00377

Ultradeep Analysis of Pretherapy HIV predicts Genetically Complex Reservoir

Kristi Huik[1], Junko Hattori[1], Valerie Boltz[1], Jason Rausch[1], Wei Shao[2], Mary F. Kearney[1], John Coffin[3], Frank Maldarelli[1]

[1]HIV Drug Resistance Program, National Cancer Institute, Frederick, MD 21702, USA, [2]Leidos Biomedical Research, Frederick, MD 21702, USA, [3]Tufts University, MA, USA

BACKGROUND

- cART suppresses but does not eradicate HIV infection
- Infected cells that persist are infected prior to ART initiation
- Measuring the genetic characteristics of HIV populations is essential to understanding the formation of HIV reservoirs that persist during cART
- Analysis of plasma HIV ultrasensitive single genome sequencing (uSGS), yields large HIV sequence datasets with the same, low PCR error and recombination rate as standard SGS
- Higher number of single genome sequences enables detect changes in allele frequencies that are the basis for replicating population size estimates

AIM

- **To determine HIV replicating population size and *in vivo* recombination rate**
- **To modify uSGS to characterize cell associated HIV RNA derived from peripheral blood lymphocytes**

METHODS

1. **HIV-1 cDNA tagging with primer-ID in RT or DNA tagging with primer-ID in linear PCR step**



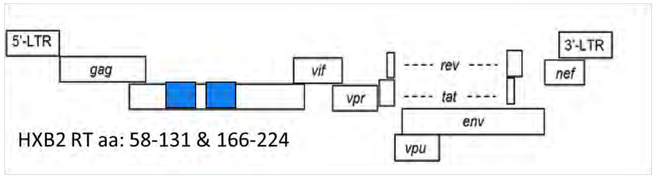
2. **Amplification and ligation of Illumina adapters (the library preparation)**



3. **Sequences after 250 bp pair-end sequencing**



4. **Sequences after quality control, concatenation and trimming**



Concatenated and trimmed final sequence of 400 bp

Modified from Boltz *et al.*, 2016

Replicating population size calculations[2,3]

- Changes in allele frequencies (F)

$$F = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_i - y_i)^2}{\left[(1 - x_i) + (1 - y_i)\right]/2 - (x_i y_i)}$$

- Replicating population size (N_e) - calculated for each site from each subject, S1 and S2 are sample sizes at two time points

$$N_e = \frac{t}{2\left(F - \frac{1}{S1} - \frac{1}{S2}\right)}$$

n – number of alleles per locus

x_i, y_i – allele frequencies at two time points

t – number of generations between sampling (1 day/generation)

Using primer-ID-based next generation sequencing we determined that **HIV replicating population sizes are large** and undergo **frequent recombination**.

RESULTS

- **Study included chronically infected ART naïve male participants (Table 1)**

Table 1. Sociodemographic and clinical data of study participants

Subject ID	Gender	Age (years)	Risk factor	Race/Ethnicity	HIV-1 VL $\log_{10}$ (copies/mL)	CD4+ cell count (cells/μL)	Minimum time since infection (days)	Infection status	ART status at sampling
FMGA-007	Male	26.6	MSM	UNK/Hispanic	4.18	490	135	Chronic	Naïve
FMGA-003	Male	19.5	MSM	UNK/Hispanic	4.99	270	337	Chronic	Naïve
FMGA-010	Male	29.7	MSM	White	4.69	546	71	Chronic	Naïve
FMGA-008	Male	41.9	MSM	American Indian	4.32	506	3907	Chronic	Naïve
FMGA-012	Male	51.1	MSM	White	4.3	672	336	Chronic	Naïve
AVBIO-141	Male	43.5	MSM	White	3.87	424	122	Chronic	Naïve

- **In total 17,172 sequences with median of 1,252 were retrieved**

Table 2. The overview of sequenced samples and retrieved sequences ordered by time between sampling

Subject ID	Minimum time since infection (days)	Sample type	Time between sampling (days)	Input viral RNA t=1 & t=2 (copies)	# of sequences t=1 & t=2	APD (%)
FMGA-003	337	Plasma RNA	36	48,500 & 20,000	2,374 & 1,754	0.3
FMGA-010	71	Plasma RNA	125	62,000 & 127,200	498 & 2,063	0.75
FMGA-012	336	Plasma RNA	168	19,800 & 3,300	509 & 54	1.2
AVBIO-141	131	Plasma RNA	168	179,900 & 50,300	2,387 & 750	0.72
FMGA-008	3,907	Plasma RNA	301	20,600 & 207,600	128 & 3,405	1.21
FMGA-007	135	Plasma RNA	695	15,200 & 50,000	3,165 & 85	1.1
Median	**235.5**	**N/A**	**168**	**49,250**	**1,252**	**0.925**

- **Replicating population sizes exceed 10^7 (Figure 1)**
- **Very small difference in allele frequencies resulted in negative replicating population size values indicating to huge population sizes**

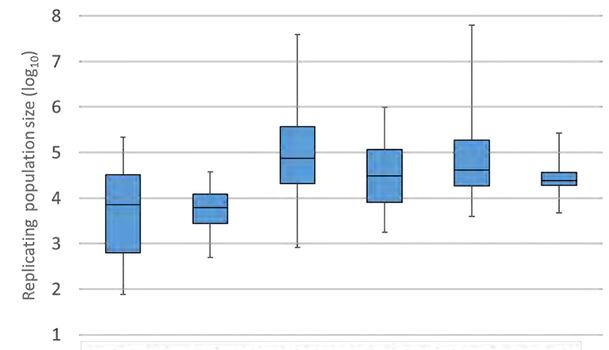


Figure 1. Box-plots of replicating population sizes of each study participants based on 400 bp sequence from reverse transcriptase region. Note. Up to 94% sites resulted in negative replicating population size values

- **Highest calculated population sizes are likely the result of changes in allele frequencies due to genetic drift and lowest the result of selection**
- **Highest values approximate replicating size most closely**
- **Prior to cART, few polymorphisms are linked indicating that HIV recombination is common *in vivo* (Figure 2)**

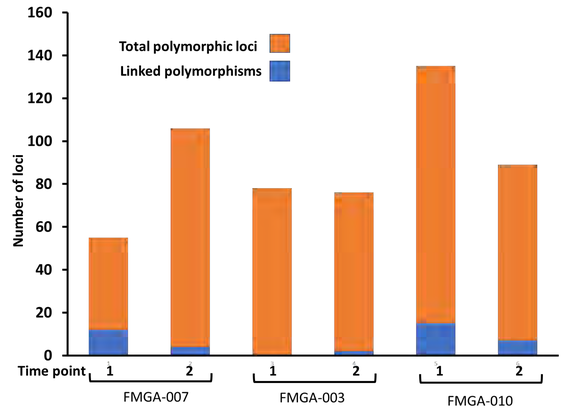


Figure 2. The analysis of linkage of three study participants. Two time points each we determined the total number of polymorphisms ( ), and polymorphisms that were linked (Fisher's exact p<0.001 after Bonferroni correction ( ).

Note. An average of 8% of all loci (40/499) were linked at any single time point.
Only one of these linked loci was found to be linked at both time points for any participants.
Prior to cART, virtually all linked loci become unlinked within short period (<9 months).

- **Using uSGS, 946 sequences were recovered from the peripheral blood lymphocytes of one donor**

CONCLUSIONS

- **Prior to ART, HIV populations are**
  - **large (>10^6-10^7/person)**
  - **compose of variants that undergo frequent recombination.**

Abbreviations
APD – average pairwise distance; ART – antiretroviral therapy; ca – cell associated; N_e – replicating population size; SGS – single genome sequencing; uSGS – ultrasensitive single genome sequencing; VL – viral load

References
1. Boltz et al., Retrovirology 2016
2. Waples *et al.*, Genetics 1989
3. Maldarelli *et al.*, J Virology 2013

Acknowledgments
We thank the study participants.

HIV Dynamics and Replication Program
NCI-Frederick

Contact information:
Kristi Huik
Kristi.huik@nih.gov
HIV Dynamics & Replication Program
NCI-Frederick (NIH)