Plasma virome of HIV infected subjects from the US and Uganda: higher levels of anelloviruses in AIDS patients

Linlin Li[1,2], Xutao Deng[1,2], Jeffrey Martin[2], Steven Deeks[2], and Eric Delwart[1,2]

1. Blood Systems Research Institute, San Francisco, CA; 2.UCSF,CA

University of California San Francisco

Contact: linlinli99013@gmail.com
delwarte@medicine.ucsf.edu

Introduction

HIV infection leads to progressive immune deficiency. Depletion of CD4+ T lymphocytes is one of the distinctive features of HIV infection. As a result of CD4+ depletion and immunodeficiency, AIDS patients lost control of the viral and microbial flora.

Here, we studied the viromes in the plasma of HIV patients with low vs. high CD4 counts from USA and Uganda, by means of next-generation sequencing (NGS). Analyzing the plasma virome in HIV/AIDS patients will help to understand the consequence of immunosuppression on commensal viruses and endogeneous retroviruses in humans.

Methods

Frozen plasma samples from HIV infected patients were obtained from the AIDS Research Program at UCSF. Viral nucleic acids were enriched by filtration and nuclease treatment before extraction and RNA/DNA libraries were constructed using ScriptSeq kit. Each sample was labeled with a different barcode. A library of 24 plasma samples was sequenced using the MiSeq Illumina platform.

Paired-end reads of 250 bp generated were debarcoded using vendor software from Illumina. An in-house analysis pipeline running on a 32-nodes Linux cluster was developed to process the data. Clonal reads were removed and low sequencing quality tails were trimmed using Phred quality score 10 as the threshold. Adaptors were trimmed using the default parameters of VecScreen which is NCBI Blastn with specialized parameters designed for adaptor removal. The cleaned reads were de-novo assembled using SOAPdenovo2. The assembled contigs, along with singlets were aligned to an in-house viral proteome database using BLASTx. The significant hits to virus were then aligned to an in-house non-virus-non-redundant (NVNR) universal proteome database using BLASTx. Hits with more significant adjusted E-value to NVNR than to virus were removed. A web-based graphical user interface (GUI) was developed to present users with the virus hits, along with taxonomy information and processing meta-information.

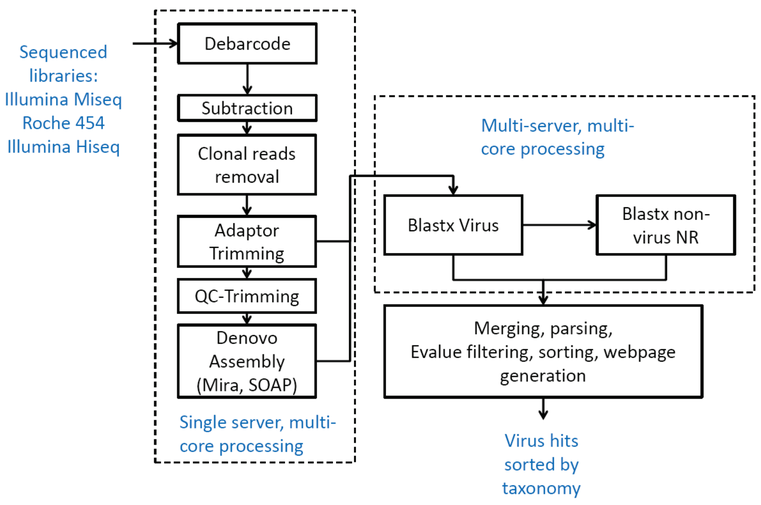

Figure 1. Sequence data processing pipeline

Results

Overview of the NGS run

Viral nucleic acids from plasma samples of 24 HIV infected subjects were sequenced by the MiSeq Illumina platform, generating ~ 22 million of 250 bp pair-end reads. After filtering out the duplicate and low quality sequences, ~6 million of unique sequences were trimmed, sorted, assembled, then put through our analytical pipeline. The resulting contigs and singlets (>50 bp) were then translated, and analyzed for viral protein similarities using BLASTx with an e-value of $<1x10^{-10}$.

Table 1: Summary of the NGS MiSeq Run

Sample Cohort	USA CD4<200/ul	USA CD4>200/ul	Uganda CD4<200/ul	Uganda CD4>200/ul
Patient No.	6	6	6	6
Total reads	5,581,356	2,603,212	5,801,302	5,619,624
Unique reads	946,560	653,389	2,335,111	1,605,910
Percentage	17%	25%	40%	29%

Blood virome in HIV patients

The plasma virome of HIV/AIDS patients contained viral sequences from HIV, GBV-C, HCV, anellovirus, and human endogenous retrovirus (HERV). HIV sequences were identified in 6 low CD4 patients with virus load >110,000 copies/ml. No HBV co-infection were detected. Anellovirus sequences were found in 75% (18/24) of the HIV positive subjects. HERV sequences were detected in 21/24 samples.

Table 2: Sample information and blood virome of the 24 HIV infected patients

Blood samples		Original reads					Standardized reads				
ID	CD4	TTV	HIV	GBV-C	HCV	HERV	TTV	HIV	GBV-C	HCV	HERV
AS10-10508	4	71	55			60	103	80			87
AS04-01159	4	2	14			41	3	24			71
AS06-06468	4	2	13			6	6	40			19
AS11-16494	4	41			2	86	95			5	200
AS03-00205	5	967		566		534	720		421		397
AS03-05969	6	196				36	267				49
AS12-08878	803	2				549	2				615
AS08-00064	817	1		163		10	3		468		29
AS08-03339	837					8					71
AS07-00787	838					191					260
AS12-08598	844					82					164
AS06-10195	856					18					74
MBA1465	10	772	13	1		1777	279	5			643
MBA1120	11	16	4			1538	6	1			570
MBA1516	11	88				15	173				29
MBA1089	12	180				649	135				488
MBA1470	12	2				1238	1				668
MBA1478	13	104	16			193	108	17			201
MBA1549	720	75		11			79		12		
MBA1581	740					2					4
MBA1218	743										
MBA1548	747	17				45	26				68
MBA1256	748	21				1075	5				248
MBA1005	751	1					5				

Samples labeled with "AS" were from USA, while samples with "MBA" were from Uganda. 2. Viral reads were counted with an e-value of $<1x10^{-10}$. Standardization was performed to account from the different number of reads from each sample.

HIV

Table 3: HIV reads subtyping (www.ncbi.nlm.nih.gov/Projects/genotyping/formpage.cgi)

ID	Reads	Subtype
AS10-10508	55	B
AS04-01159	14	B
AS06-06468	13	B
MBA1465	13	A1
MBA1120	4	06_CPX, 01_AE
MBA1478	16	10_CD, D, J

GBV-C

GBV-C has been classified into 7 genotypes and with distinct geographical distributions (G1 Africa; G2 Europe and America; G3 Asia JPN &CHN; G4 SE Asia; G5 only South Africa; G6 Indonesia; G7 CHN.). GBV-C co-infection with HIV-1 has been associated with slower disease progression, and longer survival term (G2 & 5). We found GBV-C in 4/24 samples, including USA (AS03-00205, AS08-00064) G2; Uganda (MBA1549) G1.

Anellovirus

The majority of the anellovirus reads were matched to alpha-anellovirus. The dominant anellovirus species in USA samples were different from those in Uganda. The anellovirus reads were significantly high in US low than in high CD4 group (Exact Wilcoxon test, $p<0.05$).

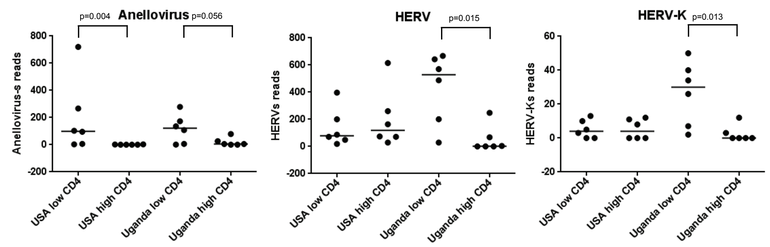

Figure 2. Comparison of standardized reads for anellovirus, HERV, and HERV-K in four groups (The Exact Wilcoxon test was used to compare median reads of anellovirus, HERV and HERV-K in HIV/AIDS patients.).

HERV

HERV comprise 8% of the human genome. Integrated HERV genomes commonly contain mutations, deletions, or are even reduced to single long terminal repeat (LTR) elements by homologous recombination. Under normal circumstances, HERVs are functionally defective or controlled by host factors. HERV-K represents the youngest and most active family.

The majority of HERV reads were matched to Class I (Gamma) group. HERV and HERV-K sequence reads were significantly high in Uganda low than in high CD4 group (Exact Wilcoxon test, $p<0.05$).

Conclusion

In this study, we examined the plasma virome in low and high CD4 HIV infected subjects from the US and Uganda. Immunodeficiency was associated with higher plasma levels of anelloviruses. HERV levels were also higher in low CD4 subjects in Uganda but not the US.