HIV Proviruses with Identical Sequences Arise from Cell Expansion and Infection by a Common Ancestor Virus

Sean C. Patro[1], Aurelie Niyongabo[1], Shuang Guo[2], Jason W. Rausch[3], Michael J. Bale[1], Andrew Musick[1], Xiaolin Wu[2], Liliana R. Perez-Rodriguez[4], Wei Shao[2], Eli A. Boritz[4], Steven G. Deeks[5], Frank Maldarelli[1], Stephen H. Hughes[1], John M. Coffin[6], Mary F. Kearney[1]

[1]HIV Dynamics and Replication Program, National Cancer Institute, Frederick, MD, [2]Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD, [3]Basic Research Laboratory, National Cancer Institute, Frederick, MD, [4]Virus Persistence and Dynamic Section, Vaccine Research Center, National Institute of Allergy and Infectious Diseases, Bethesda, MD, [5]Department of Medicine, University of California San Francisco, San Francisco, CA, [6]Department of Molecular Biology and Microbiology, Tufts University, Boston, MA

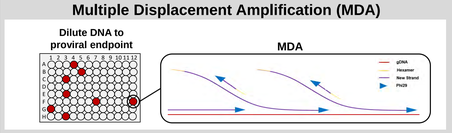
HIV Dynamics and Replication Program
Frederick National Laboratory for Cancer Research

Abstract

- HIV-1 proviruses with identical sequences can arise from the proliferation of single infected cells or from genetic bottlenecks leading to viral clones that spread prior to ART. PBMC were obtained from 5 donors on ART and analyzed by Multiple-Displacement Amplification (MDA) Single-Genome Sequencing (below) .
- We examined integration site landscape of 27 sets of identical sub-genomic sequences to determine if each
  i) had identical integration sites (cell clones)
  ii) had different integration sites (viral clones)
  iii) resulted from some with identical integration sites and some with different sites (both cell and viral clones).

Methods

Multiple Displacement Amplification (MDA)



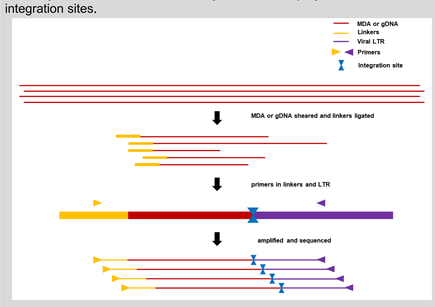
Random hexamers anneal to genomic DNA and are amplified with Phi29 for non-biased amplification. Hexamers anneal to growing strands leading to branching. New strands displace original strand.

Primary Screening and Phylogenetic Analysis

Positives are identified using nested P6-PR-RT (1.5kb) or *env* (550 bp) amplification and Sanger sequencing and analyzed with MEGA7. Neighbor-joining trees are built based on the number of nucleotide differences between sequences and are used to determine identical sub-genomic sequences.

Integration Site Sequencing

Integration site analysis (ISA) sequences the host-proviral junction and maps the integration site within the human genome, identifying distinct and identical integration sites.



Donor Characteristics

Participant identifier (PID)	Gender	Maximum pre-ART viral load[a] (c/mL)	CD4+ nadir (cells/μL)	Approximate years of infection before initial ART	Current ART regimen	Total years of ART experience	Years of viral suppression on ART at sampling	Plasma viral load on ART (c/mL)	CD4+ on ART (cells/μL)
Patient 1[b]	Male	412,033	16	unknown	TDF, FTC, RTG	13	1.0, not suppressed	<50, 134	137, 153
1683	Male	134,406	452	2.0	FTC, TDF, RTV, DRV	5.7	5.4	<40	1348
2669[c]	Male	unknown	205	≤22.0	ABC, 3TC, DTG	5.5	4.3	<40	681
3162[c]	Male	171,000	147	11.0	RTV, DRV, ABC, DTG, 3TC	20	11.2	<40	558
V-0034[d]	Female	unknown	unknown	unknown	EFV, FTC, TDF	11	11	<40, <40, <40	662, 791, 581

[a] Abbott Diagnostics RealTime HIV-1 PCR Assay
[b] Two time points, 6 months apart
[c] ART failure due to drug resistance prior to suppression on new regimen including dolutegravir
[d] Combined result of 3 time points within 36 weeks

Results: Integration Site Landscape of Identical Sub-genomic Proviral Sequences during ART

Figure 1: MDA-SGS NJ Trees

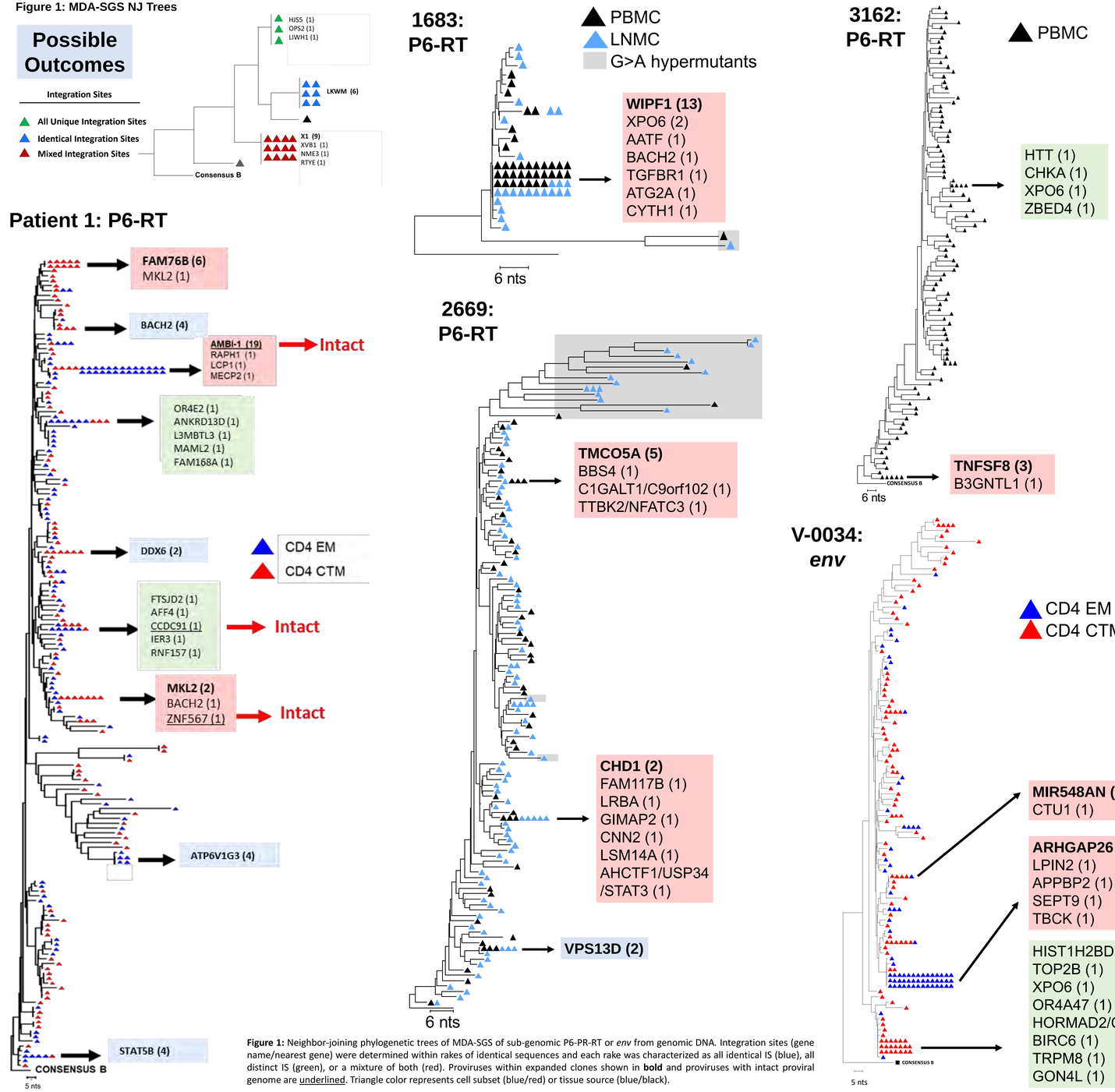


Figure 1: Neighbor-joining phylogenetic trees of MDA-SGS of sub-genomic P6-PR-RT or *env* from genomic DNA. Integration sites (gene name/nearest gene) were determined within rakes of identical sequences and each rake was characterized as all identical IS (blue), all distinct IS (green), or a mixture of both (red). Proviruses within expanded clones shown in **bold** and proviruses with intact proviral genome are underlined. Triangle color represents cell subset (blue/red) or tissue source (blue/black).

Summary

Figure 2: Integration Site Landscape within Rakes of Identical Sub-genomic sequences

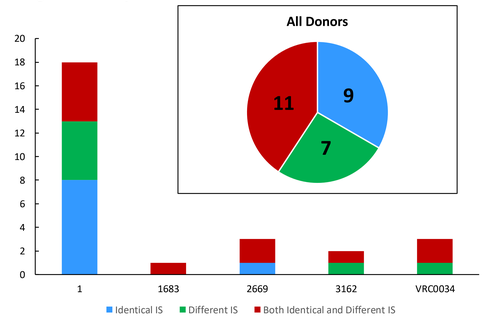
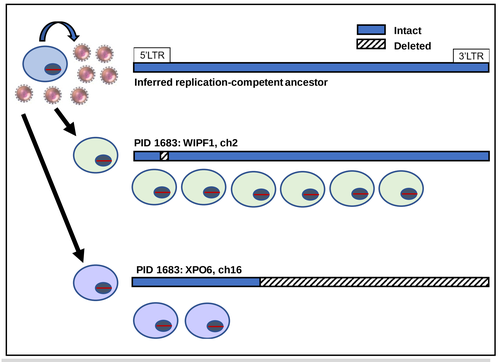


Figure 2: Integration site identity within rakes of identical sub-genomic sequences. For each PID, the number of rakes containing all identical IS (blue), all distinct IS (green) and a mixture of identical and distinct (red) are shown. Pie chart depicts the total observed rakes of identical sequences.

Viral Paleontology

Figure 3: Intact viral ancestors can be inferred from proviruses archived in clones



Conclusions

- Rakes of identical sequences can be made up of a highly expanded clone, all unique IS, or a mixture, even with the same individual. **Sub-genomic SGS is insufficient to establish IS clonality.**
- Characterization of highly-related/near-identical proviruses allows for reconstruction of historical, replication-competent viruses, such as transmitted-founder viruses.
- MDA-SGS revealed a clone (like AMBI-1) that contains a replication-competent provirus and can be targeted to intact proviruses that compose the HIV-1 reservoir donors fully suppressed on ART.

Acknowledgements: This project has been funded by intramural NCI funds to CCR, IATAP funding from OAR, and federal funds from the NCI under contract HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.