Single-cell multi-omics profiling uncovers the immune heterogeneity in HIV-infected immunological non-responders

Xiaosheng Liu,[a,b,c] Leidan Zhang,[c] Xiaodi Li,[c] Ling Chen,[c] Lianfeng Lu,[c] Yang Yang,[c] Yuanni Wu,[c] Liyuan Zheng,[c] Jia Tang,[c] Fada Wang,[c] Yang Han,[c] Xiaojing Song,[c] Wei Cao,[c,d,e] and Taisheng Li[a,b,c,d,e,*]

[a]School of Basic Medical Sciences, Tsinghua University, 100084, Beijing, China
[b]Centre for Life Sciences, Tsinghua University, 100084, Beijing, China
[c]Department of Infectious Diseases, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, 100730, Beijing, China
[d]State Key Laboratory for Complex, Severe, and Rare Diseases, Peking Union Medical College Hospital, 100730, Beijing, China
[e]Center for AIDS Research, Chinese Academy of Medical Sciences, 100730, Beijing, China

eBioMedicine
2025;115: 105667
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105667

Summary

Background Immunological non-responders (INRs) are people living with HIV-1 who fail to achieve full immune reconstitution despite long-term effective antiretroviral therapy (ART). This incomplete recovery of CD4$^+$ T cells increase the risk of opportunistic infections and non-AIDS-related morbidity and mortality. Understanding the mechanisms driving this immune dysfunction is critical for developing targeted therapies.

Methods We performed single-cell RNA sequencing (scRNA-seq) and single-cell VDJ sequencing (scVDJ-seq) on peripheral blood mononuclear cells (PBMCs) from INRs, immune responders (IRs), and healthy controls (HCs). We developed scGeneANOVA, a novel mixed model differential gene analysis tool, to detect differentially expressed genes and pathways. In addition, we developed the Viral Identification and Load Detection Analysis (VILDA) tool to quantify HIV-1 transcripts and investigate their relationship with interferon (IFN) pathway activation.

Findings Our analysis revealed that INRs exhibit a dysregulated IFN response, closely associated with CD4$^+$ T cell exhaustion and immune recovery failure. The scGeneANOVA tool identified critical genes and pathways that were missed by traditional analysis methods, while VILDA showed higher levels of HIV-1 transcripts in INRs, which may drive the heightened IFN response. These findings support a potential contribution of IFN signalling in INR-related immune dysfunction.

Interpretation Our study provides new insights into the pathogenic mechanisms behind immune recovery failure in INRs, suggesting that IFN signalling might be involved in the development of CD4$^+$ T cell exhaustion. The identification of key genes and pathways offers potential biomarkers and therapeutic targets for improving immune recovery in this vulnerable population.

Funding This study was supported by the grants from Special Research Fund for the Central High-level Hospitals of Peking Union Medical College Hospital (Grant No. 2022-PUMCH-D-008), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (Grant No. 2021-I2M-1-037), National Key Technologies R&D Program for the 13th Five-year Plan (Grant No. 2017ZX10202101-001). The funders played no role in the design, experiment conduction, data analysis and preparation of the manuscript of this work.

Keywords: Human immunodeficiency virus-1; Immunological non-responders; Interferon; Acquired immunodeficiency syndrome; CD4$^+$ T cells; Single cell RNA sequencing

Introduction

Chronic human immunodeficiency virus type 1 (HIV-1) infection typically leads to the overactivation and functional exhaustion of CD4$^+$ T cells, eventually resulting in the progressive acquired immunodeficiency syndrome (AIDS).[1,2] HIV remains a major global health issue, affecting millions and causing numerous new infections annually. Antiretroviral therapy (ART) has significantly

*Corresponding author. School of Basic Medical Sciences, Tsinghua University, 100084, Beijing, China.
E-mail address: litsh@263.net (T. Li).

Research in context

Evidence before this study

Immunological non-responders (INRs) are individuals living with HIV-1 who do not achieve complete immune recovery despite effective antiretroviral therapy (ART), increasing their vulnerability to infections and non-AIDS-related conditions. Previous studies have suggested that chronic immune activation, particularly interferon (IFN) signalling, plays a role in immune dysfunction in INRs. However, the exact molecular mechanisms, especially regarding the trajectory of $CD4^+$ T cell exhaustion, have not been fully elucidated.

Added value of this study

This study combined single-cell RNA sequencing (scRNA-seq) and single-cell VDJ sequencing (scVDJ-seq) to analyse peripheral blood mononuclear cells (PBMCs) from INRs, immune responders (IRs), and healthy controls (HCs). We introduced scGeneANOVA, a mixed model differential gene analysis tool, which revealed important differentially expressed genes and pathways overlooked by traditional methods. Through RNA velocity and TCR information analysis, we identified that $CD4^+$ T cells in INR patients are more likely to follow an exhaustion trajectory, a process closely linked to IFN activation. Additionally, the Viral Identification and Load Detection Analysis (VILDA) tool confirmed that INRs exhibit higher levels of HIV-1 transcripts, which may further drive the dysregulated IFN response.

Implications of all the available evidence

This study provides critical insights into the immune dysfunction seen in INRs, emphasizing the potential role of IFN signalling in driving $CD4^+$ T cell exhaustion. The identification of specific genes and exhaustion-related pathways offers potential biomarkers and therapeutic targets for restoring immune function in this population. These findings suggest that strategies aimed at modulating IFN signalling may help improve immune reconstitution in INRs, potentially leading to better clinical outcomes.

improved the prognosis for people living with HIV (PLWH) by effectively suppressing viral replication, increasing $CD4^+$ T cell counts, improving the CD4/CD8 ratio, and repairing immune dysfunction.[3,4] However, approximately 10–40% of PLWH do not achieve adequate immune reconstitution of $CD4^+$ T cell counts despite sustained virological suppression with ART.[5] These individuals are termed immunological non-responders (INRs). INRs face a significantly higher risk of progressing to AIDS, experiencing non-AIDS-related events such as cardiovascular diseases and cancers, and have higher mortality rates compared to immunological responders (IRs).[6–8] The mechanisms and biomarkers of INR remain poorly understood, posing a significant challenge in HIV/AIDS treatment.[9]

Persistent chronic immune activation and systemic inflammation are major contributors to $CD4^+$ T cell loss in INRs. $CD4^+$ T cells in INRs exhibit higher levels of immune activation, exhaustion, and senescence markers.[10–12] Our previous bulk RNA-seq analysis demonstrated that INRs exhibit sustained, enhanced expression of the interferon (IFN) response pathway,[13] a phenomenon further confirmed by Tang et al.[14] Recent studies propose a "pathological proliferation" hypothesis in INRs, suggesting that the over-proliferation of poor-quality $CD4^+$ T cells result in incomplete immune reconstitution.[15] Additionally, widespread abnormal immune activation is observed across various immune cells in INRs, including $CD8^+$ T cells, regulatory T cells (Tregs), double-negative T cells (DNT), mucosa-associated invariant T cells (MAIT), natural killer (NK) cells, B cells, and myeloid cells.[6,7,16] These findings raise questions about whether INRs exhibit a distinct profile of aberrant immune activation and how this leads to $CD4^+$ T cell loss.

Single-cell RNA sequencing (scRNA-seq) provides a granular view of gene expression at the individual cell level, offering crucial insights into the immune system of PLWH.[17,18] Unlike bulk RNA sequencing, scRNA-seq unveils cellular heterogeneity and dynamics, which are essential for understanding immune dysregulation.[19] Recent applications of this technology have identified significant differences in the immune cell profiles of INRs. For example, Li et al. reported diminished levels of MAIT cells and indications of mitochondrial dysfunction.[20] Wang et al. observed the heightened proportions of $GNLY^+CD4^+$ and $CD8^+$ effector T cells have been observed, correlating with poor immune reconstitution.[17] Furthermore, Jia et al. noticed the impaired diversity of the B cell receptor (BCR) repertoire and differentiation in naïve B cells of INRs through single-cell immune repertoire sequencing (scVDJ-seq).[21] These insights underscore the pivotal role of scRNA-seq in elucidating the mechanistic underpinnings of immune recovery failure. However, to fully comprehend the mechanisms driving immune dysregulation and incomplete immune recovery in INRs, comprehensive and large-scale single-cell analyses are essential.

In this study, we employed scRNA-seq and scVDJ-seq to comprehensively analyse the immunological characteristics of peripheral blood mononuclear cells (PBMCs) from INRs. We developed a mixed model differential gene analysis approach, scGeneANOVA, to examine the gene expression profiles of INRs. Our findings revealed a strong association between IFN response and exhaustion signatures during $CD4^+$ T cell differentiation. These results were further validated

using bulk RNA-seq cohorts and public datasets. Additionally, we developed the Viral Identification and Load Detection Analysis (VILDA) tool, which demonstrated that PBMCs from INRs harbour more HIV-1 transcripts, leading to elevated IFN signatures. This study provides a high-resolution transcriptional landscape of peripheral immune cells in INRs, underscoring the impact of dysregulated IFN response on $CD4^+$ T cell differentiation, and enhances the understanding of the mechanisms underlying incomplete immune reconstitution.

Methods

Study design

This study utilized scRNA-seq, scVDJ-seq, and bulk RNA-seq to investigate the immunological diversity of INRs, IRs, and healthy controls (HCs). The primary scRNA/VDJ-seq cohort consisted of 20 individuals, including seven INRs, nine IRs, and four HCs. Additionally, scRNA/VDJ-seq data from 13 external individuals (three INRs, five IRs, and five HCs),[21,22] were incorporated to enhance the statistical power and expand the cohort to 33 individuals in total (Table S1). The bulk RNA-seq cohort consisted of 60 individuals, including 23 INRs, and 37 IRs (Table S2). Of note, there was no overlap between participants in the scRNA/VDJ-seq cohort and those in the bulk RNA-seq cohort.

For both scRNA/VDJ-seq and bulk RNA-seq samples generated in this study, INRs were defined as HIV-positive patients on ART for over 4 years, with virologic control (VL <20 copies/mL) for 3.5 years but low $CD4^+$ counts (<350 cells/μL). IRs met the same criteria but had $CD4^+$ counts >500 cells/μL (Figure S1). To minimize confounding factors affecting immune function, individuals with the following conditions were excluded from the study: (1) Co-infections with hepatitis B virus (HBV), hepatitis C virus (HCV), or Mycobacterium tuberculosis (TB). (2) Other immunosuppressive diseases such as cancer or autoimmune disorders requiring immunosuppressive therapy. (3) Concurrent use of immunomodulatory drugs apart from ART. (4) Acute infections within one month prior to sample collection. HCs were HIV-negative individuals with no known immunological disorders or chronic conditions.

This study was approved by the Independent Ethics Committee of Peking Union Medical College Hospital, with approval ID 1-23PJ1189.

PBMC isolation and quality control

Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll–Paque PLUS, followed by washing and red blood cell lysis. Cell viability and quality control were assessed using acridine orange/propidium iodide (AO/PI) dual fluorescent staining on the Countstar Rigel (S2) instrument. The quality control criteria were as follows: cell concentration of 1000–1200 cells/μL, cell volume >60 μL, cell viability >90%, aggregation rate <15%, and cell diameter <30 μm. After confirming that the cell concentration and viability met the required standards, the cell suspension was placed on ice and processed for the 10× Genomics single-cell transcriptomics workflow within 30 min.

scRNA-seq library preparation

scRNA-seq, scTCR-seq, and scBCR-seq libraries were prepared in Biomarker Technologies (Qingdao, China) using the 10× Genomics Chromium platform, loading at least 10,000 cells per channel. Reverse transcription, cDNA amplification, and gene expression libraries were constructed, followed by sequencing on an Illumina HiSeq X Ten. FASTQ data were processed with FastQC and CellRanger (v7.1.0) for alignment and quantification. Seurat was used for data filtering, normalization, PCA, UMAP, clustering, and integration of TCR/BCR clonotypes. Cells with fewer than 1000 genes, 500 UMIs, or >10% mitochondrial genes were excluded. Platelet genes ("PPBP", "PF4", "GNG11", "CAVIN2") were removed and doublets were removed with DoubletFinder. Batch effects were corrected using the Harmony package.[23]

TCR/BCR repertoire data were processed with CellRanger VDJ and STARTRAC was used to analyse T and B cell dynamics across cell types.[24] The STARTRAC analysis provided quantitative information on T and B cells sharing the same immune repertoire, including clonal expansion (identical information within the same cell type), developmental transitions (identical information between different cell types), and cross-tissue migration (identical information between different tissues).[25]

Differential gene analysis

For cell cluster annotation, differentially expressed genes (DEGs) were identified using the Seurat FindAllMarkers function. To compare gene expression between groups, we developed scGeneANOVA, a mixed model approach that combines fold change (FC) calculations with rigorous statistical testing at the sample level. FCs are calculated using average expression levels across samples, followed by a pseudobulk strategy where gene expression values are analysed using ANOVA and Tukey's post-hoc test. DEGs are identified based on adjusted P-values (P-adj) from the post-hoc test. scGeneANOVA tool is available at https://github.com/rrrrroentsen/scGeneANOVA.

Gene set enrichment analysis

GSEA was performed using the fgsea package (v1.26.0). DEGs were ranked by average log2 fold change, and Hallmark gene sets were obtained from MSigDB via the msigdbr package (v7.5.1). Pathways with adjusted P-values <0.05 and normalized enrichment score (NES) > |1.0| were considered significant. Additionally, Sample-Level Enrichment Analysis (SLEA) was conducted using SLEA software (v2.3.1) to calculate Z-scores for each

sample, determining the relative importance of pathways at the sample level.

Gene set scoring algorithm

The AUCell package (v1.22.0) was used to assess gene set activity in single-cell RNA-seq data. Gene expression rankings were built using AUCell_buildRankings, followed by calculation of AUC scores with AUCell_calcAUC. Thresholds for active gene sets were explored with AUCell_exploreThresholds, assigning cells to active states. These AUC scores were integrated into Seurat metadata, and UMAP plots were generated to visualize the scores, with median UMAP coordinates used to label cell types.

RNA velocity and pseudotime analysis

RNA velocity analysis was conducted using scvelo (v0.1.25), with Seurat metadata (UMAP coordinates, clusters, cell types) integrated into the loom file. After filtering and normalization, RNA velocities were computed using scv.pp.moments function, and velocity graphs were visualized on UMAP using scv.pl.velocity_embedding and scv.pl.velocity_embedding_stream function. Single-cell trajectory analysis was performed using Slingshot (v2.8.0)[26] and Monocle2 (v2.28.0). Slingshot inferred developmental trajectories in diffusion map space, and pseudotime was calculated for each cell. Monocle2 identified highly variable genes and performed differential gene testing along pseudotime, with results visualized in pseudotime and branched heatmaps, highlighting dynamic cell state transitions.

Cell chat analysis

Cell–cell communication networks were analysed using the CellChat package (v1.6.1).[27] Potential interactions were identified using the human CellChat database, and the dataset was filtered for genes present in both databases. Overexpressed genes and interactions were identified, and communication probabilities were computed, accounting for population size and filtering low-expression genes. Pathway-level communication was aggregated, and network visualizations displayed the number and strength of interactions. Centrality measures were calculated, and cell type signalling roles were visualized using scatter plots. Interaction data were exported for further analysis.

Bulk RNA-seq and data analysis

The bulk RNA-seq pipeline includes quality control, alignment, quantification, and normalization. MultiQC (v1.9) performs initial quality control by generating metrics and visualizations. Reads are aligned to the GRCh38 reference genome using HISAT2 (v2.1.0), with SAM files converted to BAM and sorted via Samtools (v1.10). Gene expression levels are quantified using HTSeq based on the GTF annotation file. To account for sequencing depth and gene length, raw counts are normalized to TPM and FPKM, ensuring comparability across samples for downstream analysis.

VILDA analyses

VILDA was developed for comprehensive analysis of viral infections in bulk RNA-seq data from INRs. After inputting FASTQ data, reads were aligned to the human GRCh38-2020-A genome using STAR, obtaining both mapped and unmapped reads. Unmapped reads were aligned to a viral genome database (14,328 genomes) using Minimap2, and viruses with low read counts (<2) or high dropout rates (>20%) were filtered out. The VILDA tool is available at https://github.com/rrrrroentsen/VILDA.

Measurement of HIV-1 total DNA and cell-associated RNA

HIV-1 total DNA and cellular RNA were extracted from isolated PBMC samples using the QIAamp DNA Mini Kit (Qiagen, Valencia, California) and the HiPure Total RNA Plus Mini Kit (Magen, Guangzhou, China), respectively. These extracts were then amplified and quantified using the HIV-1 total DNA and cell-associated RNA (ca-RNA) quantitative detection kits (SUPBIO, Guangzhou, China) on the Roche LightCycler 480 (LC480) real-time PCR system. The quantification of HIV total DNA and caRNA was adjusted by the proportion of $CD4^+$ T cells in PBMCs and expressed as the number of DNA/RNA copies per 1×10^6 cells.

Statistical analyses

Cell proportions and gene expression levels were presented as mean and standard deviations. Differences between the three groups were analysed using ANOVA followed by Tukey's post-hoc test. Correlations were assessed using the Spearman test. A P-value of less than 0.05 was considered statistically significant. Unless otherwise noted, all statistical analyses and graphing were performed using SPSS (v25.0), R software (v4.3.1), R package ggplot2 (v3.5.1) and GraphPad Prism (v9.0.0).

Role of the funding source

This study was supported by the grants from Special Research Fund for the Central High-level Hospitals of Peking Union Medical College Hospital (Grant No. 2022-PUMCH-D-008), Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (Grant No. 2021-I2M-1-037), National Key Technologies R&D Program for the 13th Five-year Plan (Grant No. 2017ZX10202101-001). The funders played no role in the design, experiment conduction, data analysis and preparation of the manuscript of this work.

Results

A single-cell RNA-seq atlas of peripheral immune cells in INRs

To investigate the immunological diversity of INRs at single-cell resolution, we performed scRNA/VDJ-seq on PBMCs from 20 individuals, including 7 INRs, 9 IRs, and 4 HCs. Additionally, we incorporated scRNA/VDJ-

seq data from 13 external individuals (3 INRs, 5 IRs, 5 HCs) sourced from previous studies,[21,22] resulting in a combined cohort of 33 individuals (10 INRs, 14 IRs, and 9 HCs) (Fig. 1a, Table S1). For the 20 individuals recruited for this study, the average age of INRs was 42 ± 6.8 years, predominantly male, with an average ART duration of 6.9 ± 3.6 years. INRs had significantly lower median $CD4^+$ T cell counts (214, IQR 195–-258 cells/μL) compared to IRs (817, IQR 784–1058 cells/μL) and HCs (698, IQR 512–890 cells/μL) (Table S1).

After quality control and batch effect correction, we merged data from 232,369 cells across 33 samples, producing a comprehensive immune cell atlas (Fig. 1b). Sequencing saturation averaged 80.6% (Table S3). TCR/BCR profiles were integrated with scRNA-seq data (Fig. 1c). Eight major immune compartments were identified: $CD4^+$ T cells ($CD3D^+CD4^+$), $CD8^+$ T cells ($CD3D^+CD8A^+$), γδ T cells ($CD3D^+TRDC^+$), natural killer cells (NK, $GNLY^+NKG7^+$), B cells ($MZB1^+CD79^+$), plasma cells ($MZB1^+CD38^+$), monocytes (LYZ^+CD14^+), and dendritic cells (LYZ^+FLT3^+) (Fig. 1d, Figures S2 and S3). Based on top DEGs, we further identified 51 subclusters within these compartments (Fig. 1e, Tables S4–S6).

The proportion of $CD4^+$ T cells in INRs (10.4 ± 4.6%) was significantly lower than in IRs (27.5 ± 7.0%) and HCs (26.2 ± 9.0%) (Padj <0.0001), while the proportion of $CD8^+$ T cells was higher in both INRs (43.6 ± 11.0%) and IRs (36.0 ± 9.1%) compared to HCs (23.8 ± 4.5%) (Padj <0.0001) (Fig. 1f). Other cell type frequencies showed no significant differences across groups, indicating that variations were primarily in the $CD4^+$ and $CD8^+$ T cell compartments. This pattern was consistent across experimental batches (Figure S4), emphasizing the key immunological features of INRs—reduced $CD4^+$ and elevated $CD8^+$ T cell proportions compared to IRs and HCs.

scGeneANOVA and GSEA identified the enrichment of MHC-II and IFN response pathways in INRs

After annotating the single-cell transcriptomic map, we performed differential gene expression analysis across patient groups (Tables S7 and S8). Our study found that commonly used methods—Seurat's FindMarkers (Wilcoxon rank-sum test, Table S7) and edgeR-LRT (pseudobulk test, Table S8)[28]—were not well-suited for this research. As shown in Figure S5b, AC004448.2 had the biggest fold change value and lowest adjusted P-value in Wilcoxon rank-sum test. However, ANOVA showed no significant intergroup difference, as the expression of AC004448.2 was significantly elevated in a specific INR patient, causing false-positive results in both methods (Figure S5c). These results highlighting the susceptibility of single-cell and pseudobulk methods to individual sample effects.

To address this issue, we developed a mixed model differential gene analysis approach (scGeneANOVA), detailed in the Methods section (Fig. 2a). Using the scGeneANOVA function, we identified 33 new differentially expressed genes that were not detected by the single-cell and pseudobulk methods, including ITGA4, LAG3, BHLHE40, CD4, EEF1A11, and SERINC5 (Fig. 2b–c, Table S9). The above genes have been previously reported to be closely associated with HIV infection and INR phenotype.[29–32]Besides, scGeneANOVA successfully excluded AC004448.2 as differentially expressed gene, indicating better sensitivity of the mixed model. These results collectively demonstrate that the mixed model differential gene algorithm based on the scGeneANOVA function is more suitable for our study design compared to single-cell and pseudobulk analysis models.

Using the scGeneANOVA analysis method, we obtained differential gene results across different cell types in INRs relative to IRs (Fig. 2d, Table S10). Frequency analysis revealed that the HLA-C gene had the highest frequency across different cell types, followed by ITGA4, ACTN1, and SNHG29 genes (Fig. 2e). GO enrichment analysis of all upregulated differentially expressed genes indicated significant changes in biological processes such as MHC class II protein complex and MHC protein complex (Fig. 2f).

Using the results from scGeneANOVA, we conducted gene set enrichment analysis (GSEA) on the entire gene sets, utilizing hallmark pathways from MSigDB. Among these enriched inflammatory pathways, the IFN-γ response was the most significantly enriched pathway with the lowest P-value disparity between INRs and IRs (Fig. 2g). Notably, the predominance of the IFN-γ response was primarily attributed to $CD4^+$ T cells and $CD8^+$ T cells (Fig. 2h). We further investigated the expression of the IFN-γ response pathway at the individual cell level by calculating the AUC scores for each cell. Results confirmed that INRs had upregulated IFN responses in both $CD4^+$ and $CD8^+$ T cells, contributing to the enhanced expression of IFN-related genes in the global PBMC transcriptional profiles (Fig. 2i).

Upregulated innate immune cell activation profiles in INRs

Innate immune cells, including myeloid cells and innate lymphoid cells, are major components of the immune system's first line of defence.[33] Based on canonical myeloid markers, we identified eight myeloid subclusters (Figures S6, S7). Despite significant heterogeneity in the proportions of myeloid cells among individual samples, there were no significant differences between INRs and IRs in the proportions of these subclusters (Figure S7b).

Using the scGeneANOVA algorithm, we discovered significant differential gene expression in pDCs, including HLA-C, LGALS1, CD83, and JAK2, suggesting a more activated and mature phenotype in

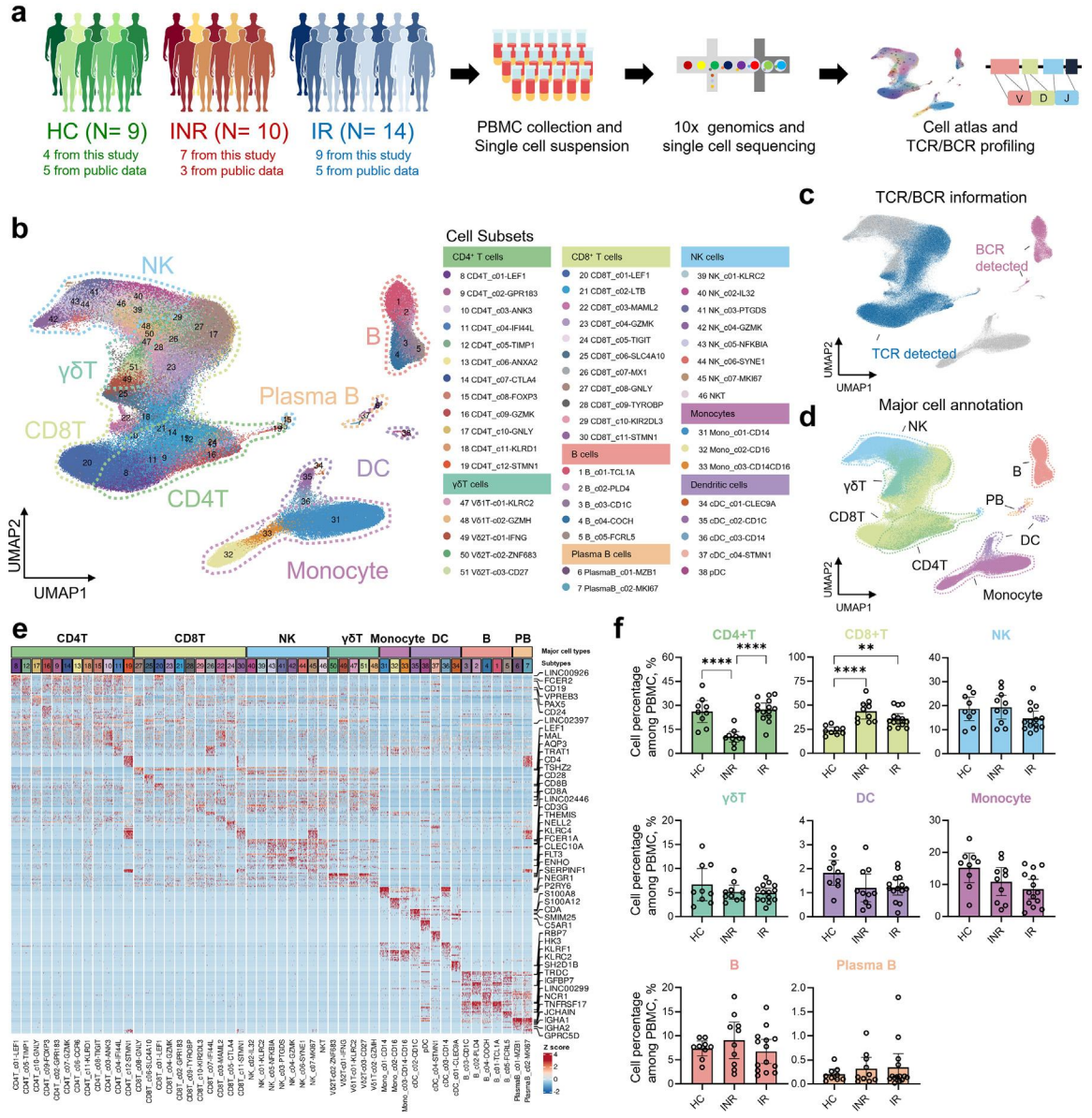

***Fig. 1:* Characterization of the immune cell profiles of INRs by scRNA/VDJ-seq.** (a) Study design overview showing the cohort composition: healthy controls (HC, N = 9), immunological non-responders (INR, N = 10), and immunological responders (IR, N = 14). PBMCs were collected and subjected to single-cell RNA and TCR/BCR sequencing. (b) UMAP visualization of the minor cell subsets identified from the scRNA-seq data. The right panel details the specific cell subsets identified within each minor cell type. (c) UMAP plot showing the detection of TCR and BCR sequences, highlighting cells with detectable TCR and BCR. (d) UMAP plot annotated with major cell types, demonstrating the clustering of different immune cell populations. (e) Heatmap showing the expression of key marker genes across different immune cell subsets. Each column represents a single cell, and each row represents a gene, with the colour intensity indicating the expression level. (f) Bar plots comparing the percentage of each major cell type among total PBMCs across HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01, ***P < 0.001).

pDCs from INRs (Figure S7c and d). Furthermore, CD83 expression in monocytes and classical dendritic cells of both INRs and IRs was significantly higher than in HCs, indicating widespread myeloid cell activation in PLWH (Figure S7e). We also found that CD83 expression levels in pDCs were significantly positively correlated with overall PBMC interferon activation levels (Pearson's R = 0.4556, P-value = 0.0077) and significantly negatively correlated with $CD4^+$ T cell proportions (Pearson's R = –0.4795, P-value = 0.0047) (Figure S7f). This finding underscores the critical role of pDCs, the primary producers of type I interferon in the body, in the INR phenotype in HIV infection.[34,35]

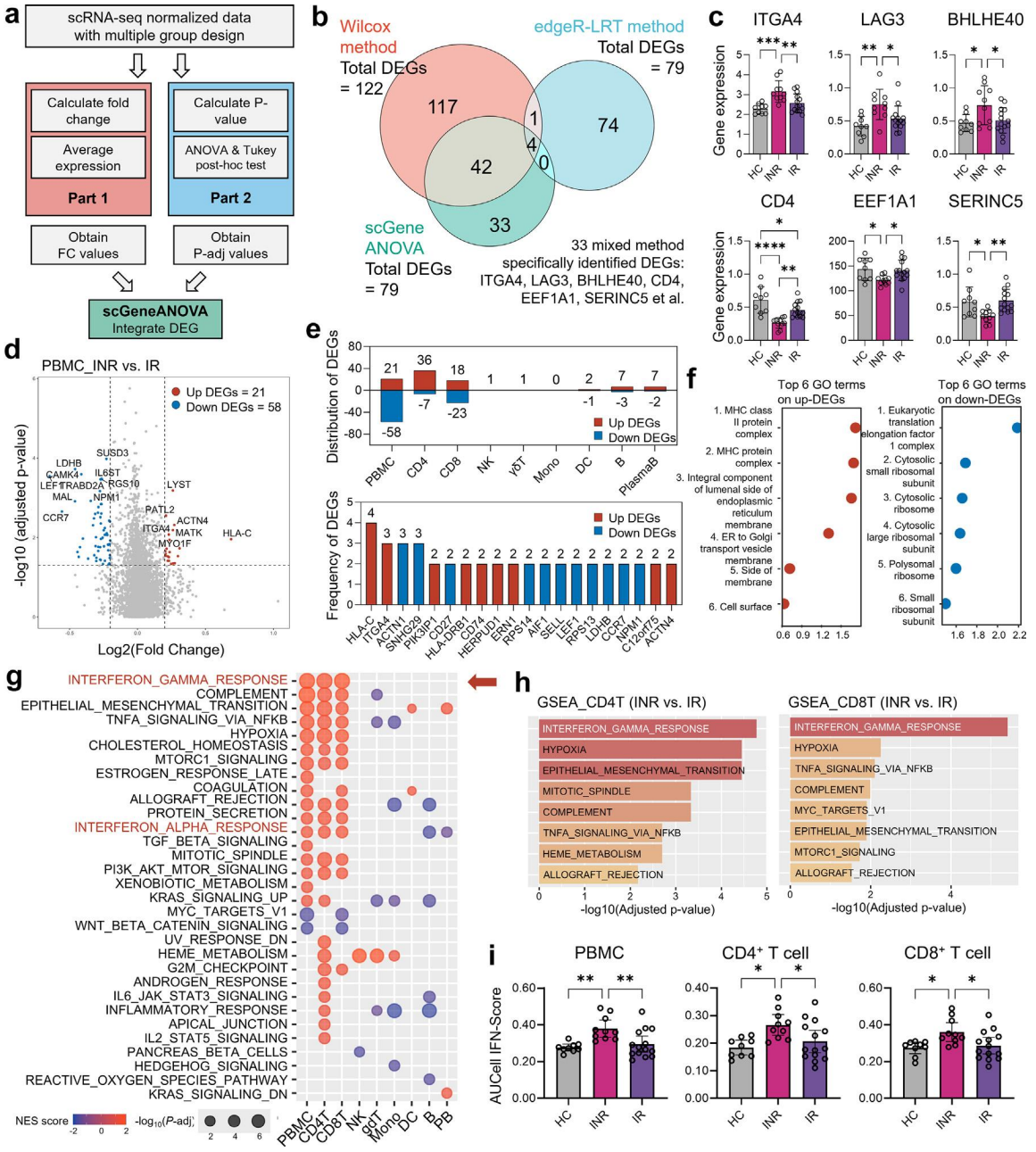

Fig. 2: **scGeneANOVA analysis and GSEA pathway enrichment analysis**. (a) Workflow diagram summarizing the multi-level scRNA-seq data analysis approach, including single-cell and pseudo-bulk methods to identify differentially expressed genes (DEGs). (b) Venn diagram showing the overlap of DEGs identified by single-cell, pseudo-bulk, and mixed methods. A total of 33 DEGs were uniquely identified using the mixed method. (c) Box plots showing the expression levels of selected DEGs in PBMCs across HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01, ***P < 0.001). (d) Volcano plot of DEGs in PBMCs between INR and IR groups, highlighting upregulated (red) and downregulated (blue) genes. The right panel shows the distribution of DEGs across different cell types. (e) Bar plots illustrating the distribution and frequency of DEGs across various cell types. The distribution plot of DEGs (upper panel) shows the number of identified DEGs in each cell type, while the frequency plot of DEGs (lower panel) represents the occurrence frequency of the identified DEGs across different cell types. Up-regulated DEGs are labelled in red colour, and down-regulated DEGs are labelled in bule colour. (f) Top 6 enriched Gene Ontology (GO) terms for upregulated (left) and downregulated (right) DEGs in INR vs. IR comparisons, with terms related to immune response and cellular components. (g) Heatmap showing the Normalized Enrichment Score (NES) of various pathways in PBMCs and specific cell subsets between INR and IR groups. Pathways like interferon gamma response and epithelial mesenchymal transition are prominently enriched. (h) GSEA results for $CD4^+$ T cells and $CD8^+$ T cells between INR and IR groups, displaying significant enrichment of interferon gamma response and other pathways. (i) Bar plots comparing the AUCell interferon score across PBMCs, $CD4^+$ T cells, and $CD8^+$ T cells in HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01).

Besides myeloid cells, we identified eight NK subclusters and five γδ T cell subclusters, with no significant differences between INRs and IRs in these subclusters (Figures S8, S9). However, IRs showed higher proportions of Vδ1T-c01-KLRC2 and NK_c01-KLRC2 cells and lower proportions of Vδ2T-c04-ZNF683 and NK_c03-PTGDS cells compared to HCs (Figure S9b). These results indicate that despite optimal CD4$^+$ T cell recovery, HIV-1 infected patients still have distinct innate immune profiles compared to healthy individuals.

Additionally, scGeneANOVA results revealed that the LYST gene is differentially expressed in five NK cell subclusters: NK_c03-PTGDS, NK_c04-GZMK, NK_c05-NFKBIA, NK_c06-SYNE1, and NK_c07-MKI67 subgroups, which belong to the CD56^{bright}CD16$^-$ NK cells and cycling NK cells (Figure S9c). LYST is known to regulate multiple aspects of NK cell lytic activity, and we found that the expression levels of LYST in the NK cells were significantly negatively correlated with CD4$^+$ T cell proportions (Pearson's R = –0.5225, P-value = 0.0018) (Figure S9d). This finding is consistent with previous studies, indicating that NK cells in INRs exhibit significantly enhanced activation and cytotoxic function.[36–38]

INRs exhibit extensive heterogeneity in CD4$^+$ T cell subclusters

Further clustering of CD4$^+$ T cells yielded 12 distinct subclusters, including naïve (T_{Naive}, CD4T_c01-LEF1), central memory (T_{CM}, CD4T_c02-GPR183), effector memory (T_{EM}, CD4T_c03-ANK3, CD4T_c04-IFI44L, CD4T_c05-TIMP1, CD4T_c06-CCR6, CD4T_c07-GZMK), effector (T_{EFF}, CD4T_c10-GNLY), killer Ig-like receptors (T_{KIR}, CD4T_c11-KLRD1), regulatory (Treg, CD4T_c09-FOXP3), exhausted (T_{EX}, CD4T_c08-TIGIT), and cycling-phenotype ($T_{Cycling}$, CD4T_c12-STMN1) (Fig. 3a–d). Notably, CD4T_c07-GZMK, CD4T_c05-TIMP1, and CD4T_c06-CCR6 were also identified as Th1, Th2, and Th17-like cells, respectively.

Significant differences in the proportions of CD4$^+$ T cell subtypes were observed (Fig. 3e). INRs exhibited lower proportions of CD4T_c01-LEF1 cells, indicative of a decrease in naive CD4$^+$ T cells, and higher proportions of CD4T_c02-GPR183, CD4T_c08-TIGIT, and CD4T_c12-STMN1 cells, suggesting an increase in central memory, exhausted, and cycling phenotypes compared to IRs and HCs (Fig. 3f). Notably, the proportions of CD4T_c05-TIMP1, CD4T_c06-CCR6, and CD4T_c07-GZMK cells were lower in IRs compared to HCs, indicating that despite achieved optimal CD4$^+$ T cell recovery, IRs have a distinct proportion of memory phenotype CD4$^+$ T cells than HCs (Fig. 3f).

scGeneANOVA results indicate that CD4$^+$ T cells in INRs exhibit significant differential gene expression, including MHC response genes such as ITGA4, HLA-C, HLA-DRB1, and CD74, as well as exhaustion markers TOX, LAG3, TIGIT, and HAVCR2 (Fig. 3g–i). By examining previously defined signature genes, we observed distinct functional states among each CD4$^+$ T cell subset. The CD4T_c01-LEF1 subset exhibited the highest naive score, while the CD4T_c09-FOXP3 and CD4T_c08-TIGIT subsets had the highest exhaustion scores. The CD4T_c07-GZMK and CD4T_c10-GNLY subsets showed elevated inflammation scores, and the CD4T_c10-GNLY and CD4T_c11-KLRD1 subsets demonstrated the highest cytotoxic scores (Figure S10).

These functional distinctions were also evident when comparing CD4$^+$ T cells between different patient groups. CD4$^+$ T cells in INRs exhibited the highest levels of inflammatory and exhaustion scores, whereas the levels of naive and cytotoxic scores in INRs were lower than those in IRs. These differences were contributed by various cell subsets and were consistent across them (Figure S10). These shifts in subpopulation frequencies and gene expression indicate extensive heterogeneity of CD4$^+$ T cells in INRs, suggesting a skew towards a more differentiated and exhausted state.

IFN response highly involved in the exhaustion trajectory of CD4$^+$ T cells

To elucidate the developmental transitions of CD4$^+$ T cell clusters, we applied RNA velocity analysis and Slingshot to construct the developmental trajectories of 12 CD4$^+$ T cell clusters. We identified two prominent directional streams originating from the naïve population (CD4T_c01-LEF1) towards the exhausted population (CD4T_c08-TIGIT) and the effector population (CD4T_c10-GNLY), respectively (Fig. 4a). Additional Monocle2 analysis also demonstrated the two directions of CD4$^+$ T cell differentiation (Fig. 4b). Along the trajectory leading to exhausted cells, the expression of CTLA4 in CD4$^+$ T cells specifically increased. Conversely, along the trajectory towards effector cells, the expression of GNLY specifically increased (Fig. 4c). We designated these two CD4$^+$ T cell differentiation pathways as the "Exhaustion trajectory" and the "Effector trajectory".

We then performed STARTRAC analysis using the scTCR-seq information (Fig. 4d–e). By comparing the STARTRAC-expa index derived from TCR sequence data, we observed that the effector population (CD4T_c10-GNLY) exhibited a significantly higher expansion capability compared to other cell types, a phenomenon consistent with findings from other studies.[39,40] Furthermore, through a comparison of the STARTRAC-tran index derived from TCR sequence data, we noted that the effector population (CD4T_c10-GNLY) in INRs showed significantly lower state transition capability (Fig. 4f). In conjunction with our prior discovery of elevated levels of exhausted CD4$^+$ T cells (CD4T_c08-TIGIT) in INRs, these observations suggest an increased tendency toward an exhaustion-prone differentiation trajectory in the CD4$^+$ T cells of INRs, with fewer cells differentiating into effector cells.

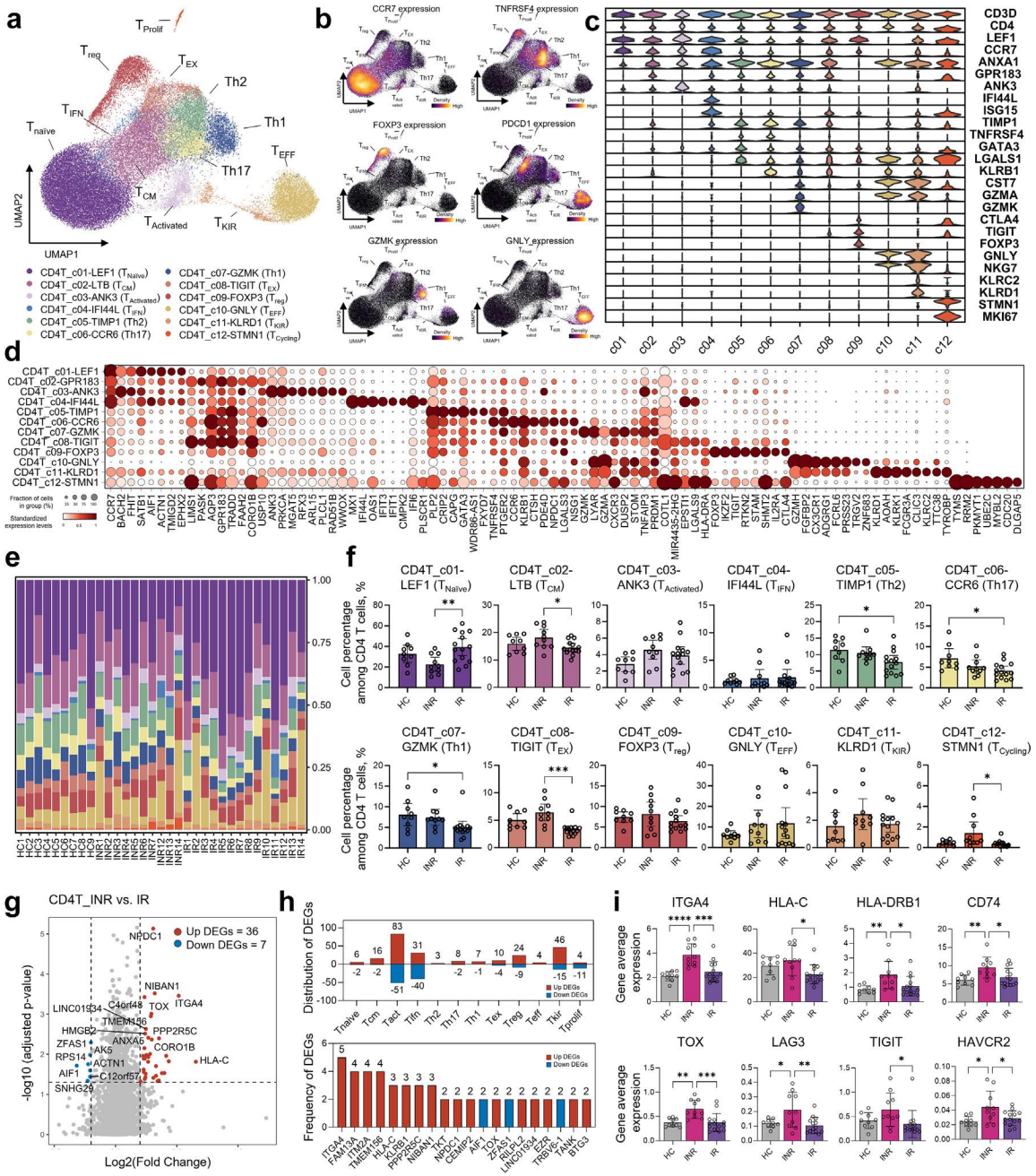

***Fig. 3*: Extensive heterogeneity in CD4⁺ T cell subclusters of INRs**. (a) UMAP plot showing the clustering of CD4⁺ T cell subclusters. Each subset is colour-coded and labelled accordingly. (b) UMAP plots showing the expression of specific marker genes across the identified CD4⁺ T cell subclusters, with higher expression levels indicated by darker colours. (c) Violin plots showing the distribution of key marker gene expressions across different CD4⁺ T cell subclusters. (d) Heatmap displaying the expression of selected marker genes across various CD4⁺ T cell subclusters. Each row represents a gene, and each column represents a cell, with colour intensity indicating expression levels. (e) Stacked bar plots showing the proportion of each CD4⁺ T cell subcluster among total CD4⁺ T cells across HC, INR, and IR groups. Each bar represents a group, and colours represent different subclusters. (f) Bar plots comparing the percentage of specific CD4⁺ T cell subclusters among total CD4⁺ T cells across HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01, ***P < 0.001). (g) Volcano plot of DEGs in CD4⁺ T cells between INR and IR groups, highlighting upregulated (red) and downregulated (blue) genes. Key genes are labelled, including ITGA4 and HLA-C. (h) Bar plots illustrating the distribution and frequency of DEGs across various CD4⁺ T cell subclusters. The distribution plot of DEGs (upper panel) shows the number of identified DEGs in each cell type, while the frequency plot of DEGs (lower panel) represents the occurrence frequency of the identified DEGs across different cell types. Up-regulated DEGs are labelled in red colour, and down-regulated DEGs are labelled in bule colour. (i) Box plots showing the expression levels of selected DEGs in CD4⁺ T cells across HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01).

To investigate the underlying mechanism driving the distinct transcriptional programs of the two trajectories, we analysed the gene expression kinetics of $CD4^+$ T cells with Monocle2 package. Ordered by pseudotime, we identified 154 DEGs which were clustered into two main groups (Fig. 4g). Notably, DEGs from cluster 1 (n = 90) primarily consisted of interferon-stimulated genes (ISGs) and highly enriched in IFN response pathway (Ranked as No. 3 in the Table S11), while cluster 2 (n = 64) was mainly composed of effector and cytotoxic genes (Fig. 4h). Additionally, pseudotime analysis highlighted a strong association between IFN and exhaustion signatures during $CD4^+$ T cell differentiation in the exhaustion trajectory (Fig. 4i). Notably, we observed that the expression levels of TIGIT and OAS1 in INRs were markedly higher compared to IRs and HCs, particularly during the later stages of the exhaustion trajectory, underscoring the distinct differences between INRs and IRs (Fig. 4j). Furthermore, the IFN signature scores in $CD4^+$ T cell were significantly positively correlated with exhausted cell percentage (Pearson's R = 0.5442, P-value = 0.0011) and significantly negatively correlated with $CD4^+$ T cell proportions (Pearson's R = −0.5500, P-value = 0.0009) (Fig. 4k).

To validate these findings, we established a new PLWH cohort consisting of 23 INRs and 37 IRs (Fig. 5a and b, Table S8). Flow cytometry showed higher levels of PD-1^+(encoded by PDCD1) and $TIGIT^+$ $CD4^+$ T cells in INRs, whereas there were no significant differences in LAG-3 and TIM-3 (encoded by HAVCR2) levels (Fig. 5c). Bulk RNA-seq revealed an elevated IFN gene signature in INRs, which correlated with increased exhausted $CD4^+$ T cells (Pearson's R = 0.4113, P-value = 0.0011) and lower $CD4^+$ T cell proportions (Pearson's R = −0.5063, P-value<0.0001) (Fig. 5d–g, Figure S11).

Using public RNA-seq data (GSE195543),[41] we confirmed that type-I IFN treatment induced exhaustion markers (PDCD1, LAG3, HAVCR2) in naïve $CD4^+$ T cells (Fig. 5h–i). Specifically, PDCD1 expression was upregulated at 4 h of IFN treatment, followed by LAG3 (16 h) and HAVCR2 (96 h) at later time points (Fig. 5i). This temporal pattern highlights the dynamic development of IFN-induced $CD4^+$ T cell exhaustion. Furthermore, knockdown of IRF or STAT transcription factors significantly reduced IFN-β-induced HAVCR2 expression (Fig. 5j), supporting the role of IFN in driving $CD4^+$ T cells into terminal exhaustion in non-HIV conditions.

Decreased naïve $CD8^+$ T cell proportion and enhanced MHC-I reaction in INRs

Similar to $CD4^+$ T cells, we identified 11 $CD8^+$ T cell subsets (Figures S12, S13). Notably, there was a significant decrease in the proportion of naïve $CD8^+$ T cells in INRs (Figure S13a and b). However, no significant differences were observed in the proportions of other cell types between the groups. STARTRAC analysis also revealed no differences in $CD8^+$ T cell dynamic development between INR and IRs.

scGeneANOVA results indicate that $CD8^+$ T cells in INRs exhibit significantly higher expression of ITGA4, PATL2, HLA-C, HLA-DRB1, and HLA-DPA1 genes (Figure S13c and d). Although LAG3 expression was elevated in INRs, other exhaustion markers (e.g., TOX, TIGIT, HAVCR2) did not show significant differences (Figure S13e). This was confirmed by flow cytometry, which showed a significantly higher proportion of exhausted $CD4^+$ T cells in INRs, but no significant difference in $CD8^+$ T cell exhaustion compared to IRs (Figure S14).

To further explore $CD8^+$ T cell function in INRs, we conducted cell communication analysis (Figure S15a). CellChat results revealed that $CD8^+$ T cells in INRs expressed more interaction genes, with the most significant differences observed in MHC-I-related molecules (Figure S15b and c). Both the signal reception and emission intensities of $CD8^+$ T cells in INRs were higher compared to IRs and HCs (Figure S15d). Although no difference in the overall number of cell interactions, the interaction intensity of INRs was significantly higher (Figure S15e). Taken together, these results suggest that the role of IFN-γ in $CD8^+$ T cells of INRs is likely to primarily involve upregulating the expression of MHC class I molecules rather than promoting their exhaustion.

Significant developmental deficiency of naive B cells in INRs

By focussing on humoural immune cells, we identified five B cell subclusters (Figure S16a–c). Compared to IRs, INRs exhibited a higher proportion of B_c01-TCL1A cells and lower proportions of B_c03-CD1C cells (Figure S16d). scGeneANOVA results also reflected changes in these cell subpopulations (Figure S16e). We found that B cells from INRs significantly overexpressed naive characteristic genes such as PLPP5, BTG1, YBX3, and FCRL1, while underexpressing functional genes like CD27 and TNFRSF13B (Figure S16g). This aligns with previous findings of dysregulated memory B cells in INRs.[21,42]

To understand B cell subpopulation dynamics, we conducted RNA velocity analysis, revealing a trajectory from naive to memory B cells (Figure S17a). Using STARTRAC, we analysed BCR transitions and found that the transition index from B_c01-TCL1A to B_c02-PLD4 was significantly lower in INRs compared to IRs, suggesting developmental deficiencies in naive B cells (Figure S17b–d). Correlation analysis showed that the proportion of UsmB (B_c03-CD1C) cells was negatively correlated with interferon activation (Pearson's R = −0.4233, P = 0.0141) and positively correlated with $CD4^+$ T cell proportions (Pearson's R = 0.3704, P = 0.0339) (Figure S17e). These findings suggest that B cell developmental impairment, along with interferon

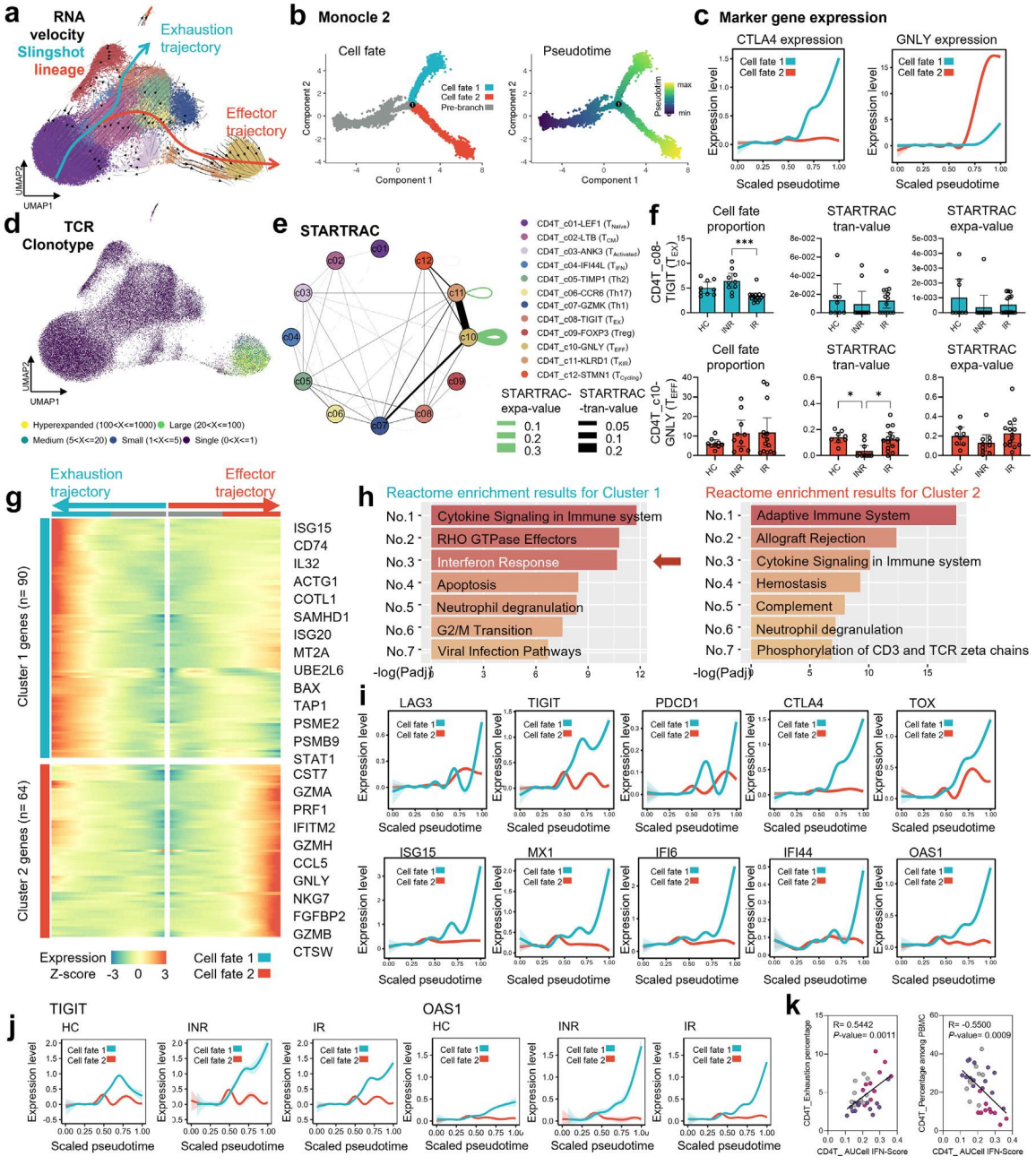

Fig. 4: **Exhaustion tendency and IFN signatures in CD4$^+$ T cells of INRs**. (a) RNA velocity plot with Slingshot lineage tracing showing the exhaustion and effector trajectories of CD4$^+$ T cells. Each cell is colour-coded according to its predicted trajectory. (b) Trajectory analysis using Monocle2, depicting cell fate determination (left) and pseudotime (right) of CD4$^+$ T cells along the exhaustion and effector lineage. (c) Expression of selected marker genes (e.g., CTLA4 and GNLY). during the increased pseudotime along two different lineages. (d) UMAP plots showing the distribution of TCR clonotypes within CD4$^+$ T cells. Each clonotype is colour-coded and labelled accordingly. (e) Network diagram illustrating the TCR information among different CD4$^+$ T cell clusters, with self-lines representing expansion score and the interaction-lines representing transition score. (f) Bar plots comparing the percentage, the STARTRAC expansion score and the STARTRAC transition score of CD4T_c08-TIGIT (T_{EX}) and CD4T_c10-GNLY (T_{EFF}) across HC, INR, and IR groups. Statistical significance was determined using ANOVA followed by Tukey post-hoc tests (*P < 0.05, **P < 0.01, ***P < 0.001). (g) Heatmap displaying the expression dynamics of key genes along the exhaustion and effector trajectories in CD4$^+$ T cells. Each row represents a gene, and colour intensity indicates expression levels. (h) Reactome pathway enrichment results for two major clusters (Cluster 1: exhaustion trajectory; Cluster 2: effector trajectory) in CD4$^+$ T cells. (i and j) Line plots showing the expression levels of selected genes along the pseudotime in CD4$^+$ T cells. Lines represent different cell fates (Cell fate 1: exhaustion, Cell fate 2: effector). (k) Scatter plots showing the correlation between exhaustion percentage of CD4$^+$ T cells and PBMC AUCell interferon score (left) and CD4$^+$ T cell percentages (right) with the respective correlation coefficients and P-values.

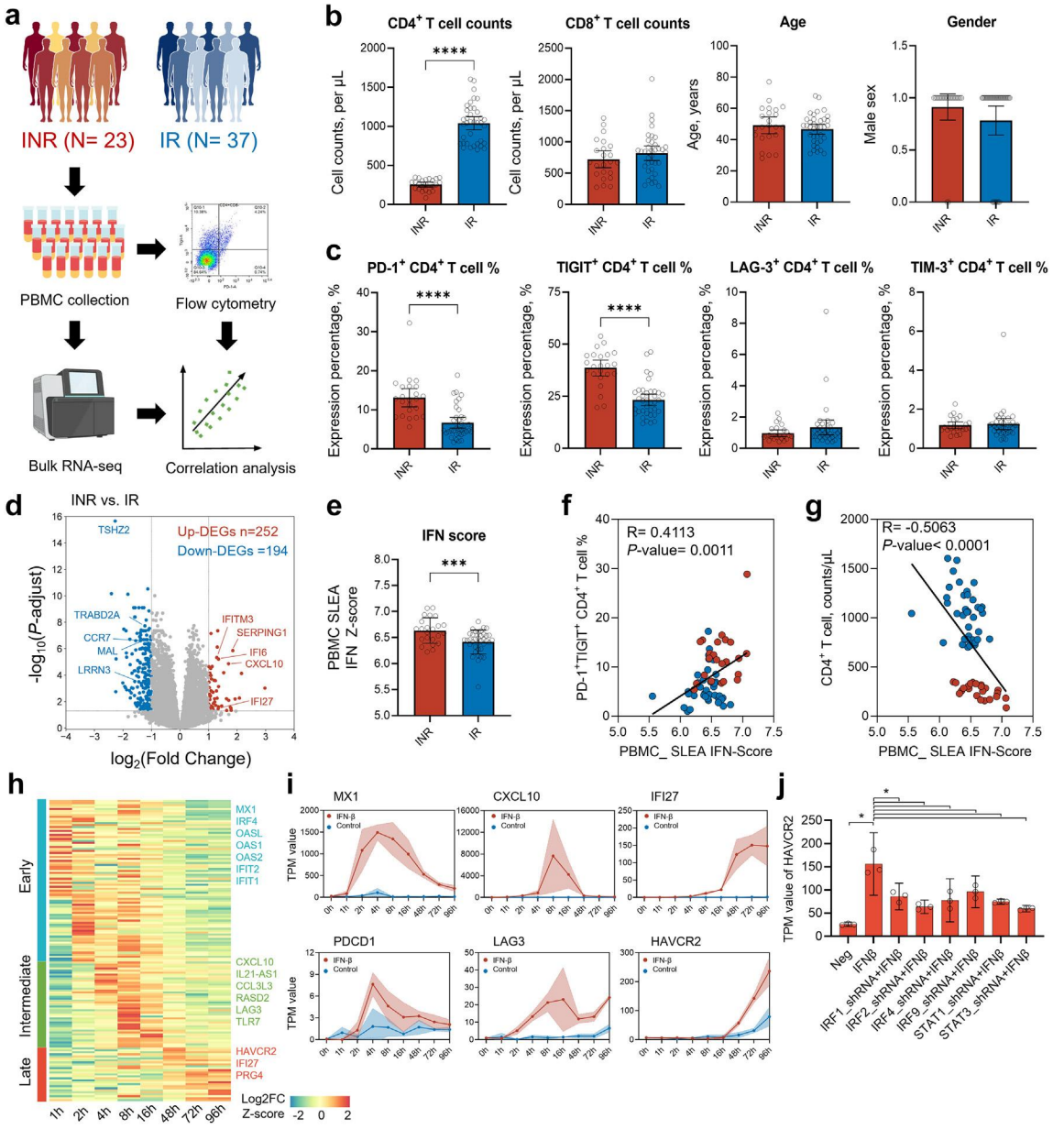

Fig. 5: **Immune dysfunction in INRs associated with IFN signalling and CD4$^+$ T cell exhaustion**. (a) Schematic of the study design. Peripheral blood mononuclear cells (PBMCs) were collected from 23 INRs and 37 immunological responders (IRs). Flow cytometry and bulk RNA-seq were performed, followed by correlation analysis. (b) Comparison of CD4$^+$ and CD8$^+$ T cell counts, age, and gender differences between INRs and IRs. (c) Comparison of the expression of PD-1, TIGIT, LAG-3, and TIM-3 in CD4$^+$ T cells between INRs and IRs. (d) Volcano plot depicting differentially expressed genes (DEGs) between INRs and IRs. (e) Comparison of the Sample-Level Enrichment Analysis (SLEA) Z-score for IFN genes between INRs and IRs. (f) Scatter plots showing the correlation between PBMC IFN score and exhaustion percentage of CD4$^+$ T cells. (g) Scatter plots showing the correlation between PBMC IFN score and CD4$^+$ T cell counts. For panels (f) and (g), the red dots represent for INRs, and bule dots represent for IRs. (h) Heatmap showing the expression patterns of DEGs across early, intermediate, and late time points in IFN-β-treated human primary CD4$^+$ T cells. (i) Temporal expression profiles of selected interferon-stimulated genes (ISGs) (MX1, CXCL10, IFI27) and exhaustion markers (PDCD1, LAG3, HAVCR2) across different time points. Red represents the IFN-β-treated group, and blue represents the control group. (j) shRNA-mediated knockdown of IRF and STAT family transcription factors demonstrating the reduction of HAVCR2 expression following IFN-β treatment. For panels (h), (i), (j), the transcriptomic raw data were obtained from GSE195542. For panels (b), (c), (e), and (j), data are represented as mean ± SD. Group differences were analysed using an unpaired t-test with two-tailed P-values. Significance levels are indicated as follows: *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001.

activation, is linked to incomplete immune reconstitution.

Higher HIV-1 viral transcripts and heightened IFN response in INRs: insights from VILDA analysis

The IFN response is vital for host defence against viral infections, primarily through the induction of interferon-stimulated genes (ISGs).[43] To investigate the mechanisms driving the heightened IFN response observed in INRs, we developed the VILDA tool. VILDA enables rapid and accurate detection of transcribed reads from up to 14,328 viral types using bulk RNA-seq data (Fig. 6a). We utilized the VILDA tool to analyse our previously published bulk RNA-seq datasets.[44] Using this approach, we identified a strong correlation between SIV viral transcript frequencies derived from VILDA and plasma SIV viral loads measured by RT-qPCR in rhesus monkeys (Figure S18).

In this current analysis, VILDA revealed that an average of 96.95% of reads from peripheral blood transcriptome sequencing fastq files were successfully mapped to the human reference genome, with the remaining 3.05% representing unmapped data (Fig. 6b). Secondary analysis of these unmapped reads identified 308 viral types, with 10 meeting our criteria for high read count and low dropout rate (Fig. 6c and Table S12). These included the BeAn 58,058 virus, human endogenous retrovirus K113 (HERV-K113), Esparto virus, Emiliana huxleyi virus 86, Choristoneura fumiferana granulovirus, Ictalurid herpesvirus 1, Escherichia phage DE3, Enterobacteria phage P7, Parainfluenza virus 5, and HIV-1 virus.

When we integrated VILDA results with SLEA for pathway enrichment, five viruses were found to be significantly associated with IFN pathway activation: BeAn 58,058 virus, HIV-1, Escherichia phage DE3, HERV-K113, and Enterobacteria phage P7 (Fig. 6d). Notably, BeAn 58,058 virus and HIV-1 showed a positive correlation with IFN activation, while the others exhibited a negative correlation. Further analysis across different patient groups revealed that INRs had significantly higher frequencies of BeAn 58,058 virus and HIV-1 transcripts compared to HCs (Fig. 6e). Among these viruses, only HIV-1 had higher expression levels in INRs compared to IRs, and it was exclusively detected in PLWH, with no detection in HCs.

Given the low plasma viral load in PLWH patients on long-term ART, the detection of HIV-1 transcripts by VILDA likely reflects the presence of the HIV-1 reservoir within PBMCs rather than circulating viral particles in plasma. qPCR validation further confirmed that INRs have higher levels of total HIV-1 total DNA and caRNA per $CD4^+$ T cell compared to IRs (Fig. 6f). Additionally, there were significant positive correlations between SLEA IFN gene set scores and levels of HIV-1 total DNA (Pearson's R = 0.6443, P-value = 0.0007) and HIV-1 caRNA (Pearson's R = 0.6669, P-value = 0.0004) (Fig. 6g, Figure S19). These findings suggest that the size and transcriptional activity of the HIV-1 reservoir are closely linked to IFN signalling activation in INRs, with higher levels of HIV-1 transcripts likely driving the enhanced IFN response observed in these patients (Fig. 6h).

Discussion

This study provides a detailed single-cell resolution map of peripheral immune cells in INRs. Previous research identified the elevated IFN signalling pathway in INR patients.[13,14,45] In the clinical phase II trial, higher dosage of LLDT-8 treatment notably resulted in increased $CD4^+$ T cell counts and reduced IFN-γ-induced protein 10 (IP-10) levels.[46] Building upon these findings, this study employs scRNA/VDJ-seq to further validate the association between the IFN response and $CD4^+$ T cell depletion, and delving deeper into the underlying mechanisms. The key findings of this study include the enhanced IFN response signature in $CD4^+$ T cells of INRs, which correlated strongly with T cell exhaustion markers. Our trajectory and pseudotime analyses further delineated the developmental paths of $CD4^+$ T cells in INRs, highlighting the progression towards an exhausted state driven by IFN response. The dysregulated IFN response is likely to underly impaired immune recovery in INRs despite ART. Our results emphasized the pathogenetic role of IFN response, providing insights for potential therapeutic targets of INRs.

Questions remained that which specific interferon signature has a more profound impact on incomplete immune reconstitution in INRs. In this study, the SLEA score primarily reflects Type I IFN characteristics, and our findings reveal a stronger association between $CD4^+$ T cell exhaustion and canonical Type I ISGs. However, these results should not overshadow the potential contributions of Type II IFN signatures. Previous studies have demonstrated a close correlation between higher plasma IFN-γ levels and lower $CD4^+$ T cell counts in PLWH.[47] Given the intricate crosstalk between the downstream signalling pathways of Type I and Type II IFNs,[48,49] further experiments are essential to elucidate their respective roles and underlying causal mechanisms in immune reconstitution failure in INRs.

$CD4^+$ T cell exhaustion is a multifactorial and dynamic process, rather than being solely driven by IFN signalling.[50] Using publicly available data, we demonstrated a strong association between IFN signalling and the expression of PDCD1 and HAVCR2 in $CD4^+$ T cells *in vitro*. However, flow cytometry analysis in patient samples revealed a significant difference only in PD-1 (encoded by PDCD1) expression, while TIM-3 (encoded by HAVCR2) showed no significant difference. This observation suggests the presence of a more complex gene regulatory network in patients. In addition to IFN

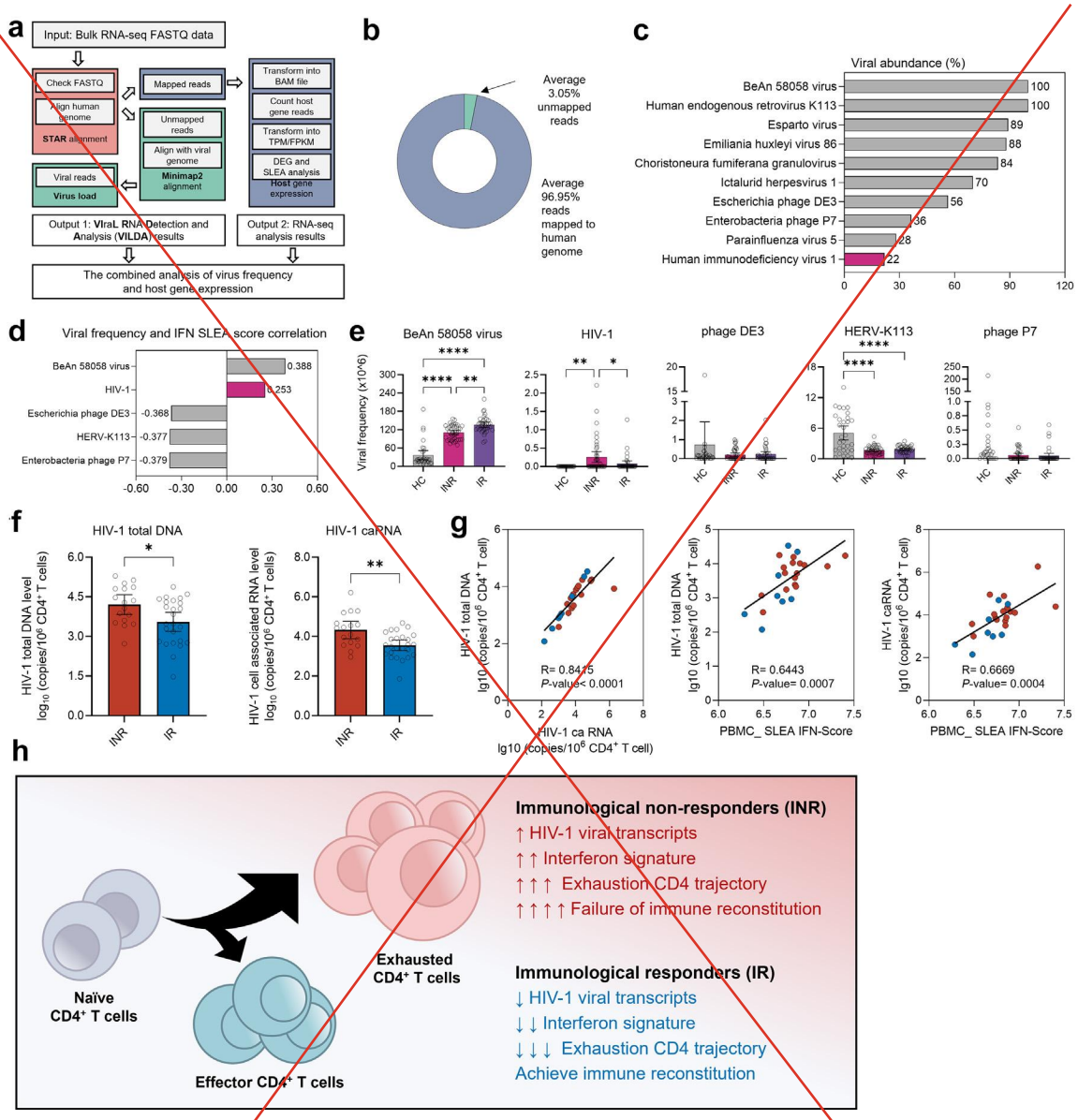

Fig. 6: **Viral identification using VILDA and correlation with IFN expression profile in INRs**. (a) Schematic workflow of the Viral Identification and Load Detection Analysis (VILDA) tool. Bulk RNA-seq FASTQ data were processed to map reads to the human genome using STAR alignment, with unmapped reads subsequently aligned to viral genomes using Minimap2. The analysis outputs included both viral identification and host gene expression, allowing for combined analysis of viral frequency and host gene expression. (b) Pie chart showing the average percentage of reads mapped to the human genome (96.95%) and unmapped reads (3.05%) from the RNA-seq data. (c) Bar chart displaying the viral abundance (percentage) of the top 10 viruses detected by VILDA across all samples. (d) Correlation between viral frequency and IFN SLEA (Sample- Level Enrichment Analysis) Z-score, showing that BeAn 58,058 virus and HIV-1 positively correlate with IFN signalling, while Escherichia phage DE3, HERV-K113, and Enterobacteria phage P7 negatively correlate with IFN signalling. (e) Comparison of viral frequency among healthy controls (HCs), INRs, and immunological responders (IRs) for BeAn 58,058 virus, HIV-1, Escherichia phage DE3, HERV-K113, and Enterobacteria phage P7. (f) Quantification of HIV-1 total DNA and cell-associated RNA (caRNA) levels in INRs and IR. Data are represented as mean ± SD. Group differences were analysed using an unpaired t-test with two-tailed P-values. Significance levels are indicated as follows: *P < 0.05, **P < 0.01, ***P < 0.001, and ****P < 0.0001. (g) Scatter plots showing the correlation between HIV-1 total DNA, HIV-1 caRNA levels, and PBMC_SLEA IFN-Score. The red dots represent for INRs, and bule dots represent for IRs. (h) Graphical scheme of revealing the role of HIV-1 transcripts and IFN signature in $CD4^+$ T cell differentiation of INRs.

responses, we identified several pathways, including hypoxia and PI3K/AKT/mTOR signalling, as significantly associated with $CD4^+$ T cells. These pathways have been described as associated with the $CD4^+$ T cell exhausted phenotype.[51,52] The significant association of these pathways with $CD4^+$ T cells suggests a potential interplay between metabolic and signalling dysregulation and the maintenance of T cell exhaustion in INRs.[6] These findings raise the possibility that the activation of the IFN response may not be the sole or primary driver of immune reconstitution failure. Instead, it could represent a downstream consequence of broader dysregulated metabolic and signalling pathways.[53]

Furthermore, our results show that while myeloid cells and NK cells in INRs do not show significant differences in cell proportions, their activation and functional indicators are markedly elevated. Although $CD8^+$ T cells in INRs also show significant enrichment of IFN response pathways, their exhaustion levels are not significantly higher. Instead, these cells demonstrate a notable enhancement in MHC-I-mediated cell interactions. Recent studies further support these findings, showing that total $CD8^+$ T cells in INRs exhibit significantly increased expression of HLA-DR and CD38, indicative of activation, but only a limited increase in $PD1^+$ subpopulations.[10] Combined with our another study, we hypothesize that the enhanced MHC-I expression in $CD8^+$ T cells of INRs might reflect the persistence of HIV-1 reservoir cells.[54] Further studies are needed to evaluate the specific functional roles of enhanced MHC-I expression of $CD8^+$ T cells.

Additionally, we observed developmental impairments in B cells of INRs, consistent with recent findings from other studies.[21,34–38,42] Notably, our analysis did not identify a separate Breg cell cluster, potentially due to their inclusion within the UsmB (B_c03-CD1C) subpopulation, as Breg also exhibit high CD1C expression.[55] Breg cells are known to regulate IFN-γ production of $CD4^+$ T cells, contributing to immune suppression and maintaining immune tolerance.[56] This immunoregulatory role may explain the observed negative correlation between the frequency of UsmB (B_c03-CD1C) cells and IFN scores in our study. Further studies are warranted to delineate the specific role of Breg cells and their interactions with other immune cell populations in INRs. Taken together, these phenomena further underscore that INRs exhibit widespread immune cell developmental impairments. Their immune dysfunction is not limited to the failure of $CD4^+$ T cell reconstitution but includes extensive immune activation and functional impairments.

This study also developed two bioinformatic tools, including scGeneANOVA and VILDA. The application of our mixed model differential gene analysis approach, scGeneANOVA, revealed a unique set of differentially expressed genes and pathways, emphasizing the altered immune landscape in INRs. Through our analysis, we found that scGeneANOVA is more suitable for this study compared to Seurat's Wilcoxon algorithm and the pseudo-bulk edgeR-LRT method, as it better captures the unique and accurate gene expression characteristics in INRs. However, the application of scGeneANOVA requires attention to two key considerations. First, the method assumes that gene expression data approximates a normal distribution, which necessitate prevalidation or transformation of the count data. Second, scGeneANOVA is better suited for large-scale single-cell cohorts with experimental designs involving three or more groups. Under such conditions, it effectively reduces Type I errors and enhances analytical accuracy. These features make it particularly advantageous for complex experimental designs and large-scale single-cell data analyses. Therefore, we recommend combining scGeneANOVA with different DEG analysis methods to provide a more comprehensive analysis, and identify truly important and critical gene expression patterns.

In addition, we developed the VILDA tool to explore the mechanisms driving the heightened IFN response observed in INRs. VILDA enabled us to rapidly and accurately detect transcribed reads from up to 14,328 viral types using bulk RNA-seq data. Our analysis revealed that INRs harbour higher frequencies of viral transcripts, particularly HIV-1, which were positively correlated with IFN activation levels. These findings suggest that the elevated IFN response in INRs may be driven by persistent HIV-1 viral reservoirs within PBMCs. Similar studies conducted in different countries have also reported higher levels of viral reservoirs in INRs, indicating this may be a common characteristic of these patients.[57,58] This further implicates viral persistence in the immune dysfunction observed in INRs and highlights the potential of targeting viral reservoirs as a therapeutic strategy.

Interestingly, our analysis also detected 100% prevalence rates of BeAn 58,058 virus (a type of poxvirus) and human endogenous retrovirus K113 (HERV-K113) across all samples, consistent with findings from previous studies.[59–62] These results reinforce the notion that certain viruses, like HERV-K113, are ubiquitously present in human populations and may play a role in modulating immune responses. The consistent detection of these viruses across different studies underscores the reliability of VILDA in identifying viral transcripts. Besides, VILDA could be a valuable tool for studying viral dynamics in various species, not just humans. However, it should be noted that further validation of the viral candidates identified by VILDA is essential. Techniques such as PCR and WB are necessary to confirm the transcriptional or translational activity of these viruses, thereby strengthening the biological relevance of the findings.

This study has several limitations. First, a key consideration in interpreting our findings is the inherent association between baseline $CD4^+$ T cell

depletion and immune reconstitution failure. In our cohort, nearly all INR patients had very low baseline $CD4^+$ T cell counts. However, some IR patients with low baseline $CD4^+$ T cell counts (e.g., IR6 and IR9) still achieved successful $CD4^+$ T cell recovery, suggesting that factors beyond low baseline counts—such as chronic IFN activation and viral persistence—may also contribute to immune exhaustion in INR patients. Second, our scRNA-seq analysis was limited to peripheral blood immune cells and was conducted as a cross-sectional study, which lacks the temporal dimension necessary to fully capture the dynamics of immune cell changes over time. Access to tissue samples, such as lymph nodes, and longitudinal tracking could provide a more comprehensive understanding of IFN's impact on $CD4^+$ T cell development. Third, while we integrated scVDJ-seq data to explore immune repertoire diversity, our study lacks spatial context and chromatin accessibility information, which are crucial for understanding the positional relationships of interacting cell types and the regulatory mechanisms driving immune dysfunction. Future studies should consider the application of spatial transcriptomics and scATAC-seq to elucidate these aspects and further investigate the role of IFN-related transcription factors in $CD4^+$ T cell state remodelling. Finally, while VILDA identified higher levels of viral transcripts in INRs, particularly HIV-1, our study's scope was limited by the relatively small sample size of our bulk RNA-seq cohort. As a result, our findings should be considered exploratory and warrant further validation in larger, prospective cohorts. Additionally, integrating VILDA with other virome-wide approaches could provide a more detailed picture of the viral landscape in INRs and its contribution to immune dysfunction.

In conclusion, this study provides a comprehensive understanding of the mechanisms driving immune dysfunction in INRs, particularly highlighting the potential role of chronic IFN signalling and viral persistence in $CD4^+$ T cell exhaustion and incomplete immune reconstitution. Our study uncovers potential biomarkers and therapeutic targets, offering new avenues for improving immune recovery in this vulnerable population. Future research should prioritize validating these biomarkers and exploring therapeutic strategies to modulate the IFN response, aiming to mitigate T cell exhaustion and enhance immune reconstitution in INRs. Together, these findings offer new insights into the mechanisms underlying immune dysfunction in INRs, emphasizing the potential role of chronic IFN signalling and viral persistence in driving $CD4^+$ T cell exhaustion and incomplete immune reconstitution. This study not only advances our understanding of INR pathophysiology but also lays the groundwork for developing targeted therapeutic strategies to improve immune recovery in this vulnerable population.

Contributors

Conceptualization, Xiaosheng Liu. Methodology, Xiaosheng Liu. Software, Xiaosheng Liu. Validation, Xiaosheng Liu. Formal analysis, Xiaosheng Liu. Investigation, Xiaosheng Liu. Resources, Xiaosheng Liu, Leidan Zhang, Xiaodi Li, Ling Chen, Lianfeng Lu, Yang Yang, Yuanni Wu, Liyuan Zheng, Jia Tang, Fada Wang, Yang Han, Xiaojing Song. Data Curation, Xiaosheng Liu. Writing - Original Draft, Xiaosheng Liu. Writing - Review & Editing, Xiaosheng Liu, Wei Cao, Taisheng Li. Visualization, Xiaosheng Liu. Supervision, Xiaosheng Liu, Taisheng Li. Project administration, Xiaosheng Liu, Taisheng Li. Funding acquisition, Taisheng Li. Xiaosheng Liu and Taisheng Li have accessed and verified the data, and Xiaosheng Liu and Taisheng Li were responsible for the decision to submit the manuscript.

Data sharing statement

Single cell RNA-seq data reported in the present paper have been deposited in the National Genomics Data Center (accession number HRA008420).

Declaration of interests

The authors have declared that no conflict of interest exists.

Acknowledgements

We thank Prof. Liguo Zhang from the Key Laboratory of Infection and Immunity, Institute of Biophysics, Chinese Academy of Sciences, Prof. Linqi Zhang from the School of basic medical sciences, Tsinghua University, Prof. Yuxian He from the Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, for their research support and guidance. We also thank Ms. Lina Wang for her assistance with financial auditing and reimbursement. We especially thank all the PLWH participants from the AIDS research center, Peking Union Medical College Hospital.

During the preparation of this work, the authors used ChatGPT-4o to checking code and manuscript. After using this tool or service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105667.

References

1 Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384(9939):258–271.
2 Doitsh G, Greene WC. Dissecting how CD4 T cells are lost during HIV infection. *Cell Host Microbe*. 2016;19(3):280–291.
3 Li TS, Tubiana R, Katlama C, Calvez V, Ait Mohand H, Autran B. Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. *Lancet*. 1998;351(9117):1682–1686.
4 Autran B, Carcelain G, Li TS, et al. Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science*. 1997;277(5322):112–116.
5 Yan L, Xu K, Xiao Q, et al. Cellular and molecular insights into incomplete immune recovery in HIV/AIDS patients. *Front Immunol*. 2023;14:1152951.
6 Zhang W, Ruan L. Recent advances in poor HIV immune reconstitution: what will the future look like? *Front Microbiol*. 2023;14:1236460.
7 Yang X, Su B, Zhang X, Liu Y, Wu H, Zhang T. Incomplete immune reconstitution in HIV/AIDS patients on antiretroviral therapy: challenges of immunological non-responders. *J Leukoc Biol*. 2020;107(4):597–612.
8 Young J, Psichogiou M, Meyer L, et al. CD4 cell count and the risk of AIDS or death in HIV-Infected adults on combination antiretroviral therapy with a suppressed viral load: a longitudinal cohort study from COHERE. *PLoS Med*. 2012;9(3):e1001194.
9 Liu J, Ding C, Shi Y, et al. Advances in mechanism of HIV-1 immune reconstitution failure: understanding lymphocyte subpopulations and interventions for immunological nonresponders. *J Immunol*. 2024;212(11):1609–1620.

10 Vos W, Navas A, Meeder EMG, et al. HIV immunological non-responders are characterized by extensive immunosenescence and impaired lymphocyte cytokine production capacity. *Front Immunol*. 2024;15:1350065.
11 Zhao J, Schank M, Wang L, et al. Mitochondrial functions are compromised in CD4 T cells from ART-controlled PLHIV. *Front Immunol*. 2021;12:658420.
12 Lorvik KB, Meyer-Myklestad MH, Kushekar K, et al. Enhanced gut-homing dynamics and pronounced exhaustion of mucosal and blood CD4(+) T cells in HIV-infected immunological non-responders. *Front Immunol*. 2021;12:744155.
13 Liu X, Lin L, Lu L, et al. Comparative transcriptional analysis identified characteristic genes and patterns in HIV-infected immunological non-responders. *Front Immunol*. 2022;13:807890.
14 Tang S, Lu Y, Sun F, et al. Transcriptomic crosstalk between viral and host factors drives aberrant homeostasis of T-cell proliferation and cell death in HIV-infected immunological non-responders. *J Infect*. 2024;88(5):106151.
15 Zhang Y, Ji J, Xie K, et al. Pathological proliferation: a potential mechanism for poor CD4(+) T cell recovery in people living with HIV. *Front Cell Infect Microbiol*. 2024;14:1344778.
16 Bono V, Augello M, Tincati C, Marchetti G. Failure of CD4+ T-cell recovery upon virally-effective cART: an enduring gap in the understanding of HIV+ immunological non-responders. *New Microbiol*. 2022;45(3):155–172.
17 Wang XM, Zhang JY, Xing X, et al. Global transcriptomic characterization of T cells in individuals with chronic HIV-1 infection. *Cell Discov*. 2022;8(1):29.
18 Kazer SW, Aicher TP, Muema DM, et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. *Nat Med*. 2020;26(4):511–518.
19 Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol*. 2018;18(1):35–45.
20 Li H, Tang Y, Wang Y, et al. Single-cell sequencing resolves the landscape of immune cells and regulatory mechanisms in HIV-infected immune non-responders. *Cell Death Dis*. 2022;13(10):849.
21 Jia J, Zhao Y, Yang JQ, et al. Naïve B cells with low differentiation improve the immune reconstitution of HIV-infected patients. *iScience*. 2022;25(12):105559.
22 Wen W, Su W, Tang H, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov*. 2020;6:31.
23 Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods*. 2019;16(12):1289–1296.
24 Zhang L, Yu X, Zheng L, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. 2018;564(7735):268–272.
25 Zhang C, Li J, Cheng Y, et al. Single-cell RNA sequencing reveals intrahepatic and peripheral immune characteristics related to disease phases in HBV-infected patients. *Gut*. 2023;72(1):153–167.
26 Street K, Risso D, Fletcher RB, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics*. 2018;19(1):477.
27 Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun*. 2021;12(1):1088.
28 Squair JW, Gautier M, Kathe C, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun*. 2021;12(1):5692.
29 Blondin-Ladrie L, Fourcade L, Modica A, et al. Monocyte gene and molecular expression profiles suggest distinct effector and regulatory functions in Beninese HIV highly exposed seronegative female commercial sex workers. *Viruses*. 2022;14(2):361.
30 Cano-Ortiz L, Luedde T, Münk C. HIV-1 restriction by SERINC5. *Med Microbiol Immunol*. 2023;212(2):133–140.
31 Li D, Rawle DJ, Wu Z, et al. eEF1A demonstrates paralog specific effects on HIV-1 reverse transcription efficiency. *Virology*. 2019;530:65–74.
32 Liu Q, Lusso P. Integrin α4β7 in HIV-1 infection: a critical review. *J Leukoc Biol*. 2020;108(2):627–632.
33 Carpenter S, O'Neill LAJ. From periphery to center stage: 50 years of advancements in innate immunity. *Cell*. 2024;187(9):2030–2051.
34 Zhang Z, Cheng L, Zhao J, et al. Plasmacytoid dendritic cells promote HIV-1-induced group 3 innate lymphoid cell depletion. *J Clin Invest*. 2015;125(9):3692–3703.
35 Li G, Cheng M, Nunoya J, et al. Plasmacytoid dendritic cells suppress HIV-1 replication but contribute to HIV-1 induced immunopathogenesis in humanized mice. *PLoS Pathog*. 2014;10(7):e1004291.
36 Luo Z, Li Z, Martin L, et al. Increased natural killer cell activation in HIV-infected immunologic non-responders correlates with CD4+ T cell recovery after antiretroviral therapy and viral suppression. *PLoS One*. 2017;12(1):e0167640.
37 Zhang QY, Zhang X, Su B, et al. Increased early activation of CD56dimCD16dim/- natural killer cells in immunological non-responders correlates with CD4+ T-cell recovery. *Chin Med J*. 2020;133(24):2928–2939.
38 Qian F, Hu S, Zhu Y, et al. CD56(dim) NK cell is an important factor in T cell depletion of cART-treated AIDS patients. *Int J Gen Med*. 2022;15:4575–4583.
39 Collora JA, Liu R, Pinto-Santini D, et al. Single-cell multiomics reveals persistence of HIV-1 in expanded cytotoxic T cell clones. *Immunity*. 2022;55(6):1013–1031.e7.
40 Yasumizu Y, Takeuchi D, Morimoto R, et al. Single-cell transcriptome landscape of circulating CD4(+) T cell populations in autoimmune diseases. *Cell Genom*. 2024;4(2):100473.
41 Sumida TS, Dulberg S, Schupp JC, et al. Type I interferon transcriptional network regulates expression of coinhibitory receptors in human T cells. *Nat Immunol*. 2022;23(4):632–642.
42 Liu Y, Li Z, Lu X, et al. Dysregulation of memory B cells and circulating T follicular helper cells is a predictor of poor immune recovery in HIV-infected patients on antiretroviral therapy. *J Med Virol*. 2023;95(2):e28559.
43 Mesev EV, LeDesma RA, Ploss A. Decoding type I and III interferon signalling during viral infection. *Nat Microbiol*. 2019;4(6):914–924.
44 Liu X, Lv T, Li X, et al. Comprehensive transcriptomic analyses identify the immunosuppressive effects of LLDT-8 in ART-treated SIV-infected rhesus macaques. *Int Immunopharmacol*. 2024;126:111173.
45 Xin Zhang LH, Zhang R, Sun H, Zhang L, Cong J. Enhanced expression of two ISGs as potential biomarkers of prognosis in HIV-infected immunological non-responders. *Zoonoses*. 2024;4(1):975.
46 Cao W, Liu X, Han Y, et al. (5R)-5-hydroxytriptolide for HIV immunological non-responders receiving ART: a randomized, double-blinded, placebo-controlled phase II study. *Lancet Reg Health West Pac*. 2023;34:100724.
47 Watanabe D, Uehira T, Suzuki S, et al. Clinical characteristics of HIV-1-infected patients with high levels of plasma interferon-γ: a multicenter observational study. *BMC Infect Dis*. 2019;19(1):11.
48 Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol*. 2005;5(5):375–386.
49 Hopfner KP, Hornung V. Molecular mechanisms and cellular functions of cGAS-STING signalling. *Nat Rev Mol Cell Biol*. 2020;21(9):501–521.
50 Miggelbrink AM, Jackson JD, Lorrey SJ, et al. CD4 T-cell exhaustion: does it exist and what are its roles in cancer? *Clin Cancer Res*. 2021;27(21):5742–5752.
51 Liu S, Liu X, Zhang C, Shan W, Qiu X. T-cell exhaustion status under high and low levels of hypoxia-inducible factor 1α expression in glioma. *Front Pharmacol*. 2021;12:711772.
52 Wu H, Zhao X, Hochrein SM, et al. Mitochondrial dysfunction promotes the transition of precursor to terminally exhausted T cells through HIF-1α-mediated glycolytic reprogramming. *Nat Commun*. 2023;14(1):6858.
53 Roman J, Rangasamy T, Guo J, et al. T-cell activation under hypoxic conditions enhances IFN-gamma secretion. *Am J Respir Cell Mol Biol*. 2010;42(1):123–128.
54 Zhu T, Cao W, Li T. HIV DNA positively correlates with HLA-DR+CD8+ T lymphocytes over 8-year suppressive antiretroviral therapy. *AIDS*. 2023;37(8):1335–1337.
55 Bao Y, Liu J, Li Z, et al. Ex vivo-generated human CD1c(+) regulatory B cells by a chemically defined system suppress immune responses and alleviate graft-versus-host disease. *Mol Ther*. 2024;32(12):4372–4382.
56 Appelgren D, Puli S, Hellmark T, et al. Regulatory B cells are reduced in the blood in patients with granulomatosis with polyangiitis, and fail to regulate T-cell IFN-γ production. *Clin Exp Immunol*. 2023;213(2):190–201.

57 Zhang LX, Song JW, Zhang C, et al. Dynamics of HIV reservoir decay and naïve CD4 T-cell recovery between immune non-responders and complete responders on long-term antiretroviral treatment. *Clin Immunol*. 2021;229:108773.

58 Scrimieri F, Bastian E, Smith M, et al. Transcriptionally active defective HIV-1 proviruses and their association with immunological nonresponse to antiretroviral therapy. *J Infect Dis*. 2024;229(6):1786–1790.

59 Ali MJ. Alterations of lacrimal sac microbiota in failed dacryocystorhinostomy: the Lacriome paper 6. *Semin Ophthalmol*. 2024;39(4):324–329.

60 Goolam Mahomed T, Peters RPH, Allam M, et al. Lung microbiome of stable and exacerbated COPD patients in Tshwane, South Africa. *Sci Rep*. 2021;11(1):19758.

61 Villesen P, Aagaard L, Wiuf C, Pedersen FS. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology*. 2004;1:32.

62 Jha AR, Pillai SK, York VA, et al. Cross-sectional dating of novel haplotypes of HERV-K 113 and HERV-K 115 indicate these proviruses originated in Africa before Homo sapiens. *Mol Biol Evol*. 2009;26(11):2617–2626.