Using RNASeq to Characterize Viral Genetic Diversity in a Cohort of HIV-2 Monoinfected Individuals

Katherine James[1], Thushan De Silva[1,2], Shokouh Makvandi-Nejad[1], Tim Vincent[3], Hilton Whittle[4], Stephen Taylor[1], Gilean McVean[1], Sarah Rowland-Jones[1]

[1]University of Oxford, Oxford, UK, [2]University of Sheffield, Sheffield, UK, [3]MRC Laborartories, Fajara, The Gambia, [4]London School of Hygiene and Tropical Medicine, London, UK

Background

Human Immunodeficiency Virus 2 (HIV-2) is a non-pandemic, contemporary human Lentivirus with the majority of infections localized to West Africa[1]. Unlike HIV-1, disease progression following HIV-2 infection appear to be bimodal. Approximately 20% of patients progress to AIDS in a manner which is clinically indistinguishable from HIV-1 infection. Approximately 38% of patients are long term non-progressors (LTNP) and maintain undetectable viral loads and high CD4+ T cell counts in the absence of ART, sometimes for decades. For LTNP, survival rates are unaffected by HIV-2 infection.

Previous work on HIV-2 plasma genetic diversity has been limited and focused on single genes, mostly generated from patient derived clones. Developing a bias-free method of whole genome sequencing may help to elucidate whether there is a correlation between disease progression and HIV-2 viral diversity and shed light on the nature of viral replication in HIV-2 LTNP.



Figure 1: Caio, Guinea-Bissau

Why RNASeq?

Plasma samples have been collected from HIV-2 infected individuals from the Caio Community Cohort, Guinea Bissau

PCR is inherently biased by both stochastic factors and the necessity of designing sequence specific primers. Mutations in primer binding sites lead to a difference in PCR efficiency between variant viral genomes.

Ideal sequencing method involves no, or sequence independent, amplification.

RNASeq involves reverse transcription of complete RNA, meaning that amplification biases are avoided.

RNASeq does not target a particular region of the genome and so whole genome sequences are retrieved.

Methodology

1. RNA extracted from plasma/culture supernatant using modified QIAGen UltraSens Protocol and quantified using a QuBit fluorescence assay.

2. Libraries prepared with a RNASeq protocol (NEB). cDNA synthesis was primed with random hexamers.

3. Libraries were sequenced using the Illumina HiSeq platform. Samples were multiplexd 6/lane and 100bp paired-end reads were generated.

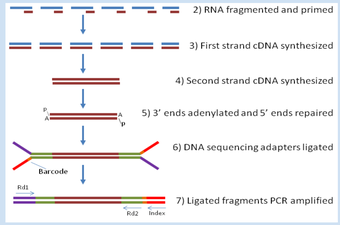

Figure 2: Schematic showing RNASeq library prep method

4. Read quality was assessed with FASTQC and low quality (Q score < 20) reads were removed. Reads were mapped to a multiple sequence alignment of all HIV-2 A sequences (downloaded from Los Alamos HIV Database) using the assembly package VICUNA (Broad Institute, ref [2])

5. Assembled whole genome sequences were visualized using IGV [3].

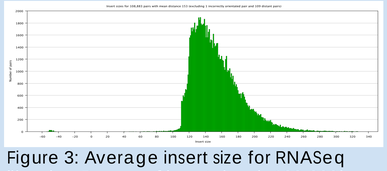

Figure 3: Average insert size for RNASeq libraries. Range of insert sizes is 100-200bp

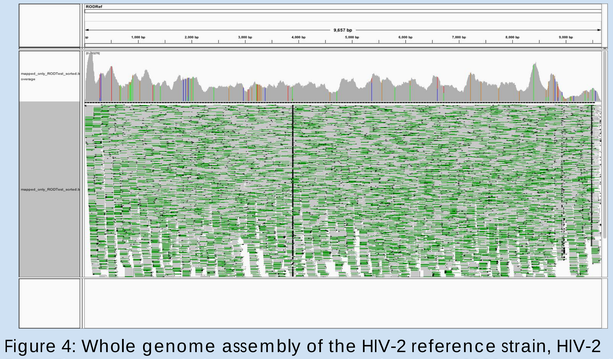

Figure 4: Whole genome assembly of the HIV-2 reference strain, HIV-2 ROD. The upper track shows coverage (scale up to 10000x). Polymorphic positions are coloured and homologous positions are shown in light grey. The lower track shows mapped read pairs.

Results

Sample	Copies of Virus	Alignment Reference	No Reads Aligning	WG Coverage	Depth	Genes Intact
TD024	5280	Pan HIV-2a	4998	93	52x	9
TD031	53591	Pan HIV-2a	9065	90	97x	9
TD062	69759	Pan HIV-2a	13304	87	147x	7
ROD	10^8	ROD	223983	91	2376x	9
ROD	10^8	Pan HIV-2a	222222	95	2258x	8
CBL-20	10^8	Pan HIV-2a	930072	87	10319x	9

Figure 5: Table showing mapping reference, number of aligned reads, coverage and depth for 3 patient samples (TD) and two HIV-2 reference strains, HIV-2 ROD and CBL-20

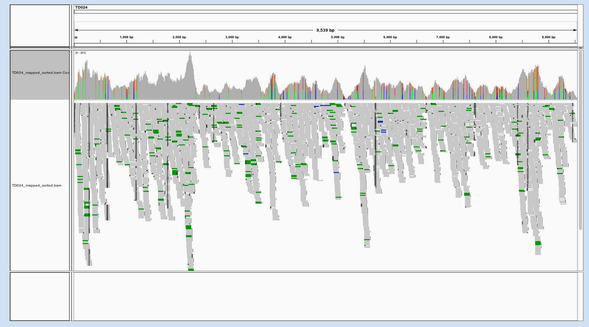

Figure 6: Whole genome assembly for patient TD024. Coverage track shows a large (1kb) region of conservation in Gag-Pol

Conclusions

We have demonstrated that RNASeq provides a feasible and valuable tool for generating HIV-2 whole genome sequences directly from patient plasma without the need for sequence-specific amplification or clonal expansion, reducing biases introduced during sample-preparation.

In addition to sequencing primary patient samples, RNASeq can also be used to generate high depth, whole genome sequences from cultured virus.

We anticipate that this novel application for RNASeq in HIV-2 research will help elucidate the complex and dynamic evolution of HIV-2 in both progressors and LTNP.

References

[1] Sharp *et al* (2011) –Cold Spring Harb. Perspect. Med, Sep 1(1)
[2] Yang *et al* (2012) – BMC Genomics 13:475
[3] Robinson *et al* (2011) – Nature Biotechnology 29:24-26

MRC The Gambia | wellcome trust | UNIVERSITY OF OXFORD