Proviral Genetic Barcoding Tracks CD4⁺ T Cell Transitions With Novel Intact HIV-1 Provirus Clones

330

Vinie Kouamou[1], Mindy Smith[1], Francesca Scrimieri[2], Brad T. Sherman[2], Cathrine A. Rehm[1], Weizhong Chang[2], Tomozumi Imamichi[2], H. Clifford Lane[1], Hiromi Imamichi[1]

[1]National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD, USA; [2]Frederick national Laboratory for Cancer Research, Frederick, MD, USA

sponsored by the National Cancer Institute

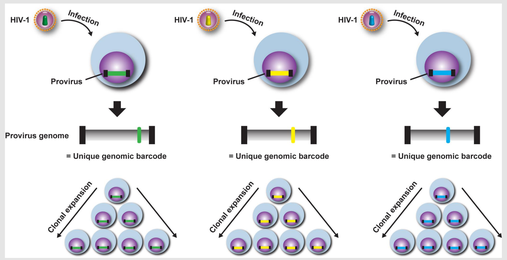
BACKGROUND

The majority of proviral DNA sequences are defective due to deletions and mutations in open reading frames. Defects in non-coding regions can also render proviruses defective; such proviruses may appear structurally intact yet potentially lack full competence. In this study, we identified such defective proviruses in two participants followed for 30 years and used them as genetic barcodes to study the dynamics and lineage relationships of CD4⁺ T cell subsets.

METHODS

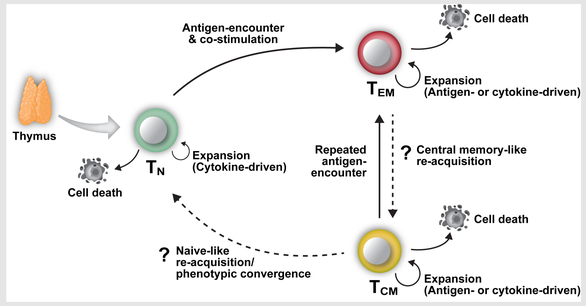
Two participants with chronic HIV-1 infection who initiated their first antiretroviral therapy (ART) regimen in 1992-1994 (pre-HAART era) were studied. Full-length HIV-1 sequencing was performed on sorted naïve, central memory, and effector memory CD4⁺ T cell subsets with longitudinal follow-up from 2008 to 2025. Expanded clones of structurally intact proviruses with defects in the 5' leader region were used as genetic barcodes to follow the transition of CD4⁺ T cells among naïve, central memory and effector memory pools. ART drug-resistance mutations (DRMs) served as temporal markers to estimate when the pool of defective proviruses was seeded.

MODEL: PROVIRAL GENOMIC BARCODES

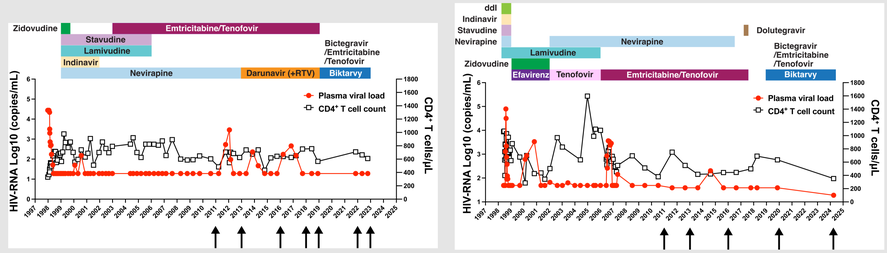


HIV-1 proviruses as stable genomic barcodes for lineage tracing. Following HIV-1 infection, each integrated provirus acquires a unique mutation pattern generated during reverse transcription, creating a permanent genomic barcode in the infected cell. This barcode is faithfully transmitted to all daughter cells during clonal expansion, enabling unambiguous identification of clonally related cells over time.

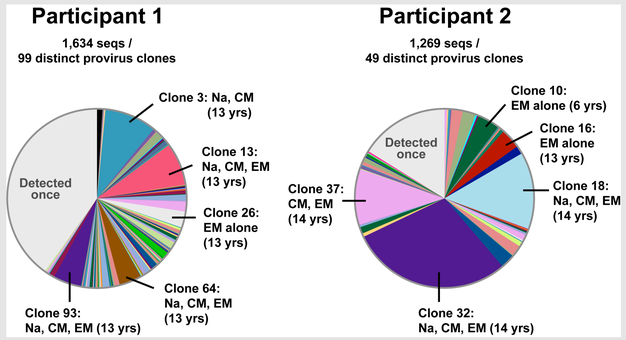
RESULTS



Human CD4⁺ T-cell differentiation, persistence, and fate. Naïve CD4⁺ T cells (T_N) arise from the thymus and undergo cytokine-driven homeostatic expansion or cell death. Antigen encounter drives differentiation into effector-memory T cells (T_{EM}), which expand, persist, transition to central-memory T cells (T_{CM}), or are lost. Repeated antigen exposure supports bidirectional transitions within the memory pool. Dashed arrows indicate hypothesized phenotypic re-acquisition or convergence pathways that remain unresolved.

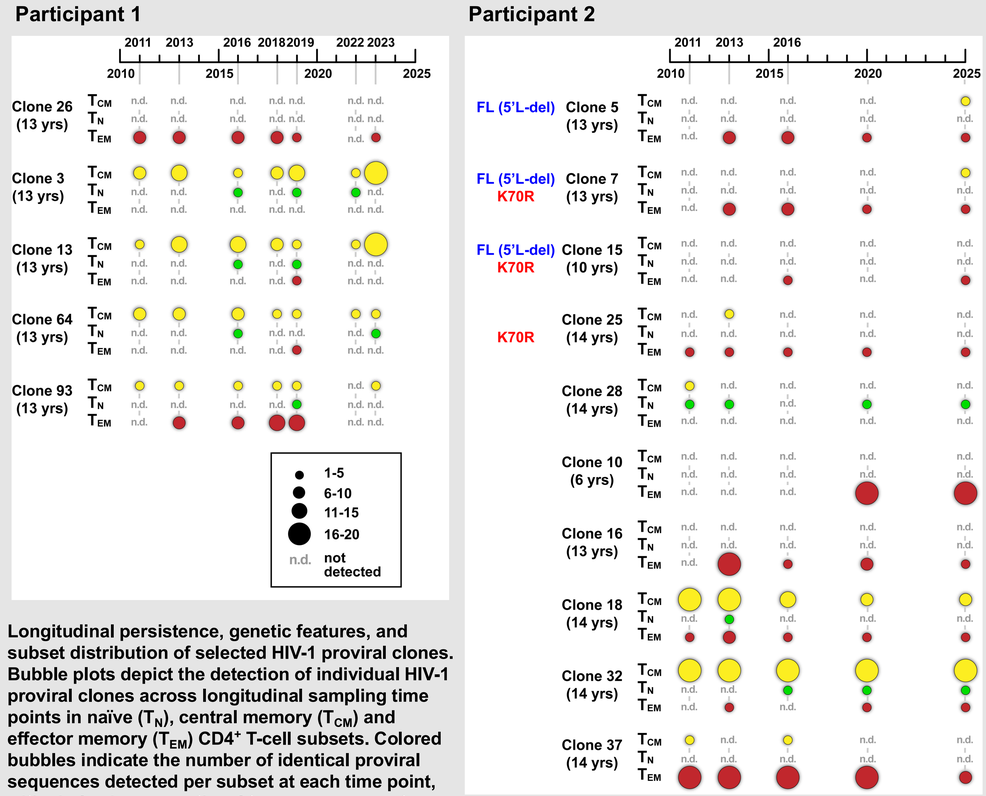


Longitudinal clinical history and sampling strategy for study participants. Plasma HIV-1 RNA levels and CD4⁺ T-cell counts are shown over time for two participants receiving suppressive antiretroviral therapy (ART). Arrows indicate time points at which peripheral blood was collected for downstream analyses. CD4⁺ T-cell subsets: naïve (T_N), central memory (T_{CM}), and effector memory (T_{EM}) were isolated by flow-cytometric sorting, followed by single-genome PCR amplification and sequencing of HIV-1 proviruses.



Distribution and persistence of clonally expanded HIV-1-infected CD4⁺ T cells. Pie charts depict the relative abundance of distinct HIV-1 proviral clones identified from longitudinally sampled CD4⁺ T-cell subsets in two participants. Each colored segment represents a unique proviral clone; grey segments indicate clones detected only once. Annotated clones highlight long-lived lineages detected across naïve (T_N), central memory (T_{CM}), and/or effector memory (T_{EM}) compartments, or restricted to T_EM, with durations indicating the span of detection (years).

RESULTS



Longitudinal persistence, genetic features, and subset distribution of selected HIV-1 proviral clones. Bubble plots depict the detection of individual HIV-1 proviral clones across longitudinal sampling time points in naïve (T_N), central memory (T_{CM}) and effector memory (T_{EM}) CD4⁺ T-cell subsets. Colored bubbles indicate the number of identical proviral sequences detected per subset at each time point, with circle size proportional to sequence counts. "FL (5'L-del)" denotes a provirus harboring a 5'-leader deletion, "K70R" denotes a provirus containing the K70R Zidovudine resistance mutation. "n.d." indicates that the clone was not detected. Clone labels specify the total duration of clone detection (years) during follow-up.

KEY TAKEAWAYS

- Integrated HIV-1 proviruses act as stable, non-diluting genomic barcodes, enabling long-term, single-cell lineage tracing in humans
- Clonally expanded CD4⁺ T-cell lineages persist for over a decade, demonstrating unexpected durability across differentiation states
- Bidirectional transitions occur among naïve, central memory, and effector memory compartments, including memory-derived clones detected within the naïve pool
- Provirus-based barcoding overcomes key limitations of conventional lineage-tracing approaches, redefining human T-cell fate and mechanisms of HIV-1 persistence during ART

Hiromi Imamichi, Ph.D., LIRID, NIAID, NIH
E-mail: himamichi@nih.gov