3′ to 5′) frames of the viral mRNA[18,19].

In the present study, we used Riboseq to delineate the translatome of HIV in infected $CD4^+$ T cells. In addition to canonical viral protein-coding sequences (CDSs), we identify 98 ARFs, corresponding to sORFs that are distributed across the HIV genome including the 5′ UTR region. Using a database of complete and unique HIV genomes, we show that most ARF aa sequences are likely conserved among HIV-1 clade B and C strains, with eight ARF aa sequences being more conserved than the overlapping aa CDS sequences. Using T cell-based assays and mass spectrometry-based immunopeptidomics, we demonstrate that 43 ARFs encode viral polypeptides. In peripheral blood mononuclear cells (PBMCs) of PLWH, ARF-derived peptides elicit potent and polyfunctional T cell responses mediated by both $CD4^+$ and $CD8^+$ T cells. Our findings broaden the spectrum of HIV immunogenic antigens that might help in the design of efficacious vaccines and might reveal the existence of viral microproteins or pseudogenes.

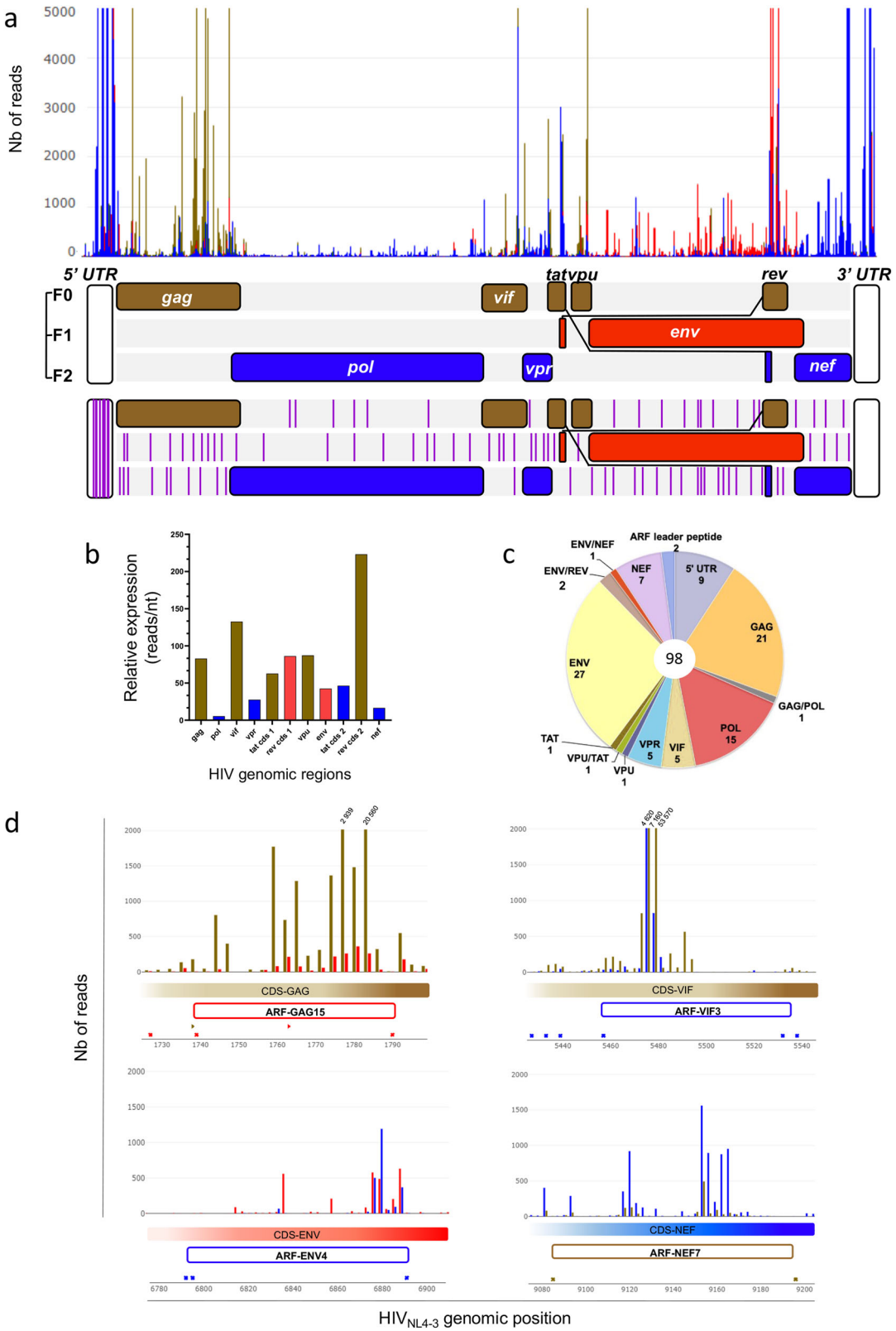
Results

HIV ribosome profiling reveals 98 alternative reading frames

To date, the existence of ARFs in the HIV genome has been highlighted using indirect approaches such as T cell-based assays and HLA footprint analyses. We thus intended to provide an unbiased assessment of actively translated viral mRNA sequences in HIV-infected cells. To this end, we infected the SupT1 $CD4^+$ T cell line with $HIV_{NL4\text{-}3}$, in three biological replicates, and analysed viral mRNA translation using Riboseq (Supplementary Fig. 1a). Twenty-four hours post-infection, the viability of the cells and the infection rates were analysed using a viability dye and a combination of anti-HIV-Gag and anti-CD4 antibodies. Combining Gag and CD4 staining allowed the direct detection of cells expressing HIV-Gag protein and the indirect assessment of cells expressing viral proteins that downmodulate CD4 expression such as HIV-Nef, -Vpu, or -Env[28]. Among the three biological replicates, the viability of the cells ranged from 44 to 71%, and the rate of infected Gag^+CD4^- and Gag^+CD4^+ cells from 54 to 73% (Supplementary Fig. 1b). The cells were harvested, and cellular and viral polysomes extracted. The ribosome-protected fragments (RPFs) were then isolated and sequenced (Supplementary Fig. 1a). Data were analysed using the RiboDoc tool[29] and inhouse bioinformatics pipelines. The length of isolated RPFs ranged between 25 and 35 nucleotides (nts). We analysed the periodicity of the reads by aligning RPFs to annotated start codons of human CDSs. The reads, in particular the 28- and 29-mers, uniformly aligned along the human CDSs with a periodicity of three nucleotides corresponding to codon triplets, thus strongly suggesting that the RPFs correspond to bona fide translation events (Supplementary Fig. 1c). The data from the three biological replicates were then combined and viral RPFs aligned to the $HIV_{NL4\text{-}3}$ genome (Fig. 1a and Supplementary Fig. 2a, b). The ribosome footprints aligned only on the forward strand throughout the viral genome. Note that in Fig. 1, each reading frame is highlighted by a specific color: brown, red, and blue for Frame 0 (F0), 1 (F1), and 2 (F2), respectively (Fig. 1a). For clarity, the RPF occupancy for the different reading frames is also presented in Supplementary Fig. 2b in three separate graphs. The number and density of RPFs were particularly high in the UTR regions of HIV RNAs (Fig. 1a). As expected, we observed patches of RPFs in the frame and covering the CDSs of known viral proteins such as HIV-Gag and -Env (Fig. 1a, in brown and red, respectively). The decrease in density and number of RPFs covering the CDS of HIV-Pol also highlighted the specific feature of the HIV-GagPol polyprotein expression, which is controlled by a -1 ribosomal frameshifting during translation elongation[30]. Based on the number of RPFs and the length of the CDSs, we calculated the relative expression of each CDS (Fig. 1b). The relative expression of HIV-GagPol polyprotein was about 6% of HIV-Gag, which is consistent with the literature[30]. Thereafter, our ribosome profiling of HIV translation identifies all known HIV CDSs and confirms specific features of the regulation of HIV translation such as ribosome frameshifting.

Strikingly, we also observed patches of RPFs overlapping known CDSs but aligning on other reading frames (e.g., patches of RPFs in frame-2 (red) overlapping HIV-Gag CDS or RPFs in frame-2 (blue) overlapping HIV-Vif CDS) suggesting translation of ARFs in the HIV genome (Fig. 1a and Supplementary Fig. 2). In our study having no information on potential start codons, we defined an ARF as a sequence between two stop codons with minimal size of ten codons (shorter sequences being unlikely to be processed to produce MHC ligands). In order to discriminate between background whole RNA sequencing and bona fide translation events, we then determined the number of reads in a frame within each potential ARF, which allowed calculating a mean of RPF/codon for each ARF. We observed that 75% of all ARFs had less than 8 RPF/codon and decided to use this value as the minimum threshold to define the bona fide translation of ARFs. Based on these criteria, we identified 98 HIV ARF sequences distributed across the viral genome (Fig. 1a bottom panel purple lines, 1c,

https://doi.org/10.1038/s41467-025-56773-2



d and Supplementary Table 1). Examples of potential ARFs that failed or succeeded in passing our selection criteria are presented in Supplementary Fig. 2c. Note that ARFs were named based on the CDS that they overlap plus a specific number (e.g., HIV-GAG15). In Fig. 1d, four examples of selected ARFs overlapping HIV-Gag, -Vif, -Env, and -Nef CDSs are highlighted, named ARF-GAG15, ARF-VIF3, ARF-ENV4, and ARF-NEF7, respectively. Note that we also analysed the presence of these 98 ARFs in each dataset of the biological replicates: 86 ARFs were common to at least two replicates, and 12 were found in only one of the replicates. Nonetheless, since these 12 ARFs exhibited a very high coverage, we decided to maintain them in our study. Interestingly, 16 out of 98 ARFs were previously described in the literature using indirect approaches[18,19,25,26] (Supplementary Table 1). We also identified nine ARFs in the 5′ UTR region of the HIV genome, commonly referred

Fig. 1 | The translational landscape of HIV-1: identification of ARFs. **a** Distribution of ribosome-protected fragments (RPFs) across the HIV-1 genome obtained from HIV_{NL4-3}-infected SupT1 cells. The number of RPFs (Nb of reads) are represented according to the position in the HIV_{NL4-3} genome. For clarity, the numbers of RPFs are truncated above 5000 reads. To illustrate the translation signal in the UTR duplicated regions, multimapped RPF are shown. RPFs translated in frames 0, +1, and +2 are represented in brown, red, and blue, respectively. HIV genomic RNA is represented below; CDSs are indicated and colored according to their reading frames from the first position of the genome. The positioning of ARFs (purple lines) are indicated in the lower RNA genome representation. The selection criteria were: (1) not corresponding to CDS sequences; (2) aa length ≥10; (3) translated between 2 stops codons, and (4) reads per codon ≥8. **b** Relative expression of HIV CDSs. Relative expressions of HIV_{NL4-3} CDSs were obtained by dividing the total read counts of each RPF by the CDS length, in nucleotide, establishing a read/nucleotide value. As in (**a**), the color codes indicate the frames in which the ARFs are localized (brown, 0; red, +1; blue, +2). **c** Pie chart of the genomic distribution of the 98 identified ARFs according to the overlapping CDS. ARFs that overlap two CDS sequences are grouped into segments labeled with the name of the two CDSs. The number of ARFs is indicated below the name. **d** Ribosome densities reveal novel viral coding regions. The number of RPFs (Nb of reads) are represented (y-axis) according to the position in HIV_{NL4-3} genome (x-axis). The color codes indicate the frames of the ARF and the relative CDS (brown, 0; red, +1; blue, +2). Filled and open rectangles indicate the relative CDS and the ARF, respectively. ARFs within gag (top left), vif (top right), env (bottom left), and nef (bottom right) CDSs. Putative stop codons are labeled according to their reading frames with crosses.

to as uORF, which is consistent with previous studies describing translation events in the 5′ UTR regions of different viral mRNA[31–34]. Note that we observed a large number of RPFs on the 5′ UTR region upstream of the major splice site (SD1) (Fig. 1 and Supplementary Figs. 2, 3).

In a second set of Riboseq experiments, using lactimidomycin (LTM), a drug that stops the initiation step of translation and thus leads to the accumulation of ribosomes at start sites, we then define the start codons of each ARF. To further enrich ribosomes at start codons, LTM was combined with puromycin (PMY) to dissociate elongating ribosomes from mRNAs[35]. As previously, SupT1 $CD4^+$ T cells were infected with HIV_{NL4-3}, in three biological replicates, cell viability and infection rates were analysed as in Supplementary Fig. 1b. Cells were treated with LTM and PMY prior harvest, RPF isolation and sequencing. The RPFs of cellular mRNAs were analyzed for periodicity and metagene. Cellular RPFs were highly enriched at start codons, thus strongly suggesting that the combined treatment of LTM and PMY allowed the accumulation of ribosomes at start codons in treated cells (Supplementary Fig. 4). Viral RPFs were then aligned to the viral genome, and RPFs in frame with the newly identified ARFs analysed. We obtained heterogenous results with 71 ARFs where multiple RPF occupancy could be observed and 27 ARFs where it was not possible to distinguish them from the background. Taking into account the three replicates, we observed that the mean intensity of RPFs for all ARFs was 62 reads, and we decided to use this value as the minimum threshold. We also excluded RPFs that accumulated 15 nucleotides downstream of stop codons. Based on these restrictive criteria, we listed the potential initiation sites that are shared by the three biological replicates (Supplementary Table 2). For 26 ARFs, a single start position was identified, while for the others, one or two patches of RPFs were observed (Supplementary Table 2). Note that in the same ARF previously defined from stop-to-stop, translation might start at different initiating sites. Interestingly, among all potential initiation codons, 18 % were closely related to the classical ATG methionine start codon (near-cognate).

Using Riboseq, we, therefore, reveal the existence of 98 ARFs distributed across the HIV genome that are actively translated in infected $CD4^+$ T cells. We confirm and extend previous work showing that other regions than annotated HIV CDSs are indeed translated.

Identified ARFs are conserved among HIV clade B and C isolates

Assuming that these ARF-encoded peptides might be a reservoir of genetic novelty and a source of T cell antigens, we analysed their aa sequence conservation among clade B and C HIV isolates. Using the Los Alamos HIV Sequence Database, we created in-house databases of clade B and C HIV sequences containing complete HIV genome without aberrant mutations within known CDSs and isolated from different individuals. We obtained 1609 and 411 clade B and C sequences, respectively. Using UGENE software, we then aligned the ARFs that we identified in the HIV_{NL4-3} genome to the databases (Fig. 2). Owing to the diversity of potential translation initiation sites for most ARFs, we decided to perform this analysis using the ARF aa sequence from stop-to-stop. The ARF-encoded peptide aa sequence conservation ranged from 48% to 98% with a median of 88% among the clade B (Fig. 2A) and from 42% to 98% with a median of 89% among the clade C (Fig. 2B). We did not notice a particular trend in aa sequence conservation when comparing the reading frames in which the ARFs are located in relation to the overlapping CDS (Supplementary Fig. 5a). In contrast, statistically significant differences in aa sequence conservation could be observed depending on the overlapping CDS (Supplementary Fig. 5b). The most significant difference in aa sequence conservation was found between ARFs overlapping *pol* and *env* genomic regions, which is consistent with the fact that *env* CDS is the most variable region. There was no difference in ARF-encoded peptide aa sequence conservation between clades B and C regardless of the reading frame from which the ARFs are translated (Supplementary Fig. 5c). Finally, we did not observe a correlation between ARF-encoded peptide aa sequence conservation and ARF relative expression (Supplementary Fig. 5d).

We then investigated whether it is the conservation of the CDS that imposes the conservation of the ARFs. To this end, we analysed the percentage of aa sequence conservation of the overlapping CDS and of the third frame (i.e., corresponding neither to the CDSs nor to the ARFs sequences). To simplify the analysis, ARFs that overlap two ORFs were excluded. We observed significantly higher conservation of CDS regions compared to the ARF and third frame, with a median of 93, 88, and 88% aa conservation, respectively (Fig. 2C and Supplementary Fig. 5e). Interestingly, the most variable ARF among clades B and C, ARF-ENV3, overlaps the highly variable V1-loop of HIV-Env glycoprotein[36] (Fig. 2). These observations further highlight that ORFs encoding the known viral proteins required for viral replication are under positive selection. In addition, there is no statistical difference in aa sequence conservation between the ARFs and the third frames (median of 88% aa conservation for both, Supplementary Fig. 5e), which might imply that there is a low or no selection pressure on ARFs. However, it is particularly interesting to note that eight ARFs sequences are more conserved than their overlapping ORF: ARF-GAG9, -VIF5, -VPU1, -ENV4, -ENV9, -ENV20, -ENV21, and -NEF3 (Fig. 2C, ARF names highlighted with black dots).

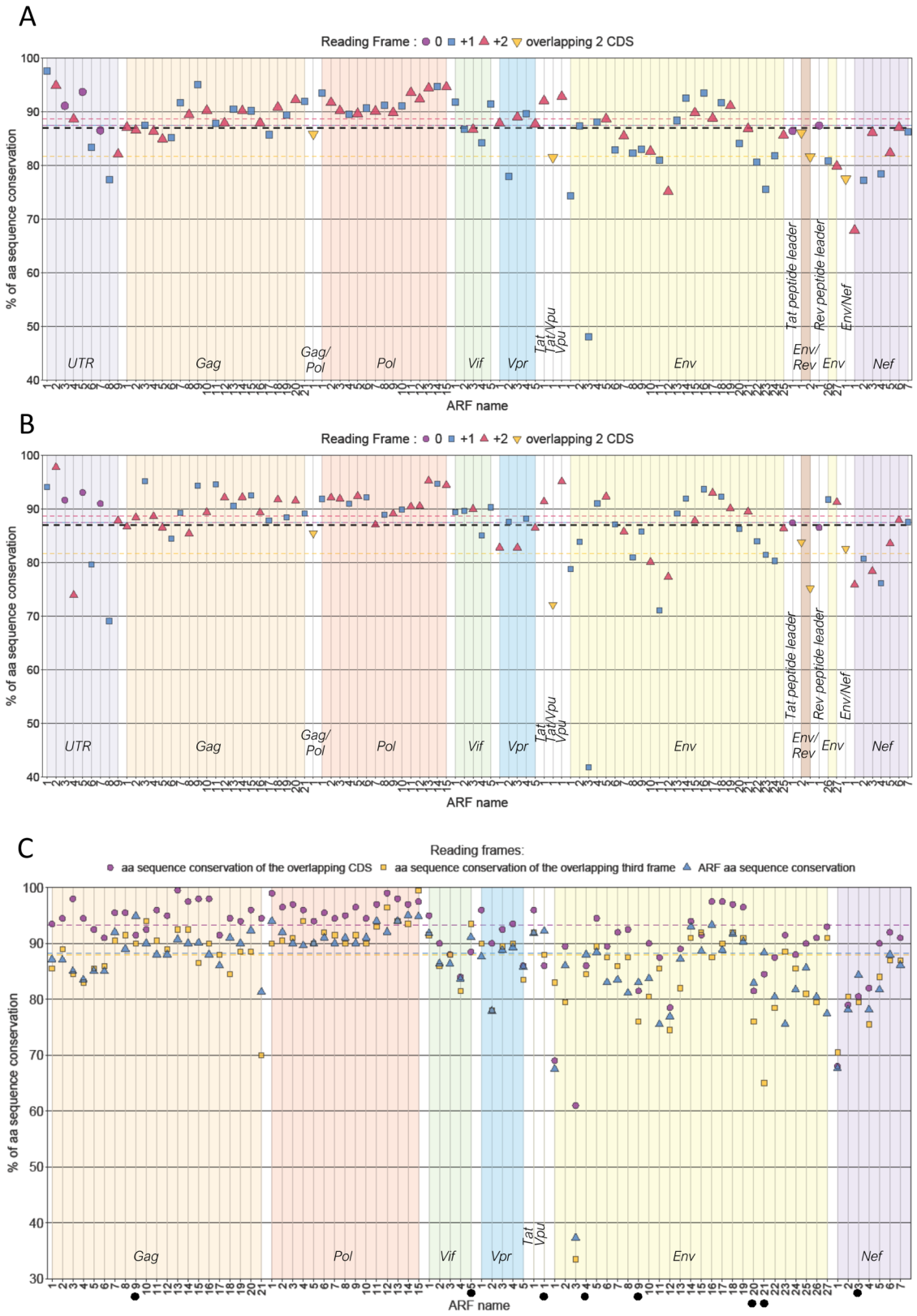
Overall, our HIV sequence analysis strongly suggests that the 98 noncanonical transcripts identified by ribosome profiling are likely conserved among HIV clade B and C clinical isolates. Although, most ARFs are probably not under selective evolutionary pressure, eight ARFs are more conserved than their overlapping CDSs.

To determine whether all these ARFs readily encode viral polypeptides, we then used two complementary approaches, monitoring immune responses to ARF-derived polypeptides using PBMCs of PLWH and using biochemical immunopeptodomics approaches to isolate directly, in infected cells, ARF-derived peptides.

42 ARFs encode viral polypeptides eliciting ARF-specific T cell responses

We first selected a set of peptides to be synthetized and then tested their capacity to elicit T cell responses using PBMCs of PLWH, using a

https://doi.org/10.1038/s41467-025-56773-2



cultured IFNγ-ELISPOT assay (Supplementary Fig. 6a). ARF-derived peptides (ARFP) were selected based on their aa sequence conservation among HIV isolates and their predicted capacity to bind HLA molecules exhibiting high prevalence in the general population (Supplementary Table 3). For ARFs with long potential aa sequences, several peptides were synthetized and tested (Supplementary Table 3). The peptide length varied in order to include as much as possible multiple potential epitopes within the same aa sequences. We then monitored T cell responses to these ARFP, using blood samples from individuals under antiretroviral treatment (ART) or individuals who naturally control viral replication without treatment, so-called elite controllers (EC) (Supplementary Table 4). In total, 96 different ARFP were synthesized and distributed in nine different peptide pools (Supplementary Table 3). As a control, we included peptides derived

Fig. 2 | Amino-acid conservation of ARF-encoded peptide sequences among HIV-1 clade B and C strains. Percentage of amino-acid (aa) sequence conservation of each identified ARFs among HIV-1 clade B (**A**) and clade C (**B**). ARFs are represented according to the reading frame of their overlapping CDSs (top legend). ARFs expressed in +1 or +2 are represented by a blue square and a red triangle, respectively. ARFs overlapping two CDSs are represented by a yellow reverted triangle and those that correspond to peptide leaders, that are expressed in the first reading frame or in the 5′UTR having no CDS, are represented by a purple circle. The horizontal lines indicate: the global median conservation in aa of ARF-encoded peptides (black dotted line), the median for ARF-encoded peptides in 0 (purple dotted line), +1 (blue dotted line), and +2 (red dotted line) frames, and the median for ARF-encoded peptides overlapping 2 CDSs (yellow dotted line). Genomic regions are indicated by background colors, with ARF names on the x-axis simplified for readability (e.g., "1" for ARF-GAG1), ranging from the 5′UTR (left) to the *nef* region (right). **C** Percentage of ARF-encoded peptide amino-acid conservation and of overlapping genomic regions, among HIV-1 clade B. Genomic regions are indicated by background colors, with ARF names on the x-axis simplified for readability (e.g., "1" for ARF-GAG1), ranging from the *gag* (left) to the *nef* region (right). The aa conservation of each encoded product of ARF (blue triangle), overlapping CDSs (purple circle), and third frame (yellow square) is presented. The colored horizontal lines indicate the corresponding median of the set. Black dots indicate ARF-encoded peptides that are more conserved than their overlapping CDS.

from canonical HIV proteins (HIV) and from non-HIV common viral peptides (CEF). The HIV-canonical peptide pool was composed of previously characterized T cell epitopes[37] (Supplementary Table 5).

PBMCs from PLWH were seeded in 24-well plates and loaded with the different peptide pools. Seven days later, a fraction of the cells was loaded with the peptide pool used for the initial culture, and T cell reactivity was monitored using an IFNγ-ELISPOT assay (Supplementary Fig. 6a, b). Although with different magnitudes, the PBMCs from all donors reacted in a specific manner to the pool of peptides derived from canonical HIV proteins (HIV+) compared to the negative control (HIV-) (i.e cells from the same T-cell cultures but loaded with DMSO-containing medium during the ELISPOT assay) (Supplementary Fig. 6b). Out of the seven samples from PLWH, six reacted to at least one ARF-derived peptide pool (Supplementary Fig. 6b). Cells from donor EC-1 reacted to eight ARFP pools. In contrast, we did not identify ARFP-specific response for donor ART-3, due to the high IFNγ secretion observed in negative controls (wells without peptide re-stimulation (−)). The response to the pool of classical HIV peptides was also particularly low in this donor (HIV+) (Supplementary Fig. 6b). Overall, several ARFP pools (including Pool-1, -2, -3, -4, and -6) induced very strong and specific IFNγ secretions in PBMCs from multiple HIV-positive donors (Supplementary Fig. 6b). The cells were then allowed to expand further by renewing the cytokine cocktail and tested on day 12, after the initial culture, for their capacity to react to individual peptides of the pools tested positive at day 7 (Fig. 3 and Supplementary Fig. 7). Note that despite the high background of IFNγ secretion observed at day 7, we tested the capacity of the cells from donor ART-3 to react to individual peptides of the pools 4, 6, and 8 because they showed a tendency to induce specific responses at day 7 (see Supplementary Fig. 6, ART-3). In addition, T cell responses where the spots appeared visually particularly large on day 7, even if not exceeding the positivity threshold in terms of spot numbers, were also tested on day 12. The reactivity of each PLWH to the individual ARFP is exhaustively presented in Supplementary Fig. 7. ARFP-induced IFNγ$^+$-T cell responses are summarized in Fig. 3 and pictures of the corresponding wells of the IFNγ-ELISPOT plates, are shown in Supplementary Fig. 7.

At day 12, with the exception of EC-2, all samples from PLWH exhibited very strong or saturated responses to classical HIV peptides. Remarkably, the PBMCs of all donors reacted to at least one individual ARFP (Fig. 3). Among the donors, the median of ARFP recognized was 6. However, responses were very heterogeneous with donor EC-1 reacting to 33 ARFP and EC-2 only to a single one (Fig. 3a). A total of 60 ARFP-specific T cell responses were identified across PLWH samples (Fig. 3a). Fourteen ARFP elicited T cell responses in samples from two different individuals (Fig. 3b). Altogether, the results using samples from PLWH revealed 46 unique ARF-derived peptide-specific T cell responses. The magnitude of T cell responses was heterogeneous between and among samples from PLWH, with an intensity ranging from 13 to above 1250 (saturated) SFU/10^6 PBMCs (Fig. 3c) with a median of 311 SFU/10^6 PBMCs, but highly significant over background (Supplementary Fig. 8a). For instance, despite exhibiting a broad ARFP-specific T cell response, the magnitude of most T cell responses from donor EC-1 were low as compared with those from other donors (Fig. 3 and Supplementary Fig. 8b). Finally, we did not observe a significant difference in the magnitude of T cell responses between ART and EC individuals (Supplementary Fig. 8b). Remarkably, ARFP-specific-T cell responses target peptides encoded by ARFs distributed all along the HIV genome sequence from the 5′ UTR to *nef* CDS (Fig. 3d). Although the magnitude of uORF-specific T cell responses seemed particularly high (Supplementary Fig. 8c), ARPF-induced T cell responses were not significantly dominated by T cells recognizing ARFP from one specific HIV genomic region, neither UTR nor CDS (Supplementary Fig. 8d).

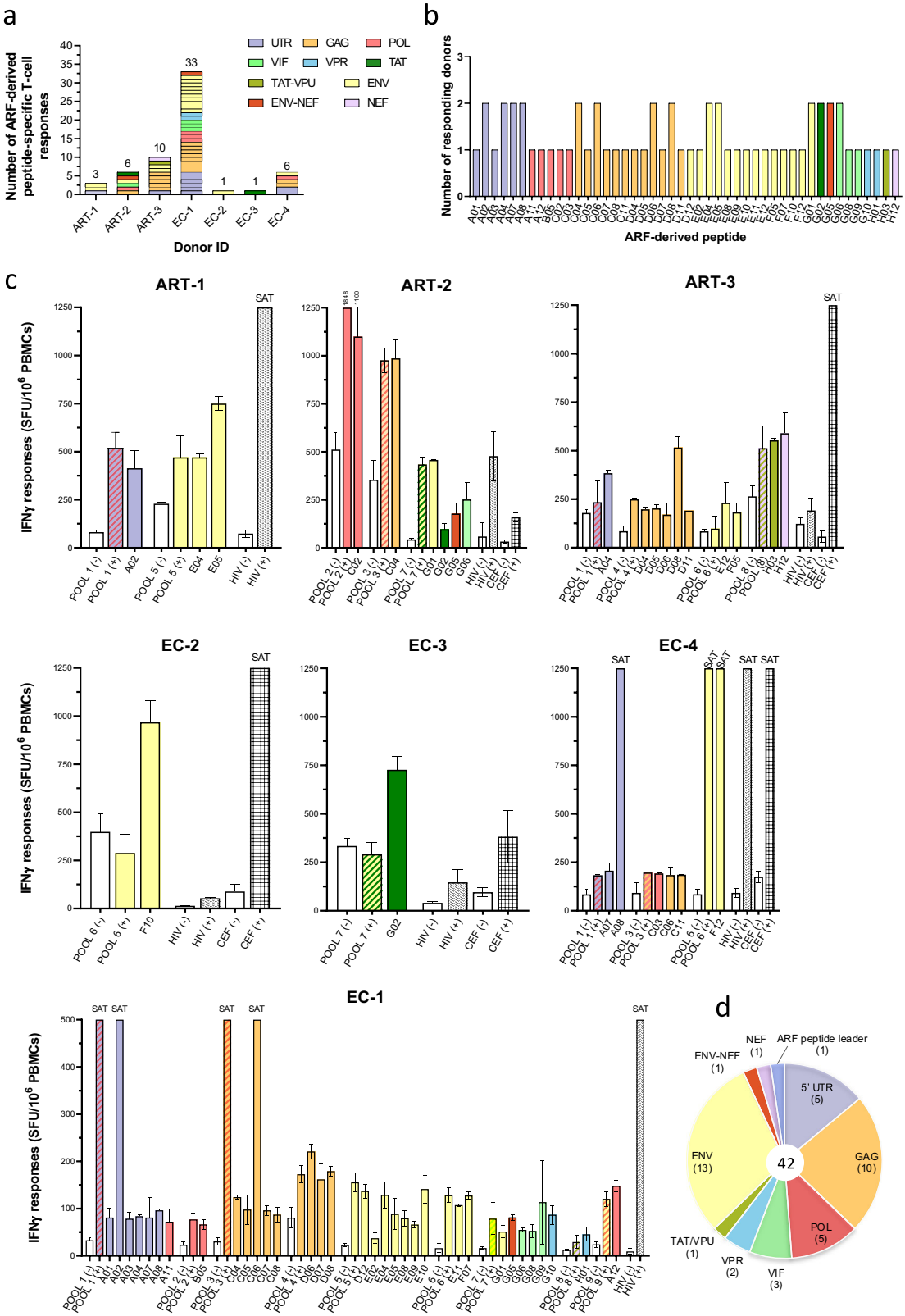
Using the exact same protocol, the entire set of ARF-derived peptide pools was also tested, at day 7, on PBMCs from non-infected donors ($n = 3$), and on day 12, peptide deconvolution was performed on pools stimulating IFNγ-responses (Supplementary Fig. 9a, b). Cross-reactivity to one peptide pool was observed for EFS 1619 and EFS 5174 samples (Supplementary Fig. 9a) but the peptide cross-reactivity was not confirmed when testing individual peptides of the pool. Donor EFS 0685 exhibited a high background on all tested peptide pools. Peptide deconvolution showed a cross-reactivity to three peptides that were then excluded from the study (Supplementary Fig. 9b).

Overall, we demonstrate here that T-cell responses from PLWH target 46 ARF-derived peptides encoded by 42 different ARFs (Fig. 3d). Of note, for most ARFP that elicited T cell responses, the corresponding aa sequence could be found at least partially in the aa sequence encoded from the translation initiation sites, identified in our LTM + PMY Riboseq experiments (see Supplementary Table 2 in red). We thus reveal that at least 42 ARFs identified in our ribosome profiling encode viral polypeptides and elicit T cell responses in vivo in the course of natural infection.

ARF-specific T cell responses are mediated by CD4$^+$ and CD8$^+$ T cells with a polyfunctional profile

We then asked whether CD8$^+$ and/or CD4$^+$ T cells mediate ARFP-specific T cell responses and studied the quality of these responses. The quality of HIV-specific T cell activation, defined as the capacity to produce multiple antiviral cytokines/chemokines, rather than the magnitude of T cell responses has been linked to disease outcome[38]. Therefore, we characterized the capacity of ARFP-specific T cells to produce MIP-1β, IFNγ, IL2, CD107a, and TNFα using Intracellular Cytokine Staining (ICS) (Fig. 4). PBMCs from two ART and four EC individuals (ART-2 was not tested in ICS due to a lack of sample availability) were seeded in 24-well plates and loaded with the individual ARFPs that induced T cell responses in the previous cultured IFNγ-ELISPOT assay (one ARFP per well). In total, we tested 47 ARF-derived peptides, 45 peptides which induced IFNγ-responses in the cultured IFNγ-ELISPOT assay plus two ARFP (one new ARFP (A05) and one (C12) inducing a slight but not significant T cell response in the first ELIPSOT) (Supplementary Table 6). In addition to the HIV+ peptide pool (HIV+), we included immunodominant peptides derived from classical HIV proteins with known HLA binding capacity[37].

https://doi.org/10.1038/s41467-025-56773-2



This pool of immunodominant peptides was adapted to the HLA allotype of each individual (Supplementary Table 7). At day 12, peptides inducing an IFNγ secretion were then used to monitor T cell activation using ICS combined with CD4, CD8, and CD3 extracellular staining (Fig. 4).

As illustrated with the ARF-derived peptide G02 and the HIV-Env-derived peptide env01, peptide loading led to high production levels of MIP-1β, IFNγ, IL2, CD107a, and TNFα by $CD4^+$ T cells as compared to the negative controls (NS_ARF and NS_ORF, respectively) (Fig. 4a). Overall, 27 peptides induced the secretion of at least one cytokine (Fig. 4b). Most T cell responses, 89%, were mediated by $CD4^+$ T cells (Fig. 4b, c). The ARF-derived peptides A05, G02, and G05 induced both $CD8^+$ and $CD4^+$ T cells (Fig. 4a right panel and 4b).

CDS-specific $CD4^+$ T cell responses were characterized by a significantly higher proportion of cells producing individual cytokines as compared with ARFP-specific $CD4^+$ T cell responses (Supplementary

Fig. 3 | Detection of T cell responses specific to ARF-derived peptides (ARFP) in PBMCs of PLWH. a Overview of ARFP-specific T cell responses. The number of ARFP-specific T cell responses is represented for each donor. The largest rectangle represents different peptides encoded by the same ARF. The color code, on the top right corner, indicates the overlapping CDS. **b** Number of responding PLWH to each ARFP stimulation. As in (**a**) the color code indicates the overlapping CDS. The rough data for IFNγ-Elispot responses are presented in Supplementary Fig. 7-related to Fig. 3. **c** IFNγ-ELISPOT deconvolution of positive pools at day 12 and for all tested donors. Only IFNγ-positive T cell responses upon individual ARFP stimulations are presented, in Spot Forming unit (SFU)/10^6 PBMCs. Intrinsic negative controls (POOL (−), cells incubated with DMSO-containing medium) are indicated by white bars for each tested pool and the respective peptides. POOL (+), HIV (+), and CEF (+) correspond to PBMCs loaded with the pool of ARF-derived, HIV-classical, and non-HIV common virus-derived peptides, respectively. The color code indicates the overlapping CDS. SAT: saturated well for which the IFNγ signal was too high to be quantified. Data are presented as mean values of technical triplicates ± SD. **d** Genomic repartition of ARF encoding immunogenic peptides. The number of ARFs is indicated in brackets. The color code is as in (**a**).

Fig. 10a) which was not the case for $CD8^+$ T cell responses (Supplementary Fig. 10b). However, although the frequency of responses against ARFP was lower than against CDS-derived peptides, ARFP stimulated polyfunctional responses in both $CD4^+$ and $CD8^+$ T cells (Fig. 4d, e). Since tri-functional T cell responses are often considered polyfunctional and might be associated with protection from disease progression[39], we focused our analysis on activated T cells with a polyfunctional profile with at least 3 functional responses. Clustering the data, we observed that 80% of ARFP-specific $CD4^+$ T cells secreted at least three cytokines simultaneously compared to 93% for canonical HIV-specific $CD4^+$ T cells (Fig. 4d). Ninety-one percent of $CD8^+$ T cell responses directed against canonical HIV peptides and 100% of ARFP-specific $CD8^+$ T cell responses harbored a tri-functional phenotype, respectively (Fig. 4e). Profiles and frequencies of cytokines responses induced by each ARF- and CDS-derived peptide for the six PLWH are presented as heatmaps for $CD4^+$ and $CD8^+$ T cells (Supplementary Fig. 10c, d). It was interesting to note that upon G02 peptide stimulation, the proportion of activated T cells secreting five cytokines simultaneously was higher than for some canonical HIV or even CEF peptide stimulations (see donor EC-3 donor, Supplementary Fig. 10c). In ART and EC individuals, we did not observe a significant difference in the capacity of ARF- and CDS-encoded peptides to induce polyfunctional cytokine productions (at least three functions) among neither $CD4^+$ nor $CD8^+$ T cells (Fig. 4f, g). There is also no significant influence of the clinical status on the polyfunctional profile of ARF- and CDS-specific T cell responses (Fig. 4f).

Finally, since the cytokine combination secreted by T cells might influence disease outcome[40], we dissected ARF- and CDS-specific T cell cytokine secretion patterns. To this end, we analyzed the 32 possible cytokine combinations (out of the five cytokines detected: MIP-1β, IL2, INFγ, CD107a, and TNFα) secreted by $CD4^+$ and $CD8^+$ T cells upon ARF- or CDS-derived peptide stimulation. Although the frequencies of $CD4^+$ T cells targeting ARFP were lower than that of CDS-derived peptides, activated $CD4^+$ T cells exhibited similar cytokine secretion patterns whether targeting ARF- or CDS-derived peptides (Supplementary Fig. 10e). The majority of polyfunctional ARFP-specific $CD4^+$ T cells produced TNFα, CD107a, IFNγ, and MIP-1β simultaneously. Among ARF- and CDS-specific $CD4^+$ T cells, the most prevalent monofunctional categories were $TNFα^+$ and $Mip\text{-}1β^+$ cells, respectively (Supplementary Fig. 10e). ARF- and CDS-specific $CD8^+$ T cell responses displayed similar pentafunctional profiles (Supplementary Fig. 10e).

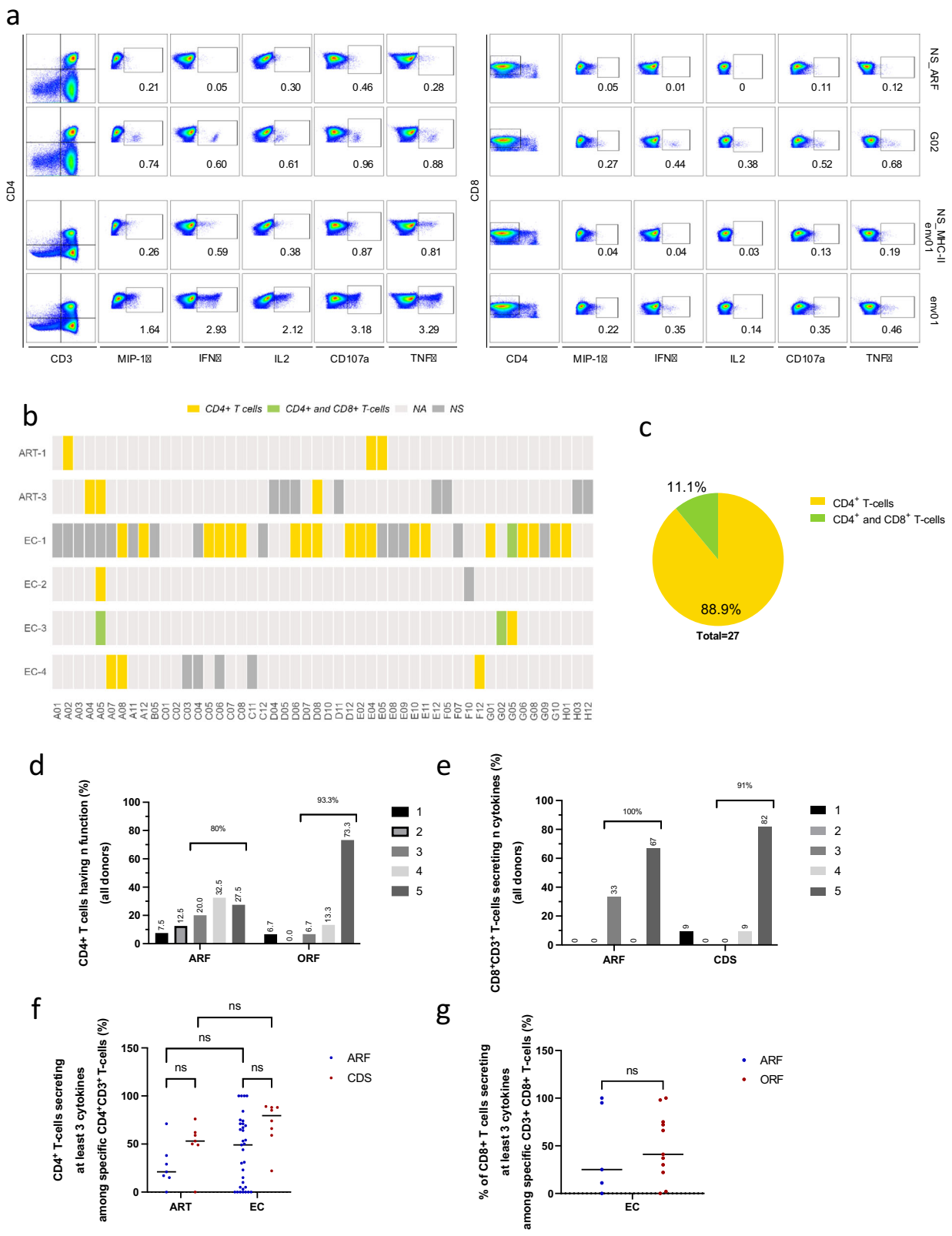
We then sought to analyze ex vivo ARF- and CDS-specific T cell responses in PBMCs of PLWH without in vitro expansion. To this end, cells from EC-1 and EC-3 were loaded with the individual peptides that induced a potent T cell activation in cultured IFNγ-ELISPOT and/or ICS (Supplementary Fig. 11). Remarkably for EC-1, four out of the six ARFP tested in ICS induced significant $CD4^+$ T cell responses and one peptide stimulated $CD8^+$ T cells (Supplementary Fig. 11a, left panel). For donor EC-3, the four ARFP induced significant $CD4^+$ T cell responses (Supplementary Fig. 11b, left panel). Overall, the frequency of ex vivo ARF- and CDS- specific T cells was very low, limiting the analysis of the polyfunctional profile of the cells (Supplementary Fig. 11a, b, pie charts on the right panels). Nonetheless, bivalent and pentafunctional ARFP-specific ex vivo T cell responses could be detected for single ARF-derived peptides (Supplementary Fig. 11).

Taken together, our multiparametric study showed that ARF-derived peptides are predominantly recognized by $CD4^+$ T cells from ART and EC donors and, to a lesser extent, by $CD8^+$ T cells. Following in vitro T cell expansion, the percentage of ARFP-specific T cell activation was lower than CDS-specific T cell responses. Nonetheless, some ARF-derived peptides induced a higher magnitude of T cell activation than CDS-derived peptides. ARF-specific T cells were readily detected ex vivo in the PBMCs of PLWH. ARF-derived peptide stimulation induced polyfunctional T cell activations.

Identification of a naturally presented HLA-A*02:01-restricted ARF-derived peptide on infected primary $CD4^+$ T cells

MHC ligands derived from sORFs that are specific or overrepresented in tumor cells were recently identified combining Riboseq and immunopeptidomic approaches[6–8]. In the context of viral infection, two studies also highlighted that HLA class I molecules can present peptides from ARFs of HCMV and SARS-CoV-2[21,22]. We thus intended to identify ARF-derived peptides presented by HLA class I and II molecules on the surface of HIV-infected cells (Supplementary Fig. 12a). We performed de novo experiments and also reanalyzed LC-MS/MS data from previously published immunopeptidomic experiments that used HIV-infected cells. To identify HIV-derived HLA-restricted peptides from all potential ARFs, the reviewed human proteome supplemented with the six-frame translated genome of $HIV_{NL4\text{-}3}$, taking into account all potential peptides ≥8 aa encompassed between two stop codons, was used as a reference. For our experiments, we used the $CD4^+$ T cell lines SupT1 (used in the Riboseq experiment) and C8166, which express HLA class I and both HLA class I and II molecules, respectively. In order to bypass the downmodulation of HLA molecule expression mediated by HIV-Nef, we used a *nef*-deficient $HIV_{NL4\text{-}3}$ isolate ($HIV_{NL4\text{-}3\Delta Nef}$)[41]. Cells were readily infected, with more than 50 and 20% of HIV-gag^+ SupT1 and C8166 cells, respectively (Supplementary Fig. 12b). For both cell lines, although we obtained a large number of HLA class I ligands (5673 for SupT1 and 7794 for C8166) and we detected known HIV-derived HLA class I ligands, we could not identify any ARF-derived HLA-restricted peptides. For C8166 cells, we additionally identified 4834 HLA class II-presented peptides from which 13 were derived from the HIV-Gag protein (Supplementary Fig. 12c). To our knowledge, these HIV-derived peptides are rare examples of natural HLA class II ligands identified in infected cells. Nonetheless, we did not detect any ARF-derived peptide. We thus reanalyse the immunopeptidomics data from our previously published work using $HIV_{NL4\text{-}3}$-infected primary $CD4^+$ T cells[42] and we identified seven HIV-derived peptides (Fig. 5a), one of which, ILINGQYSL, is derived from an ARF overlapping the *pol* gene (Supplementary Fig. 12e). Spectral validation using an isotopic labeled synthetic peptide confirmed the identification (Supplementary Fig. 12d). The ILINGQYSL peptide is annotated as a potent binder of HLA-A*02:01. Note that this ARF did not pass through the selection criteria of our initial Riboseq profiling because of its low coverage per codon (2.5 RPF/codon). However, it was readily identified in our Riboseq performed with LTM + PMY. LTM + PMY treatment led to the accumulation of RFPs at the leucine position 3222, suggesting that it might be the start codon of the ARF from which the ILINGQYSL peptide is derived (Supplementary Table 2, bottom line). This leucine is highly conserved in strains from clade B and poorly conserved for clade C (85

https://doi.org/10.1038/s41467-025-56773-2



versus 15% conservation) suggesting that other translation initiation events might occur in clade C isolates. Nonetheless, this ARF, is highly conserved (above 90% of aa conservation) among both HIV clades.

We then tested the capacity of the ILINGQYSL peptide (YSL-I) and an elongated form of YSL-I, LHSFGWVMNSILING (YSL-II), which has a high predictive score for binding to HLA-DR molecules (e.g., HLA-DRß1*03:01, *04:01, *07:01 according to the SYFPEITHI server), to activate T cell responses in PBMCs of PLWH. Using our cultured IFNγ-ELISPOT assay, we observed in PBMCs of EC-1 and EC-3 donors, intermediate to very strong T cell responses targeting the YSL-I peptide (Fig. 5b). The intracellular cytokine assay revealed that the YSL-I peptide is recognized by $CD4^+$ T cells with a polyfunctional profile equivalent to responses specific to immunodominant CDS-derived peptides (Fig. 5c, pie charts). Responses to YLS-I peptide were

Fig. 4 | ARF-derived peptide (ARFP)-specific T cell responses are mediated by CD4⁺ and CD8⁺ T cells exhibiting polyfunctional cytokine profiles. **a** Representative gating scheme for the identification of ARFP-specific T cell responses. Upper panel, gatings for G02-ARFP-specific CD4⁺ (left) and CD8⁺ (right) T cell responses are shown for the EC-3 PLWH. Lower panel, gatings for Env01-specific T cell responses as for G02. Env01 epitope is a known immunodominant peptide from HIV-Env. CD8⁺ CD4⁻ T cell populations were pre-gated on CD8⁺ CD3⁺ T cells, lived cells, and doublets were excluded. Gates for each cytokine/chemokine were set based on the negative controls (cells loaded with DMSO-containing medium, NS: Not stimulated) NS_ARF and NS_CDS for the assessment of G02-ARFP- and Env01-specific T cell responses, respectively. A figure exemplifying the gating strategy is provided in Supplementary Fig. 13.**b** Heatmap of ARFP recognized by CD4⁺ or CD8⁺ T cells from PLWH summarizing the ICS assays. Yellow and green rectangles are used to indicate CD4⁺ and both CD4⁺ and CD8⁺ T cell positive responses in ICS, as defined in (**a**), respectively. NA not applicable (i.e., not tested), NS not significant. **c** Percentage of CD4⁺ or CD8⁺ T cell responses among ARFP-specific T cell responses after in vitro stimulation with ARFP at the cohort level. Proportion of ARFP- and CDS-specific CD4⁺CD3⁺ (**d**) or CD8⁺CD3⁺ (**e**) T cells secreting 1 to 5 cytokines simultaneously. The numbers on the top indicate the percentage of polyfunctional T cells secreting at least three cytokines. ARFP-specific T cell responses from all tested donors are combined. **f** Comparison of the percentage of polyfunctional T cells secreting at least three cytokines among ARFP- and CDS-specific CD4⁺CD3⁺ T cells in the EC and ART patient groups. **g** Same as (**f**) among ARFP- and CDS-specific CD8⁺CD3⁺ T cells in the EC group. Each dot represents one peptide-specific T cell response. The lines correspond to the median responses. A two-sided Mann–Whitney test with Dunn's comparison was applied. ns: not significant. Data for YSL-I- and YSL-II-specific T cell responses are integrated in the graphs from (**d**–**g**).

particularly strong in CD4⁺ T cells from the EC-3 donor (up to 30% of CD4⁺ T cells) and were also detected, but to a weaker extent, in the cells from donors EC-1, EC-2, and EC-4 (Fig. 5d). Two donors also exhibited weak specific CD4⁺ T cell responses to the YLS-II peptide (Fig. 5e). Overall, in our cultured ELISPOT assay, 5 donors reacted to either YSL-I or YSL-II peptides (Fig. 5f). In addition, YSL-I-specific T cell responses with polyfunctional profiles were readily detected during ex vivo peptide stimulations in PBMCs of PLWH (Supplementary Fig. 11), clearly demonstrating that the YSL-ARF readily encodes viral polypeptides.

Therefore, using two complementary and independent approaches, detection of ARF-derived peptide-specific T cells in the PBMCs of PLWH and direct isolation of one HLA-bound ARF-derived peptide using mass spectrometry-based immunopeptidomics, we readily demonstrate that the HIV ARFs that we identified by ribosome profiling encode viral polypeptides capable of inducing broad T cell responses.

Discussion

We provide here the characterization of the HIV translatome in CD4⁺ T cells. We revealed the existence of at least 98 ARFs distributed across the HIV genome, including the 5′UTR, that are actively translated in infected CD4⁺ T cells. Most ARFs are likely conserved among HIV clade B and C clinical isolates. We demonstrated that at least 43 ARFs can be translated into viral polypeptides, since ARF-derived polypeptides were targeted by CD4⁺ and CD8⁺ T cells from PLWH. Remarkably, ARF-derived peptides induced a polyfunctional T cell memory response that is reminiscent of the ones targeting CDS-derived immunodominant epitopes. Finally, we identified a conserved ARF-derived epitope, ILINGQYSL, naturally presented on HIV-infected primary CD4⁺ T cells by the HLA-A*02:01 molecule. ILINGQYSL is translated from a highly conserved ARF and induces ex vivo a polyfunctional T cell activation. Using two complementary approaches, we demonstrated that the ARFs identified are actively translated into viral polypeptides with the capacity to induce potent T cell immunity.

In human cells, recent advances have revealed that sORFs are widely spread throughout the human genome, some overlapping different frames of classical ORFs or locating within the 5′ UTR of known genes[3]. Remarkably, these sORFs encode microproteins or polypeptides smaller than 100 aa[2,4]. In fact, it has been estimated that 85% of the translation products originate from non-annotated regions of the human genome and mostly from out-of-frame sequences of CDS[9,43]. We identified 98 ARFs within the HIV-1 genome, ranging from the 5′ UTR region to the *nef* gene. Nine ARFs are located in the 5′ UTR region. Several studies have recently demonstrated that polypeptides can be encoded in the 5′ UTR regions of viral genomes[8,9,32–34,44,45]. These ARFs located at the 5′ UTR are probably translation regulatory elements of downstream CDS as observed in the human genome[46].

In our dataset, in addition to the 5′ UTR, most ARFs were located within *gag, pol*, and *env* genes, which correlate with the size of the respective CDS. Previous studies, using HLA footprint or predictive ORF approaches, also found ARFs overlapping *gag, pol*, and *nef* sequences[18–20,25,26]. Only 16 HIV ARFs identified in our ribosomal profiling were in common with ARFs already described in the literature[19,25,26]. This relatively low number might be due to an overestimation of the number of ARFs by the various prediction methods or the lack of thorough annotations of some ARFs in the previous studies. It can be also related to the rigorous criteria that we applied in our study for the selection of ARFs. This is illustrated by the ARF overlapping *gag* CDS and encoding the Q9VF epitope that we previously characterized in HIV-infected cells[20,24] whose coverage (5.9 RPF/codon) here is just below our selection criteria. However, we have previously shown that HIV_{NL4-3}-infected cells are not recognized by Q9VF-specific CTL due to a proteasomal cleavage of the Q9VF epitope[20,24]. Remarkably, we also did not identify RPFs aligning in a frame on the minus strand of HIV. In particular, we did not identify the CDS encoding the antisense protein of HIV-1 (ASP) which overlaps the *env* gene, although several lines of evidence suggest that it is expressed in vivo[47,48]. The lack of ASP CDS detection might be due to specific regulation of translation from the 3′ UTR that could be down-modulated in lymphoid cells such as CD4⁺ T cells used in our study[49]. Nonetheless, we report here the identification of ARFs translated from the *vif, vpr, vpu, tat*, and *env* sequences as well as ARF overlapping *tat* and *rev* exons.

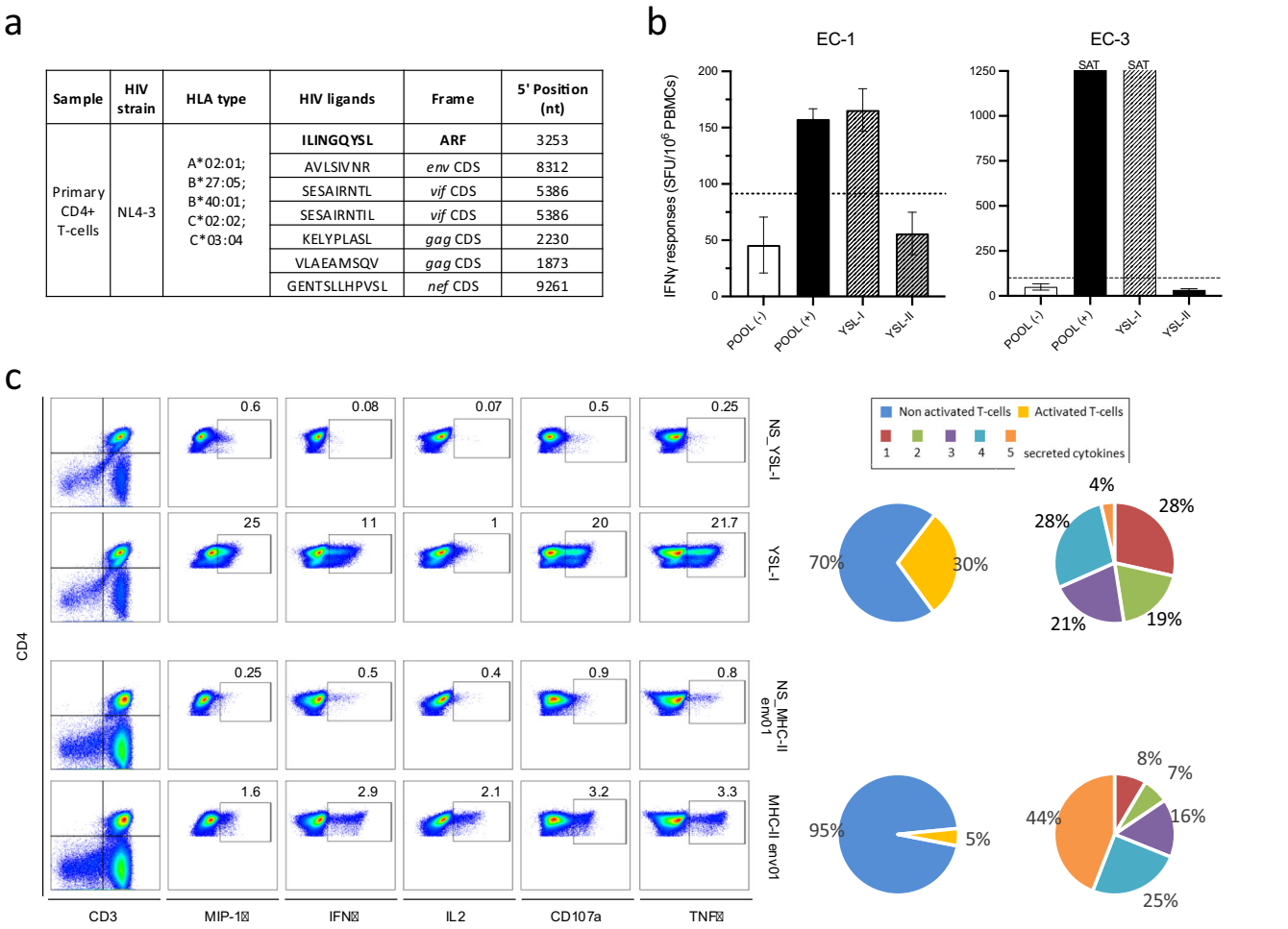
Combining ribosome profiling with LTM and PMY, we intended to define the translation initiation sites of the newly described ARFs. We obtained heterogenous results: for 26 ARFs a single start site was identified while for the others one or two initiation sites were delineated by patches of RPFs (Supplementary Table 2). Interestingly, among all potential initiation codons, 18% were closely related to the classical ATG methionine start codon. However, these results strongly suggest that a variety of mechanisms might be responsible for ARF expression. Indeed, a large part of eukaryotic and viral CDSs or ARFs initiate translation by noncanonical mechanisms such as leaky scanning, stop codon readthrough, ribosome shunting, re-initiation, IRES, and frameshifting[13]. Due to the diversity of potential translation initiation sites for most ARFs, we defined the ARF using flanking stop codons which also probably has an influence on the threshold of ARF RFP/codon and thus ARF identification.

Based on these criteria, the length of the identified ARFs ranged from 10 to 101 aa and can therefore be considered as sORF[50]. Using an in-house HIV database of clade B and C strains, we propose that the ARF-encoded aa sequences might be conserved among both clades. Overall, we observed that the overlapping CDSs are significantly more conserved than the ARFs and the third frame, suggesting that ARFs might be under a selective pressure weaker than classical CDSs. Remarkably, the ARF, encoding the ILINGQYSL peptide, is highly conserved (above 90% of aa conservation) among HIV clade B and C, which is probably explained by the fact that it overlaps the sequence encoding the catalytic site of the HIV reverse transcriptase (RT). This sequence of the RT is crucial for HIV replication and is among the most

https://doi.org/10.1038/s41467-025-56773-2

a

Sample	HIV strain	HLA type	HIV ligands	Frame	5' Position (nt)
Primary CD4+ T-cells	NL4-3	A*02:01; B*27:05; B*40:01; C*02:02; C*03:04	**ILINGQYSL**	**ARF**	3253
			AVLSIVNR	*env* CDS	8312
			SESAIRNTL	*vif* CDS	5386
			SESAIRNTIL	*vif* CDS	5386
			KELYPLASL	*gag* CDS	2230
			VLAEAMSQV	*gag* CDS	1873
			GENTSLLHPVSL	*nef* CDS	9261



conserved regions of the virus[51]. Nonetheless, 8 ARFs appeared to be more conserved than their overlapping CDS sequences. We cannot exclude, thus far, that these 8 ARFs might encode polypeptides with functional viral proprieties.

uORF in the 5′UTR of HIV could influence Gag expression. This hypothesis has been addressed in the accompanying manuscript submitted together with Emiliano Ricci. Indeed, using mutagenesis and reporter systems, we show that extensive uORF translation from HIV-1 transcripts conditions the expression of Gag CDS. Other ARFs in the HIV genome might also regulate the expression of the main HIV CDSs. They might also correspond to irrelevant by-standard translation events imposed by the tertiary structure and sequences of HIV

Fig. 5 | MS-based identification of ARF-derived HLA-presented peptides and polyfunctional T cells responses. a List of CDS- and ARF-derived peptides identified in HIV_{NL4-3}-infected primary $CD4^+$ T cells. **b** T cell responses specific to YSL-ARF-derived peptide in PBMCs of PLWH. IFNγ responses of donors EC-1 and EC-3 upon YSL-peptide stimulation, on day 12. POOL (+), YSL-I, and YSL-II correspond to cells loaded with a pool of peptides containing YSL-I and YSL-II peptides, or only with YSL-I and YSL-II peptides, respectively. POOL (−) (white bar): negative control using cells loaded with DMSO-containing medium. Data were presented as mean values of technical triplicates ± SD. Positive response: number of rough $IFNγ^+$ spots per well >20 and two times higher than the negative control. SAT: saturated well. **c** Functionality of YSL-specific $CD4^+$ T cell response from EC-3 donor. Following a 12-day in vitro expansion, ICS was performed for MIP-1β, IFNγ, IL2, CD107a, and TNFα after stimulation with YSL-I (top panel) or an Env-derived peptide (MHC-II_env01, bottom panel), used as a positive control. FACS gatings were performed as in Fig. 4 and as shown in Supplementary Fig. 13. Right panel, the left pie charts illustrate the percentage of activated (yellow) and non-activated (blue) T cells following YSL- or Env-peptide stimulation, upper and lower lines, respectively. The polyfunctional profile of activated YSL- or Env-specific T cells are shown in the right pie charts. Boolean analysis was conducted to determine the percentage of cells secreting 1, 2, 3, 4, or 5 cytokines simultaneously (red, green, purple, blue, and orange, respectively). Polyfunctional profiles of YSL-I- (**d**) and YSL-II-specific (**e**) $CD4^+$ T cell responses. As in (**c**), Left pie charts: percentage of activated (yellow) and non-activated (blue) T cells following peptide stimulation. Right pie charts: polyfunctionality of activated T cells, for all responding donors (codes are indicated on the top). Boolean analysis determined the percentage of cells secreting one or five cytokines simultaneously. The same FACS gating as in (**c**). **f** Heatmap summarizing YSL-I and YSL-II peptide-specific $CD4^+$ T cell responses in all tested PLWH. NA not applicable, NS tested but not significant.

mRNAs. To this regard, we observed a strong accumulation of reads, in the three frames, 30 nucleotides upstream of SD1, which is known to fold into a stable stem-loop structure (Supplementary Fig. 3). Interestingly, this region is also translated from various frames since we could detect T cell responses targeting peptides derived from two different ARFs that overlay this patch of RPF (Fig. 3 and Supplementary Fig. 3). Overall, we believe that the strong patch of reads in this region might be due to several factors including, the presence of SD1 in all viral transcripts, the formation of stable hairpin loops (containing SD1 and the RNA dimerization motif (DIS) upstream of SD1) but also to translation of ARFs.

As mentioned, ARFs might also encode polypeptides with biological functions. Indeed, it is becoming evident that short polypeptides (below 100 aa), that were largely ignored so far, can exert various functions in the development of viral infection. New identified ARFs might as well be seen as a genomic reservoir of unselected sequences allowing the emergence of de novo genes. Indeed, pervasive translation of short peptides derived from presumed non-coding regions might expose these ARF-encoded polypeptides to selection, allowing, from time to time, the passage of ARFs into the world of established, regulated, and selected products[52], as observed for ASP[53].

In infected cells, MHC-I epitopes originate from viral proteins but also from truncated or misfolded viral polypeptides, also called DRiPs (defective ribosomal products)[54]. DRiPs are labile products degraded shortly after translation, allowing rapid loading of MHC-I molecules and thus CTL recognition within minutes after viral infections[55]. Remarkably, DRiPs were initially identified in HIV-infected cells[56]. As discussed above, it is likely that ARF-produced viral polypeptides belong to DRiPs. In the present work, we readily demonstrate that at least 43 ARFs, identified by Riboseq, produce viral polypeptides which can induce T cell responses in PLWH. Some of these ARF-derived peptides are targeted by CTLs. Highlighting the importance of ARFP-specific T cells in the course of HIV infection, others and we previously demonstrated that ARFP-specific CTL recognize infected cells[18–20] and exert a selection pressure on the virus in vivo[18,19,24].

To identify naturally presented ARF-derived HLA ligands in HIV-infected cells, we analyzed the immunopeptidome of two infected $CD4^+$ T cell lines. Unfortunately, we could not identify any ARF-derived HLA-presented peptide. These might be due to several factors including the sensitivity of shotgun mass spectrometric discovery approaches. Indeed, even in the context of the immense technical improvements in the last decades, it remains a challenge to identify low-abundance peptides with high turnover rates. In addition, the immunopeptidome is a highly dynamic, rich, and complex assembly of peptides, which is shaped by several factors, including accessibility to the antigen processing and presentation machinery e.g., specificities of proteasomes and cellular protases or of individual HLA allotypes[57]. Nonetheless, our analysis of the immunopeptidomics data, published in our previous work[58], coming from infected primary $CD4^+$ T cells allowed us to identify one ARF-derived peptide, the HLA-A*02:01-restricted peptide ILINGQYSL, thus providing a direct demonstration that ARF-derived peptides are naturally presented by HLA molecules. Probably due to the low number of HIV-infected samples that we used in our study, we did not detect CTL responses targeting this peptide. However, $CD4^+$ T cell responses against this peptide or an elongated version of it was observed in five PLWH. Therefore, our results suggest that alternative translation events may give rise to epitopes recognized independently or concomitantly by $CD4^+$ and $CD8^+$ T cells and presented on different HLA class I and HLA class II alleles. Overall, we observed that the majority of T cell responses targeting ARF-derived peptides were mediated by $CD4^+$ T cells.

As a matter of fact, HIV-specific $CD4^+$ T cells play an important role in HIV infection. The breadth and specificity of HIV-specific $CD4^+$ T cell responses are associated with improved viral control and low viremia during acute and chronic infection, respectively[59,60]. In particular, the generation of Gag-specific $CD4^+$ T cells and their maintenance correlates with a better disease outcome and viral control, respectively[59,60]. There is also evidence of an association between certain HLA-DRß1 alleles and Gag-specific $CD4^+$ T cell activity and delayed disease progression[61]. HIV-specific $CD4^+$ T cells can also exert direct antiviral functions by eliminating infected cells and/or inhibiting viral replication[62]. Remarkably, others and we have shown that $CD4^+$ T cells can recognize peptides derived from newly-synthetized, so-called endogenous antigens in virus-infected cells, including HIV-infected cells[63,64] or in tumor cells[65]. In fact, depending on the subcellular localization, the trafficking, and the nature of the antigen itself, different pathways are involved in the degradation and presentation of endogenous antigens by MHC-II molecules. This includes components of the MHC-I processing pathway, such as proteasome and TAP[66,67], autophagy[68] and receptors regulating vesicular trafficking[69]. In the field of cancer, it was recently reported that $CD4^+$ T cells also recognize peptides derived from unannotated CDSs and neoantigens[70–73].

To get some insights on the quality of ARFP-specific T cell responses, we analyzed the polyfunctional profiles and the pattern of cytokine secretions of these CTL and $CD4^+$ T cell responses in PBMCs of PLWH. We show here that $CD4^+$ and $CD8^+$ T cell responses against ARFP exhibit a polyfunctional profile with 80% of cells secreting at least three cytokines simultaneously. The pattern of secretion does not differ between previously described T cell responses targeting immunodominant epitopes and ARFP-specific T cell responses, suggesting that the latter are as potent as classical T cell responses in controlling viral replication.

Although using a small cohort of PLWH with suppressed viral replication, we were intrigued by the large heterogeneity of ARF-specific T cell responses. Several parameters might influence the magnitude and the breadth of HIV-specific T cell responses, including the individual HLA allotype and T cell repertoire, the senescence of T cells, the antiretroviral treatment, and the size of the active viral

reservoir, defined as cells carrying a transcriptionally active viral genome[74]. Remarkably, the later has been recently linked to the magnitude and functions of HIV-specific CD4$^+$ and CD8$^+$ T cells[75,76]. Interestingly, this active reservoir is mainly composed of defective viral genomes[75], which might favor the expression of alternative reading frames and, thus, the activation of ARF-derived peptide-specific T cells. In the future, it will be of interest to study ARF-specific T cell responses with regards to the active reservoir and the emergence of new dominant T cell clonotypes after prolonged ART treatment[77].

In conclusion, using ribosome profiling, we defined here the translatome of HIV in infected CD4$^+$ T cells. We demonstrated that the HIV genome harbors ARFs that can be translated into viral polypeptides used by the immune system to initiate potent T cell responses. Both in cancer and in HIV-infected cells, ARF-derived polypeptides represent promising targets to promote the control of tumor growth or viral replication. Understanding how the translation of non-canonical genomic regions is regulated and its involvement in various cellular or viral processes will certainly help in fighting cancer and infections.

Methods

Virus and infection

VSV-G pseudotyped $HIV_{NL4\text{-}3\ XCS}$ and $HIV_{NL4\text{-}3\Delta Nef}$ were produced by transfection of 293T cells using a CAPHOS kit (Sigma-Aldrich) as in ref. 63. Viruses were pseudotyped to favor viral entry and to achieve a high infection rate within a short time. SupT1 CD4$^+$ T cells were cultured with RPMI GlutaMax 1640 (Gibco) complemented with 10% FBS (Dutscher) and 1% penicillin/streptomycin. About 1×10^9 SupT1 CD4$^+$ T cells were infected with 4055 ng of VSV-G-$HIV_{NL4\text{-}3\ XCS}$ HIV-Gag-p24 for 3 h in IMDM plus 10 mM HEPES (Gibco/Thermo Fisher Scientific) supplemented with 2 μg/ml of DEAE-dextran (Sigma). Three biological replicates were performed corresponding to 300 million cells infected at day 0. Twenty hours post-infection a fraction of the cells (2×10^5) was harvested, stained with CD4-Vio770 antibody (Miltenyi) and a Live/Dead (LVD) fixable violet dye (Invitrogen), fixed with 4% paraformaldehyde (ChemCruz), and permeabilized (PBS1X, BSA 0,5%, saponin 0,05%). Cells were then stained with an HIV-1 Gag-p24 specific antibody (KC57-PE, Beckman Coulter). Sample acquisition and analysis were performed using a BD LSRFortessa and FlowJo v10.8 Software, respectively (both from BD Life Sciences). Mock-infected cells were used as a negative control. Twenty-four hours post-infection, cycloheximide (CHX) was then added to the remaining cell cultures (100 μg/ml, 10 min, 37 °C, Sigma). Cells were washed in cold PBS containing CHX and lysed at 4 °C (10 mM Tris-HCl pH 7.5, 100 mM KCl, 10 mM Mg2+ acetate, 1% Triton X-100 and 2 mM DTT). Glass beads were then added to the cell supernatant and incubated on an orbital shaker (5 min, 4 °C). Beads were removed by centrifugation and the supernatant was quickly frozen in liquid nitrogen and stored at −80 °C. To identify the start codons, prior harvest, cells were treated with Lactimidomycin (LTM; 5 μM, 30 min, 37 °C) together with Puromycin (PMY, 25 μM) for the last 20 min. Cells were lysed in abovementioned lysis buffer containing LTM and PMY.

Ribosome profiling

Cell lysates were gently thawed on ice in the presence of an EDTA-free complete protease inhibitor cocktail (Roche) and 200 U of RNase Murine Inhibitor (NEB) were added. The absorbance of the crude extract obtained was measured at 260 nm. The extracts were digested using 5U/UA_{260nm} RNase I (Ambion) for 1 h at 25 °C. The digestion was stopped by adding a SUPErasin RNase inhibitor (500U, Invitrogen). Monosomes were then loaded on a 24% sucrose cushion and ultracentrifugated at 543,000×*g* on a TLA110 rotor at +4 °C. The concentrated monosomes were resuspended in 600 μl of lysis buffer. RNA were extracted by acid phenol at 65 °C, chloroform, and precipitated by ethanol with 0.3 M sodium acetate pH 5.2. Resuspended RNAs were loaded on 17% polyacrylamide (19:1) gel with 7 M urea and run in 1xTAE buffer for 6 h at 100 V. RNA fragments corresponding to 28–34nt were retrieved from gel and precipitated in ethanol with 0.3 M sodium acetate pH 5.2 in presence of 100 mg glycogen. rRNA were depleted using the Ribo-Zero Gold rRNA Removal kit (Illumina). The supernatants containing the ribosome footprints were recovered, and RNA were precipitated in ethanol in the presence of glycogen overnight at −20 °C. The RNA concentration was measured by Quant-iT microRNA assay kit (Invitrogen), and the RNA integrity and quality was verified using Bioanalyzer small RNA Analysis kit (Agilent). cDNA library from 100 ng RNA was prepared by the High-throughput sequencing facility of I2BC, using the NebNext Small RNA Sample Prep kit with 3′ sRNA Adapter (Illumina) according to the manufacturer's protocol with 12 cycles of PCR amplification in the last step followed by DNA purification with AMPpure XP beads cleanup. Library molarity was analyzed using a Bioanalyzer DNA Analysis kit (Agilent) and an equimolar pool of libraries was sequenced with NexTSeq 500/550 High output kit v2 (75 cycles) (Illumina) with 10% PhiX.

Bioinformatical analysis

The ribosome profiling analysis was made using the RiboDoc tool (v0.9.0)[29]. The different main steps with corresponding programs, versions and command lines used in its analysis are described below. The reference genomes are Homo_sapiens.GRCh38.104 for humans (from the Ensembl database[78]) and HIV-1-pNL4-3 XCS for HIV.

The sequencing adapters were trimmed by cutadapt v4.3[79] and the lengths of the RPFs was filter to keep reads from 25 to 35 nucleotides long as there expected length is around 30 nucleotides with the following parameters:

cutadapt -e 0.125 --max-n = 1 -m 25 -M 35 -a ${adapter_sequence} -o ${output.fastq} ${input.fastq}

The removal of the rRNA reads was made by an alignment on the rRNA sequences by bowtie2 v2.5.1[80]:

bowtie2 -x ${index.rRNA} -U ${input.fastq} --un-gz ${output.fastq}

The alignment on the genome was made with both hisat2 v2.2.1[81] and bowtie2 v2.5.1:

hisat2 -x ${hisat2_index.genome} --no-softclip -U ${input.fastq} --un-gz ${output.fastq} -S ${output.sam_hisat2}

bowtie2 -x ${bowtie2_index.genome} --end-to-end -U ${output.fastq} -S ${output.sam_bowtie2}

The selection of the reads uniquely mapped on the genome was made with samtools v1.14[82]:

samtools view -F 3844 -q 1 -h -o ${output.bam}

The counting of the reads corresponding to each transcript was done by htseqcount v2.0.2[83]:

htseq-count -f bam -t CDS -i Parent --additional-attr Name -m intersection-strict --nonunique fraction ${input.bam} ${input.gff} > ${output.txt}

The qualitative analysis was made on a transcriptome made from the genome with a selection of transcripts annotated as having a 5′UTR region only (for the human genome). This analysis and the determination of the P-site offset for each read length of every sample was made by the riboWaltz v1.2.0 package[84].

To study the reading frame of the ribosome-protected fragments (RPF), each read was represented by the coordinate corresponding to the first base of the associated ribosome's P-site. To determine where the P-site is, a P-site offset has to be defined for every read length. This step was done with the riboWaltz program[84] which looks for the first base of each read beginning of the signal.

Data sets and multiple alignments

Sequences were downloaded from the Los Alamos HIV Sequence Database (HIV-1 clades B and C) (https://www.hiv.lanl.gov/content/index). Sequences were filtered based on the presence of only one sequence per patient and excluding sequences carrying premature

stop codons within CDS. Very similar sequences and incomplete sequences (more than 10 gaps/unknowns) were additionally discarded. We obtained 1609 and 412 sequences of HIV-1 clade B and clade C, respectively. The aa sequence conservation of the overlapping CDS and third frame were analyzed using the Unipro UGENE toolkit. Multiple alignments were performed using ClustalO. Statistical analysis was performed with R (https://www.R-project.org/, version 4.2.2) using Rstudio IDE (2023.03.01.554). Comparison of population distributions were performed using Kruskal–Wallis test, Wilcoxon, and Dunn tests. Plots were generated using the ggplot2 package (v3.4.2). For each aa sequence conservation analysis of ARFs within HIV-1 clades B and C, the number of sequences used ranged from 168 to 1609 and from 47 to 411 sequences, respectively. Taking into consideration that in the UTR regions, the sequences may contain microdeletion and/or truncation, another pipeline was used for the ARF overlapping the UTRs. For each nucleotide sequence, the six phases were translated. Then, ARF aa sequences were aligned recursively to the six translated sequences. The best alignments were kept, and alignments too far (±150 nt) from the putative position were eliminated. Multiple alignments were performed using Clustal Omega.

Immunopeptidome

Isolation of HLA ligands. About 5×10^8 C8166 cells were infected with $HIV_{NL4\text{-}3\Delta Nef}$ at an MOI of 0.05 and snap frozen 22 h post-infection. HLA class I and HLA class II molecules were isolated using standard immunoaffinity purification[85], using the pan-HLA class I-specific W6/32, the pan-HLA class II-specific Tü-39, and the HLA-DR-specific L243 monoclonal antibodies (produced in-house) covalently linked to CNBr-activated Sepharose (Sigma-Aldrich). Cells were lysed in lysis buffer (CHAPS (Panreac AppliChem), complete protease inhibitor cocktail tablet (Roche) in PBS) for 1 h on a shaker at 4 °C, sonicated, and centrifuged (45 min, 4000×*g*) and incubated again for 1 h. Lysates were cleared by sterile filtration (5 μm filter unit (Merck Millipore)) and cyclically passed through a column-based setup overnight at 4 °C. Columns were washed with PBS (30 min) and ddH_2O (1 h). Peptides were eluted by 0.2% trifluoroacetic acid (TFA), isolated by ultrafiltration (Amicon filter units (Merck Millipore)), lyophilized, and desalted using ZipTip pipette tips with C18 resin (Merck).

Mass spectrometric data acquisition. For the mass spectrometric analysis peptides were loaded on a 75 μm × 2 cm PepMap nanotrap column (Thermo Fisher) at a flow rate of 4 μl/min for 10 min[85]. Subsequent separation was performed by nanoflow high-performance liquid chromatography (RSLCnano, Thermo Fisher) using a 50 μm × 25 cm PepMap rapid separation column (Thermo Fisher, particle size of 2 μm) and a linear gradient ranging from 2.4 to 32.0% acetonitrile at a flow rate of 0.3 μl/min over the course of 90 min. Eluting peptides were analysed in technical replicates in an online-coupled Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher) equipped with a nanoelectron spray ion source using a data-dependent acquisition mode employing a top-speed collisional-induced dissociation (CID, normalized collision energy 35%, HLA class I peptides) or higher-energy collisional dissociation (HCD, normalized collision energy 30%, HLA class II peptides) fragmentation method. MS1 and MS2 spectra were detected in the Orbitrap with a resolution of 120,000 and 30,000, respectively. The maximum injection time was set to 50 and 150 ms for MS1 and MS2, respectively. The dynamic exclusion was set to 7 and 10 s for HLA class I and HLA class II, respectively. The mass range for HLA class I peptide analysis was set to 400–650 m/z with charge states 2+ and 3+ selected for fragmentation. For HLA class II peptide analysis mass range was limited to 400 1000 m/z with charge states 2+ to 5+ selected for fragmentation.

Data processing. For data processing, the SEQUEST HT search engine (University of Washington[86]) was used to search the human proteome as comprised in the Swiss-Prot database (20,394 reviewed protein sequences, December 4 2020) supplemented with the six-frame translated genome of $HIV_{NL4\text{-}3}$, taking into account all potential peptides ≥8 aa encompassed between two stop codons without enzymatic restriction. Precursor mass tolerance was set to 5 ppm, and fragment mass tolerance to 0.02 Da. Oxidized methionine was allowed as a dynamic modification. The peptide spectrum matches (PSM) false discovery rate (FDR) was estimated using the Percolator algorithm[87] and limited to 5% for HLA class I and 1% for HLA class II. Peptide lengths were limited to 8–12 aa for HLA class I and to 8–25 aa for HLA class II. Protein inference was disabled, allowing for multiple protein annotations of peptides. HLA class I annotation was performed using NetMHCpan-4.1[88] and SYFPEITHI[89] annotating peptides with percentile rank below 2% and ≥60% of the maximal score, respectively.

Spectrum validation. Spectrum validation of the experimentally eluted peptide was performed by computing the similarity of the spectra with the corresponding isotope-labeled synthetic peptides measured in a complex matrix. The spectral correlation was calculated between the MS/MS spectra of the eluted and the synthetic peptide[90].

Participants and samples

Elite controllers (EC, $n = 4$) were recruited from the CO21 CODEX cohort implemented by the ANRS | MIE (Agence Nationale de recherches sur le SIDA, les hépatites virales|Maladies infectieuses émergentes). 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 treatment for more than 5 years. Efficiently treated PLWH (ART) ($n = 3$) were recruited at AP-HP Bicêtre Hospital (Kremlin Bicêtre, France). They are treated according to the standard of care for at least 1 year (mean of 10 years) and have undetectable viral loads using standard assays. PBMCs from anonymous HIV-negative blood donors ($n = 3$) were purchased from EFS (Établissement Français du sang) under the agreement number 15/EFS/022. A detailed description of the donors is provided in Supplementary Table 4, including the interquartile range for age (at the time of the study), the sex, the number of years of infection and of treatment, the drug regimen, the mean CD4 T cell counts, RNA loads and the HLA haplotypes. HIV-RNA loads were measured on-site with different real-time PCR-based assays; depending on the date of enrollment in the cohort and the assay routinely used on each site, the VL detection limit varied from 50 to 10 copies/mL.

Ethic statement

All the 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 | MIE and approved by the Ile de France VII Ethics Committee. The study was conducted according to the principles expressed in the Declaration of Helsinki.

Peptides

ARF-derived peptides were chosen based on the aa conservation among the clade B isolates and the predictions to bind frequent HLA alleles. The binding affinity of ARF-derived peptides was evaluated using the NetMHCPan-4.1 and NetMHCIIPan-4.1[88]. For the longest ARF sequences, some overlapping peptides were also designed. The synthetic peptides were then randomly distributed in 9 different pools to obtain 11 or less peptides per pool to limit peptide competition for HLA binding (Supplementary Table 3). HIV-classical epitopes were chosen based on the literature[37] (Supplementary Table 5). ARF-derived peptides and HIV-classical peptides were synthetized by Vivitide company and in-house at the Department of Immunology, University of Tübingen, Germany. The in-house production was performed with the peptide synthesizer Liberty Blue (CEM) using the 9 fluorenylmethyloxycarbonyl/tert-butyl strategy[91]. Non-HIV common peptides (from

HCMV, EBV, and influenza virus) were obtained from Mabtech (PepPool: CEF (CD8), human, 3616-1).

12-day in vitro T cell amplification prior to ELISPOT assay

PBMCs were thawed and rested 2–3 h in IMDM (Gibco/Thermo Fisher Scientific) containing 5% human AB serum (SAB, Institut Jacques Boy), supplemented with recombinant human IL2 (rhIL-2, 10 U/ml, Miltenyi) and Dnase I (1 U/ml, New England Biolabs). Cells were washed, and $5–9 \times 10^6$ PBMCs were then seeded/well in a 24-well plate in IMDM supplemented with 10% SAB, Pen/Strep, nonessential aa, and sodium pyruvate (all from Gibco/Thermo Fisher Scientific). Nevirapine (NVP, 1.2 μM, HIV reagent program) was added to inhibit potential viral replication and poly I:C (2 μg/ml, InvivoGen) to facilitate the presentation of long peptides by antigen-presenting cells[92]. PBMCs were loaded with peptide pools (each peptide at 10 μg/ml), except for the CEF pool (Mabtech) used at 5 μg/ml, and cultured overnight at 37 °C, 5% CO_2. HIV-classical peptides and/or CEF peptide pools were used as positive control for the expansion of HIV-specific and non-HIV common anti-virus-specific T cells, respectively. On days 1 and 3, T cell media were then complemented with rhIL-2 (10 U/ml) and rhIL-7 (20 ng/ml, Miltenyi) with NVP (1.2 μM), respectively, and maintained throughout the culture. On days 3, 5, 7, and 9, when the cell layer was >70% confluent, cells were split into two wells before the addition of rhIL-2, rhIL-7, and NVP. On day 7, cells were harvested, counted, and a fraction was submitted to IFN-γ ELISPOT assay using the ARF-derived peptide pools and positive controls: HIV-classical peptide and/or CEF peptide pools. The remaining cells were maintained in culture and submitted, on day 12, to IFN-γ ELISPOT assay using individual ARF-derived peptides and the controls.

Enzyme-linked immunospot assay—cultured IFN-γ ELISPOT

About $1–3 \times 10^5$ cells/well were seeded in ELISPOT plates (MSIPN4550, Millipore) in IMDM supplemented with 10% SAB, Pen/Strep, nonessential aa, sodium pyruvate 10%, and HEPES 10 mM (Gibco/Thermo Fisher Scientific), loaded with either 50, 10, or 2 μg/ml of ARF-derived peptides, HIV-classical peptides 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 was revealed using an anti-IFNγ secondary antibody conjugated with biotin (both from Mabtech) as described[24]. 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 solubilize the peptides).

Intracellular cytokine staining (ICS) assay

Cells were either cultured as described for the 12-day in vitro T cell amplification or treated directly after thawing (for the ex vivo assays). On day 12, cells were harvested, washed, and counted. About $2–10 \times 10^5$ PBMCs were then seeded/well of a 96-well U-bottom plate in IMDM supplemented with 10% SAB, Pen/Strep, nonessential aa, sodium pyruvate 10%, and HEPES 10 mM. PBMCs were immediately loaded with individual ARF-derived peptides (50 μg/ml), individual HIV- classical peptides (10 μg/ml), or CEF peptide pool (2 ug/ml). Additionally, the CD107a-BV786 antibody (BD Biosciences) was included in each condition for degranulation detection. For the ex vivo assessment of T cell activation, $2–4 \times 10^6$ PBMCs were seeded per well of a 96-well U-bottom plate using the same IMDM-supplemented medium. The same peptide loading and CD107a-BV786 antibody addition steps were performed. After 1 h at 37 °C, brefeldin A was added (5 μg/ml), and the incubation was carried out for an additional 5 h. PBMCs were washed in PBS and incubated with LVD and human Fc Block (BD Biosciences). Cells were then stained for 30 min on ice with a mix of cell surface antibodies: CD3-PE, CD8-SuperBright600 (Invitrogen), and CD4-AF750 (Beckman Coulter). Next, cells were washed, fixed with PFA 4%, and permeabilized (PBS + BSA 0.5%, saponin 0,05%). Cytokine production were detected by intracellular staining using IFN-γ-PerCP-Cy-5.5, MIP-1β-FITC, TNFα-PE-Cy7, and IL2-APC antibodies (all from BD Biosciences).

Flow cytometry and data analysis

Sample acquisition and analysis were performed using Cytoflex with CytExpert software (Beckman Coulter). The compensation matrix was calculated automatically by the CytExpert software after measuring negative and single-color controls. Sample analysis was performed using FlowJo v10.8 Software (BD Life Sciences). After the elimination of doublets, lymphocytes were gated on the side scatter area versus the forward scatter area pseudo-color dot plot, and dead cells were removed according to LVD staining. $CD4^+CD3^+$ events were gated versus individual gates made for the detection of MIP-1β, IFNγ, IL2, CD107a, and TNFα secretion. All gates were set based on the unstimulated control (cells loaded with the DMSO-containing medium used to solubilize the peptides). Individual cytokine secretions were then combined together using Boolean gating to create a full array of possible combinations of response patterns from the $CD4^+CD3^+$-cell subsets. The same procedure was used to $CD8^+CD3^+$ events. Positive responses were reported after background subtraction. Samples were considered positive when the % of activated cells was at least two times higher than negative controls.

Statistical analysis

Statistical significances (*p* values) were calculated using Prism Software (GraphPad). Statistical tests used for each individual figure are indicated in the figure legends.

Reporting summary

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

Data availability

The mass spectrometry immunopeptidomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE (Perez-Riverol et al, 2019) partner repository with the dataset identifier: Project accession: PXD043984. The Riboseq data have been deposited to the public functional genomics data repository GEO: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE239818 Source data are provided with this paper.

Code availability

The code package for this study is available and fully described in the above section “Bioinformatical analysis”.

References

1. Andrews, S. J. & Rothnagel, J. A. Emerging evidence for functional peptides encoded by short open reading frames. *Nat. Rev. Mater.* **15**, 193–204 (2014).
2. Martinez, T. F. et al. Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* **16**, 458–468 (2020).
3. Ji, Z., Song, R., Regev, A. & Struhl, K. Many lncRNAs, 5′UTRs, and pseudogenes are translated and some are likely to express functional proteins. *eLife* **4**, e08890 (2015).
4. Schlesinger, D. & Elsässer, S. J. Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins. *FEBS J.* **289**, 53–74 (2022).
5. Yewdell, J. W. MHC class I immunopeptidome: past, present, and future. *Mol. Cell. Proteomics* **21**, 100230 (2022).

6. Chong, C., Coukos, G. & Bassani-Sternberg, M. Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* **40**, 175–188 (2022).
7. Nelde, A. et al. Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens. *Cell. Mol. Life Sci.* **79**, 171 (2022).
8. Ouspenskaia, T. et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).
9. Ruiz Cuevas, M. V. et al. Most non-canonical proteins uniquely populate the proteome or immunopeptidome. *Cell Rep.* **34**, 108815 (2021).
10. Bartok, O. et al. Anti-tumour immunity induces aberrant peptide presentation in melanoma. *Nature* **590**, 332–337 (2021).
11. Smith, C. C. et al. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *J. Clin. Invest.* **128**, 4804–4820 (2018).
12. Starck, S. R. et al. Translation from the 5′ untranslated region shapes the integrated stress response. *Science* **351**, aad3867 (2016).
13. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93**, 1385–1409 (2012).
14. Bullock, T. N. & Eisenlohr, L. C. Ribosomal scanning past the primary initiation codon as a mechanism for expression of CTL epitopes encoded in alternative reading frames. *J. Exp. Med.* **184**, 1319–1329 (1996).
15. Yang, N. et al. Defining viral defective ribosomal products: standard and alternative translation initiation events generate a common peptide from influenza A virus M2 and M1 mRNAs. *J. Immunol.* **196**, 3608–3617 (2016).
16. Zanker, D. J. et al. Influenza A virus infection induces viral and cellular defective ribosomal products encoded by alternative reading frames. *J. Immunol.* **202**, 3370–3380 (2019).
17. Gaur, A. & Green, W. R. Role of a cytotoxic-T-lymphocyte epitope-defined, alternative gag open reading frame in the pathogenesis of a murine retrovirus-induced immunodeficiency syndrome. *J. Virol.* **79**, 4308–4315 (2005).
18. Bansal, A. et al. CD8 T cell response and evolutionary pressure to HIV-1 cryptic epitopes derived from antisense transcription. *J. Exp. Med.* **207**, 51–59 (2010).
19. 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).
20. 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).
21. Erhard, F. et al. Improved Ribo-seq enables identification of cryptic translation events. *Nat. Methods* **15**, 363–366 (2018).
22. Weingarten-Gabbay, S. et al. Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell* **184**, 3962–3980.e17 (2021).
23. Bansal, A. et al. Enhanced recognition of HIV-1 cryptic epitopes restricted by HLA class I alleles associated with a favorable clinical outcome. *J. Acquir. Immune Defic. Syndr.* **70**, 1–8 (2015).
24. Cardinaud, S. et al. CTL escape mediated by proteasomal destruction of an HIV-1 cryptic epitope. *PLoS Pathog.* **7**, e1002049 (2011).
25. Champiat, S. et al. Influence of HAART on alternative reading frame immune responses over the course of HIV-1 infection. *PLoS ONE* **7**, e39311 (2012).
26. Garrison, K. E. et al. Transcriptional errors in human immunodeficiency virus type 1 generate targets for T-cell responses. *Clin. Vaccine Immunol.* **16**, 1369–1371 (2009).
27. Maness, N. J. et al. AIDS virus specific CD8+ T lymphocytes against an immunodominant cryptic epitope select for viral escape. *J. Exp. Med.* **204**, 2505–2512 (2007).
28. Wildum, S., Schindler, M., Münch, J. & Kirchhoff, F. Contribution of Vpu, Env, and Nef to CD4 down-modulation and resistance of human immunodeficiency virus type 1-infected T cells to superinfection. *J. Virol.* **80**, 8047–8059 (2006).
29. François, P., Arbes, H., Demais, S., Baudin-Baillieu, A. & Namy, O. RiboDoc: a docker-based package for ribosome profiling analysis. *Comput. Struct. Biotechnol. J.* **19**, 2851–2860 (2021).
30. 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. *Nature* **441**, 244–247 (2006).
31. Bencun, M. et al. Translational profiling of B cells infected with the Epstein-Barr virus reveals 5′ leader ribosome recruitment through upstream open reading frames. *Nucleic Acids Res.* **46**, 2802–2819 (2018).
32. Finkel, Y. et al. The coding capacity of SARS-CoV-2. *Nature* **589**, 125–130 (2021).
33. Whisnant, A. W. et al. Integrative functional genomics decodes herpes simplex virus 1. *Nat. Commun.* **11**, 2038 (2020).
34. Yang, Z. et al. Deciphering poxvirus gene expression by RNA sequencing and ribosome profiling. *J. Virol.* **89**, 6874–6886 (2015).
35. Gao, X. et al. Quantitative profiling of initiating ribosomes in vivo. *Nat. Methods* **12**, 147–153 (2015).
36. Curlin, M. E. et al. HIV-1 envelope subregion length variation during disease progression. *PLoS Pathog.* **6**, e1001228 (2010).
37. *HIV Molecular Immunology* (Theoretical Biology and Biophysics Group T-10, Los Alamos National Laboratory, 2022).
38. Sauce, D., Elbim, C. & Appay, V. Monitoring cellular immune markers in HIV infection: from activation to exhaustion. *Curr. Opin. HIV AIDS* **8**, 125–131 (2013).
39. Larsen, M. et al. Evaluating cellular polyfunctionality with a novel polyfunctionality index. *PLoS ONE* **7**, e42403 (2012).
40. Vingert, B. et al. HIV controllers maintain a population of highly efficient Th1 effector cells in contrast to patients treated in the long term. *J. Virol.* **86**, 10661–10674 (2012).
41. Schwartz, O., Maréchal, V., Gall, S. L., Lemonnier, F. & Heard, J.-M. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV–1 Nef protein. *Nat. Med.* **2**, 338–342 (1996).
42. Ziegler, M. C. et al. HIV-1 induced changes in HLA-C∗03:04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS* **34**, 1713–1723 (2020).
43. Ingolia, N. T., Hussmann, J. A. & Weissman, J. S. Ribosome profiling: global views of translation. *Cold Spring Harb. Perspect. Biol.* **11**, a032698 (2019).
44. Zhang, J. & Crumpacker, C. HIV UTR, LTR, and epigenetic immunity. *Viruses* **14**, 1084 (2022).
45. Arias, C. et al. KSHV 2.0: a comprehensive annotation of the Kaposi's sarcoma-associated herpesvirus genome using next-generation sequencing reveals novel genomic and functional features. *PLoS Pathog.* **10**, e1003847 (2014).
46. Orr, M. W., Mao, Y., Storz, G. & Qian, S.-B. Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.* **48**, 1029–1042 (2020).
47. Bet, A. et al. The HIV-1 antisense protein (ASP) induces CD8 T cell responses during chronic infection. *Retrovirology* **12**, 15 (2015).
48. Vanhée-Brossollet, C. et al. A natural antisense RNA derived from the HIV-1 env gene encodes a protein which is recognized by circulating antibodies of HIV+ individuals. *Virology* **206**, 196–202 (1995).
49. Laverdure, S. et al. HIV-1 antisense transcription is preferentially activated in primary monocyte-derived cells. *J. Virol.* **86**, 13785–13789 (2012).
50. Finkel, Y., Stern-Ginossar, N. & Schwartz, M. Viral short ORFs and their possible functions. *Proteomics* **18**, 1700255 (2018).
51. Jacobo-Molina, A. et al. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-

stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl Acad. Sci. USA* **90**, 6320–6324 (1993).
52. Ruiz-Orera, J., Verdaguer-Grau, P., Villanueva-Cañas, J. L., Messeguer, X. & Albà, M. M. Translation of neutrally evolving peptides provides a basis for de novo gene evolution. *Nat. Ecol. Evol.* **2**, 890–896 (2018).
53. Cassan, E., Arigon-Chifolleau, A.-M., Mesnard, J.-M., Gross, A. & Gascuel, O. Concomitant emergence of the antisense protein gene of HIV-1 and of the pandemic. *Proc. Natl Acad. Sci. USA* **113**, 11537–11542 (2016).
54. Yewdell, J. W., Antón, L. C. & Bennink, J. R. Defective ribosomal products (DRiPs): a major source of antigenic peptides for MHC class I molecules? *J. Immunol.* **157**, 1823–1826 (1996).
55. Dolan, B. P. et al. Distinct pathways generate peptides from defective ribosomal products for CD8+ T cell immunosurveillance. *J. Immunol.* **186**, 2065–2072 (2011).
56. Schubert, U. et al. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404**, 770–774 (2000).
57. Bassani-Sternberg, M. & Coukos, G. Mass spectrometry-based antigen discovery for cancer immunotherapy. *Curr. Opin. Immunol.* **41**, 9–17 (2016).
58. Ziegler, M. C. et al. HIV-1 induced changes in HLA-C*03: 04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS* **34**, 1713–1723 (2020).
59. Ranasinghe, S. et al. HIV-specific CD4 T cell responses to different viral proteins have discordant associations with viral load and clinical outcome. *J. Virol.* **86**, 277–283 (2012).
60. Schieffer, M. et al. Induction of Gag-specific CD4 T cell responses during acute HIV infection is associated with improved viral control. *J. Virol.* **88**, 7357–7366 (2014).
61. Ranasinghe, S. et al. Association of HLA-DRB1-restricted $CD4^+$ T cell responses with HIV immune control. *Nat. Med.* **19**, 930–933 (2013).
62. Zheng, N., Fujiwara, M., Ueno, T., Oka, S. & Takiguchi, M. Strong ability of Nef-specific CD4+ cytotoxic T cells to suppress human immunodeficiency virus type 1 (HIV-1) replication in HIV-1-infected CD4+ T cells and macrophages. *J. Virol.* **83**, 7668–7677 (2009).
63. 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).
64. Veerappan Ganesan, A. P. & Eisenlohr, L. C. The elucidation of non-classical MHC class II antigen processing through the study of viral antigens. *Curr. Opin. Virol.* **22**, 71–76 (2017).
65. Fonteneau, J. F., Brilot, F., Münz, C. & Gannagé, M. The tumor antigen NY-ESO-1 mediates direct recognition of melanoma cells by CD4+ T cells after intercellular antigen transfer. *J. Immunol.* **196**, 64–71 (2016).
66. Lich, J. D., Elliott, J. F. & Blum, J. S. Cytoplasmic processing is a prerequisite for presentation of an endogenous antigen by major histocompatibility complex class II proteins. *J. Exp. Med.* **191**, 1513–1524 (2000).
67. Tewari, M. K., Sinnathamby, G., Rajagopal, D. & Eisenlohr, L. C. A cytosolic pathway for MHC class II–restricted antigen processing that is proteasome and TAP dependent. *Nat. Immunol.* **6**, 287–294 (2005).
68. Dengjel, J. et al. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. *Proc. Natl Acad. Sci. USA* **102**, 7922–7927 (2005).
69. Sarango, G. et al. The autophagy receptor TAX1BP1 (T6BP) improves antigen presentation by MHC-II molecules. *EMBO Rep.* **23**, e55470 (2022).
70. Cachot, A. et al. Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. *Sci. Adv.* **7**, eabe3348 (2021).
71. Hu, Z. et al. Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat. Med.* **27**, 515–525 (2021).
72. Kreiter, S. et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **520**, 692–696 (2015).
73. Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
74. Baxter, A. E., O'Doherty, U. & Kaufmann, D. E. Beyond the replication-competent HIV reservoir: transcription and translation-competent reservoirs. *Retrovirology* **15**, 18 (2018).
75. Dubé, M. et al. Spontaneous HIV expression during suppressive ART is associated with the magnitude and function of HIV-specific CD4+ and CD8+ T cells. *Cell Host Microbe* **31**, 1507–1522.e5 (2023).
76. Takata, H. et al. An active HIV reservoir during ART is associated with maintenance of HIV-specific CD8+ T cell magnitude and short-lived differentiation status. *Cell Host Microbe* **31**, 1494–1506.e4 (2023).
77. White, E. et al. Clonal succession after prolonged antiretroviral therapy rejuvenates CD8+ T cell responses against HIV-1. *Nat. Immunol.* **25**, 1555–1564 (2024).
78. Yates, A. D. et al. Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
79. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10–12 (2011).
80. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
81. 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).
82. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
83. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
84. Lauria, F. et al. riboWaltz: optimization of ribosome P-site positioning in ribosome profiling data. *PLoS Comput. Biol.* **14**, e1006169 (2018).
85. Nelde, A., Kowalewski, D. J. & Stevanović, S. Purification and identification of naturally presented MHC class I and II ligands. *Methods Mol. Biol.* **1988**, 123–136 (2019).
86. Eng, J. K., McCormack, A. L. & Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 (1994).
87. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **4**, 923–925 (2007).
88. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).
89. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanović, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* **50**, 213–219 (1999).
90. Toprak, U. H. et al. Conserved peptide fragmentation as a benchmarking tool for mass spectrometers and a discriminating feature for targeted proteomics. *Mol. Cell. Proteomics* **13**, 2056–2071 (2014).
91. Sturm, T. et al. Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nat. Commun.* **4**, 1616 (2013).
92. Schuhmacher, J. et al. Simultaneous identification of functional antigen-specific CD8+ and CD4+ cells after in vitro expansion using elongated peptides. *Cells* **11**, 3451 (2022).

Acknowledgements

This work was granted by the Agence Nationale de Recherches sur le SIDA (ANRS-Maladies infectieuses émergentes, A.M. and O.N.),

Sidaction for fundings (A.M. and O.N) and Agence Nationale de la Recherche (ANR-22-CE12-0041-02, O.N.). L.B., A.B., and E.L. were supported by ANRS. L.B. and M.P. were supported by Sidaction. G.B. was supported by ANR-20-IDEES-0002. We thank Remy Villette for his expertise using R, Bernard Maillere and his team for the access to the ELISPOT reader, Anne Lopes, Paul Roginski, for discussions, Frederic Suba and Clemence Richetta for access to the L3 facility of ENS-Paris-Saclay. We thank all participants of the ANRS CODEX cohort and the NIH AIDS Research and Reference Reagent Program for providing drugs. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, WA 4608/1-2, J.S.W.), the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy (EXC2180 390900677, J.S.W. and H.-G.R.), the German Cancer Consortium (DKTK, H.-G.R. and J.S.W.), the Ernst Jung Prize for Medicine (H.-G.R.), the Landesforschungspreis of Baden-Württemberg (H.-G.R.), the Wilhelm Sander Stiftung (2016.177.3, J.S.W.), the Deutsche Krebshilfe (German Cancer Aid, 70114948, J.S.W.), and the Fortüne Program of the University of Tübingen (2451-0-0, J.S.W.). The collaboration between Tübingen University and I2BC was supported by the French-German Partnership Hubert Curien Procope 2021 program, A.M. and H.-G.R.

Author contributions

A.M. conceived and designed the project together with O.N. for the Riboseq experiments. Design and performed the experiments: L.B., A.N., B.C.R., I.H., S.G., D.D., L.G., A.B., Y.V. Analysed the data: H.A., P.F., L.B., G.B., A.N., S.D., B.C.R., I.H., S.G., Y.V., J.V, M.P., S.G.-D., J.W. O.N., and A.M. Contributed reagents/materials/analysis tools: H.A., P.F., G.B., I.H., A.G., M.C.Z., C.G., M.A., L.H., E.R., E.L., R.J.-M., S.C., A.E., A.S., B.A., O.L., H.-G.R., J.W., O.N., and A.M. Wrote the paper: L.B. and A.M. with contributions from H.A., A.N., and O.N. and 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/s41467-025-56773-2.

Correspondence and requests for materials should be addressed to Olivier Namy or Arnaud Moris.

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]Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell (I2BC), 91190 Gif-sur-Yvette, France. [2]Sorbonne Université, Inserm U1135, CNRS ERL 8255, Centre d'Immunologie et des Maladies Infectieuses (CIMI-Paris), 75013 Paris, France. [3]Department of Peptide-based Immunotherapy, University and University Hospital Tübingen, 72076 Tübingen, Germany. [4]Institute of Immunology, University of Tübingen, 72076 Tübingen, Germany. [5]Cluster of Excellence iFIT (EXC2180) "Image-Guided and Functionally Instructed Tumor Therapies", University of Tübingen, Tübingen, Germany. [6]Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, 46 allée d'Italie F-69364, Lyon, France. [7]ADLIN Science, Evry-Courcouronnes, France. [8]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. [9]IRIM, UMR 9004, CNRS, Université de Montpellier, Montpellier, France. [10]Leibniz Institute of Virology, Hamburg, Germany. [11]ESPCI Paris, PSL University, Spectrométrie de Masse Biologique et Protéomique, CNRS UAR2051 Paris, France. [12]Vaccine Research Institute (VRI), INSERM-U955 (IMRB) Équipe 16, Université Paris-Est Créteil (UPEC), Créteil, France. [13]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), partner site Tübingen, 72076 Tübingen, Germany. [14]Centre Hospitalier Universitaire d'Orléans, Orléans, France. [15]Université Paris Saclay, Inserm, CEA, AP-HP, UMR1184 IDMIT, Department of Internal Medicine & Clinical Immunology, Bicêtre Hospital, Le Kremlin-Bicêtre, Bicêtre, France. [16]Clinical Collaboration Unit Translational Immunology, German Cancer Consortium (DKTK), Department of Internal Medicine, University Hospital Tübingen, 72076 Tübingen, Germany. [17]These authors contributed equally: Annika Nelde, Bertha Cecilia Ramirez, Isabelle Hatin. ✉e-mail: olivier.namy@i2bc.paris-saclay.fr; arnaud.moris@i2bc.paris-saclay.fr