Exploring the HIV reservoir with full-length provirus sequencing

Dr John-Sebastian Eden
Centre for Virus Research
& The University of Sydney

The Westmead Institute FOR MEDICAL RESEARCH

Outline

1. Background to HIV provirus sequencing
2. Origin and use of the “FLIPS” assay
3. Sequence assembly workflow
4. Provirus annotations: Intact vs defective

The Westmead Institute FOR MEDICAL RESEARCH

Bonnie Hiener
Beth Horsburgh
Sarah Palmer

Cell Reports
Resource

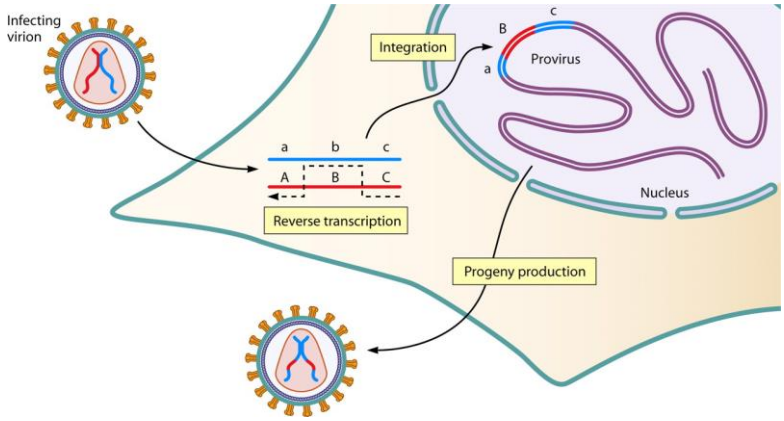
Identification of Genetically Intact HIV-1 Proviruses in Specific CD4+ T Cells from Effectively Treated Participants

Bonnie Hiener,[1,9,*] Bethany A. Horsburgh,[1] John-Sebastian Eden,[1,2] Kirston Barton,[1] Timothy E. Schlub,[3] Eunok Lee,[1] Susanne von Stockenstrom,[4] Lina Odevall,[4] Jeffrey M. Milush,[5] Teri Liegler,[5] Elizabeth Sinclair,[5] Rebecca Hoh,[5] Eli A. Boritz,[6] Daniel Douek,[7] Rémi Fromentin,[8] Nicolas Chomont,[8] Steven G. Deeks,[5] Frederick M. Hecht,[5] and Sarah Palmer[1]

[1]Centre for Virus Research, The Westmead Institute for Medical Research, The University of Sydney, Sydney, NSW 2145, Australia
[2]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
[3]Sydney School of Public Health, Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia
[4]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Karolinska University Hospital, Stockholm 171 77, Sweden
[5]Department of Medicine, University of California, San Francisco, San Francisco, CA 94110, USA
[6]Virus Persistence and Dynamics Section, Vaccine Research Center, National Institute of Allergy and Infectious Disease, NIH, Bethesda, MD 20814, USA
[7]Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Disease, NIH, Bethesda, MD 20814, USA
[8]Centre de Recherche du CHUM and Department of Microbiology, Infectiology and Immunology, Université de Montréal, Montréal, QC H2X 0A9, Canada
[9]Lead Contact
*Correspondence: bonnie.hiener@sydney.edu.au
https://doi.org/10.1016/j.celrep.2017.09.081

The Westmead Institute FOR MEDICAL RESEARCH

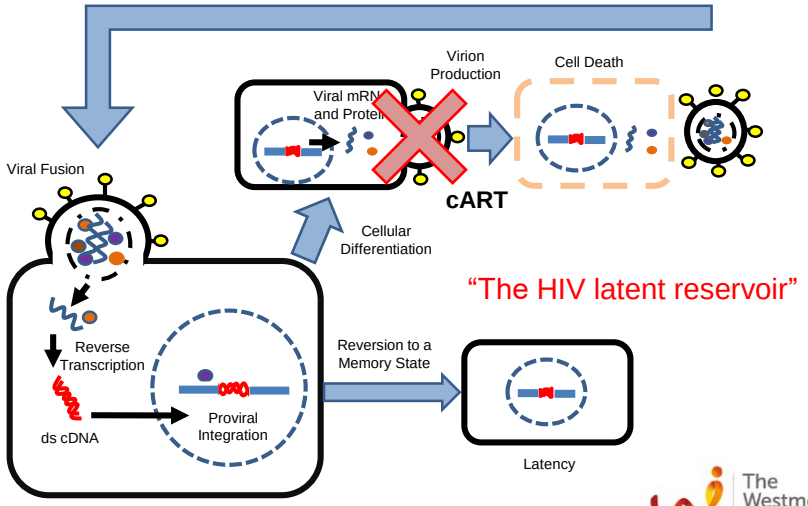
What is a Provirus?



Adewunmi Onafuwa-Nuga & Alice Telesnitsky
Microbiol. Mol. Biol. Rev. (2009) 73:451-480

The Westmead Institute FOR MEDICAL RESEARCH

HIV Lifecycle



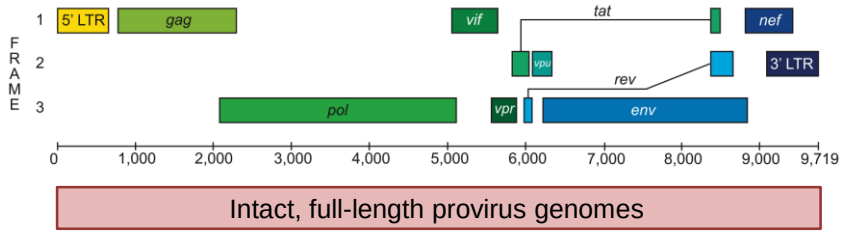
HIV Reservoir

For a cure, we need to know where the virus hides

- For example, we know that the virus remains dormant in resting, memory CD4+ T-cells.
- But...
  - which particular cell subsets?
  - where are the viable "replication-competent" viruses?
  - And, in what proportion?

Provirus sequencing

How do we known if a virus is replication-competent?



Intact, full-length provirus genomes

High-throughput sequencing looking for intact proviruses

No mutations, no deletions, no frameshift, no stop codons, no inversions

Vast majority will be defective!

The Westmead Institute FOR MEDICAL RESEARCH

The FLIPS assay

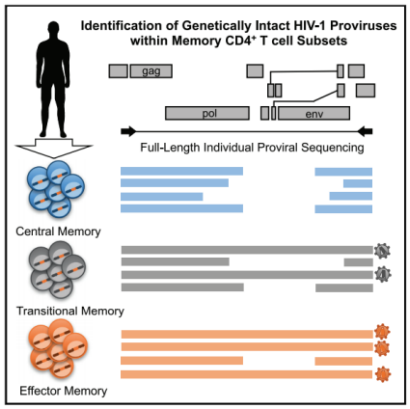
Full

Length

Individual

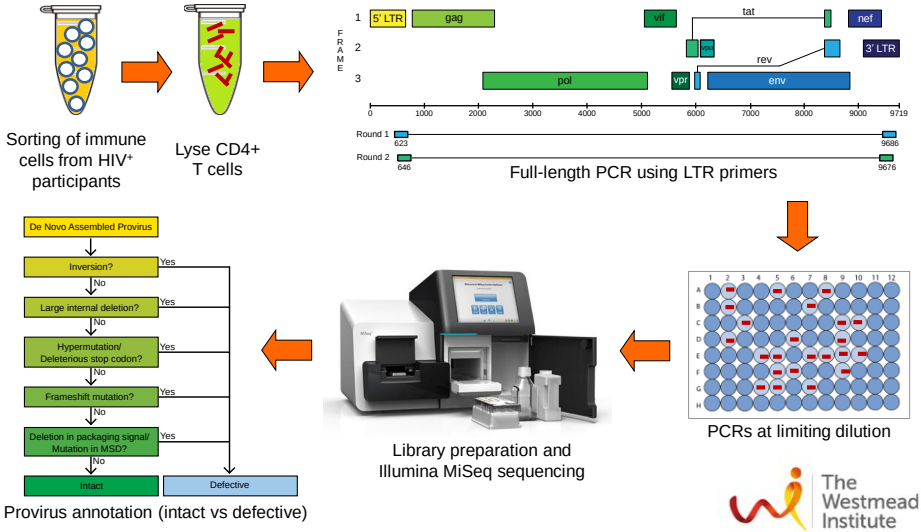
Provirus

Sequencing

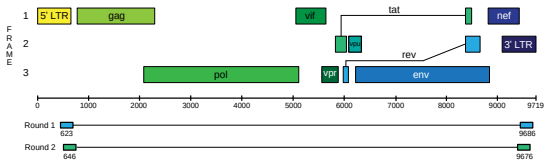


The Westmead Institute FOR MEDICAL RESEARCH

The FLIPS assay

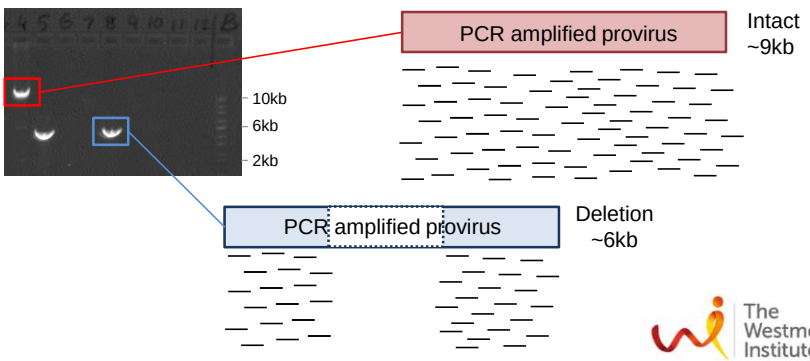


PCR using LTR primers



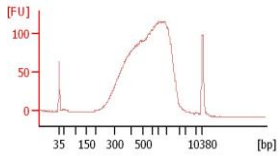
"Limiting dilution"

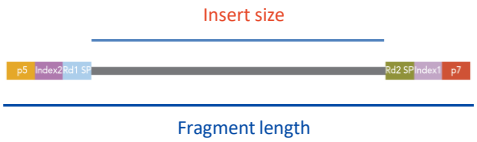
One PCR amplicon = One provirus



Illumina MiSeq sequencing

- DNA 'Tagmentation' (Tagging + Fragmentation)

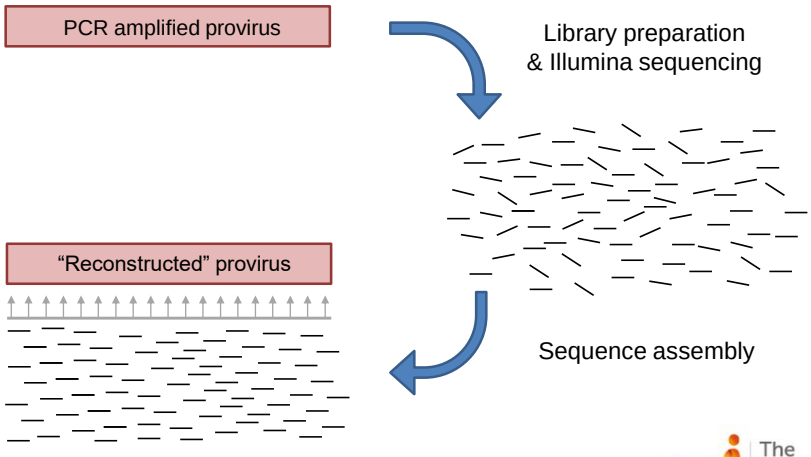




- Sequence "barcodes" allows multiple samples per run
- One MiSeq run = 20 million sequence reads (2x150nt)
- **60 to 80 "proviruses" per run**

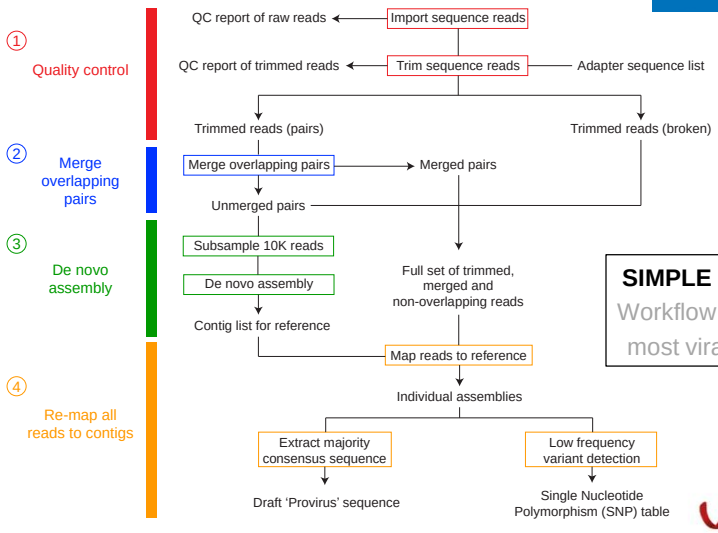
The Westmead Institute FOR MEDICAL RESEARCH

Need for sequence assembly



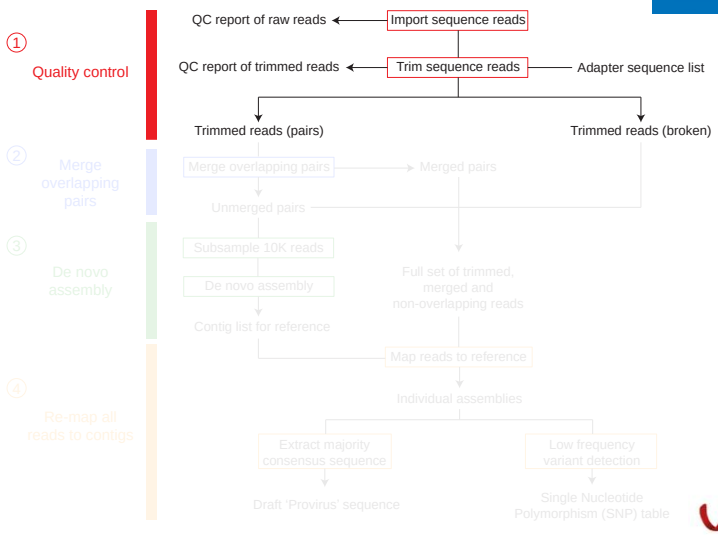
The Westmead Institute FOR MEDICAL RESEARCH

Assembly workflow



SIMPLE & RELIABLE
Workflow works well for most viral assemblies

Assembly workflow



1. Quality control

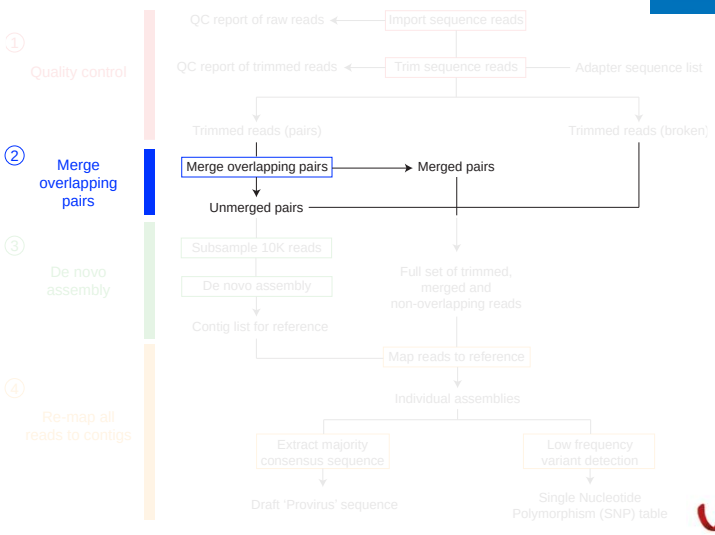
- Assess quality of the data
  - Sequencing provides per base “quality” scores
  - Score reflects a probability of error
- Trim sequences
  - Based on quality threshold (1%)
  - Of any sequence artifacts like primer or adapters
- Filter reads by length (>50bp)

Open source = FastQC, Trimmomatic & Cutadapt

The Westmead Institute FOR MEDICAL RESEARCH

Assembly workflow

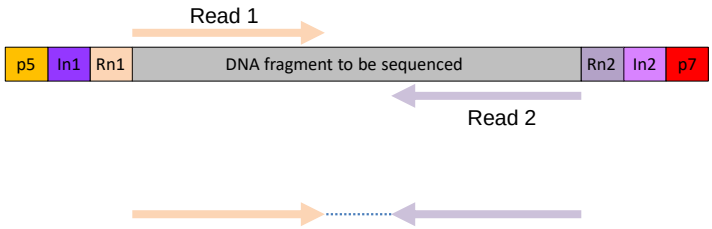
CLC Genomics Workbench



The Westmead Institute FOR MEDICAL RESEARCH

2. Merge overlapping pairs

- Paired reads are from one fragment, sequenced from each end



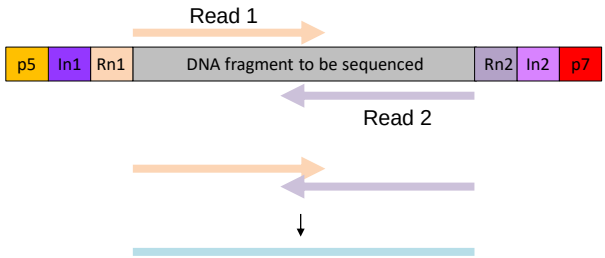
Paired reads

Open source = Flash

The Westmead Institute FOR MEDICAL RESEARCH

2. Merge overlapping pairs*

- If read length (2X150nt) more than fragment length, then reads will overlap and can be collapsed



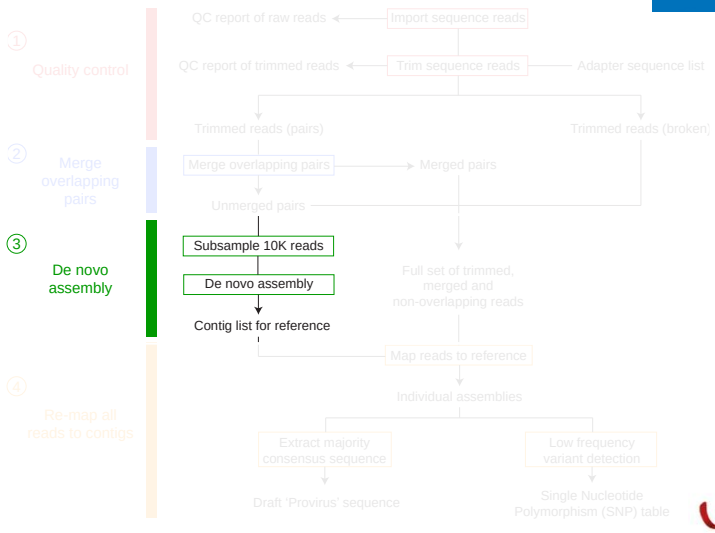
Merged reads

Open source = Flash

*Somewhat optional

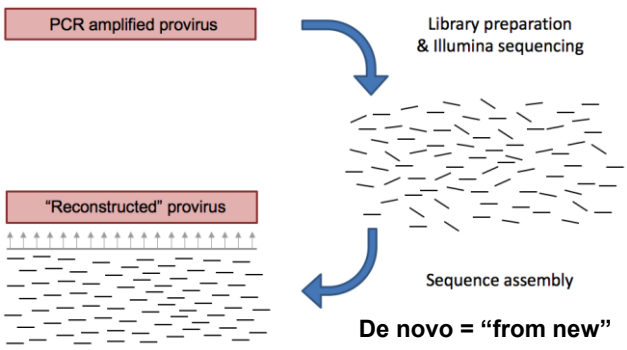
The Westmead Institute FOR MEDICAL RESEARCH

Assembly workflow



3. De novo assembly

Reconstructing the sequence from short reads



De novo = "from new"

No guides… No prior info…

Just the reads!

3. De novo assembly

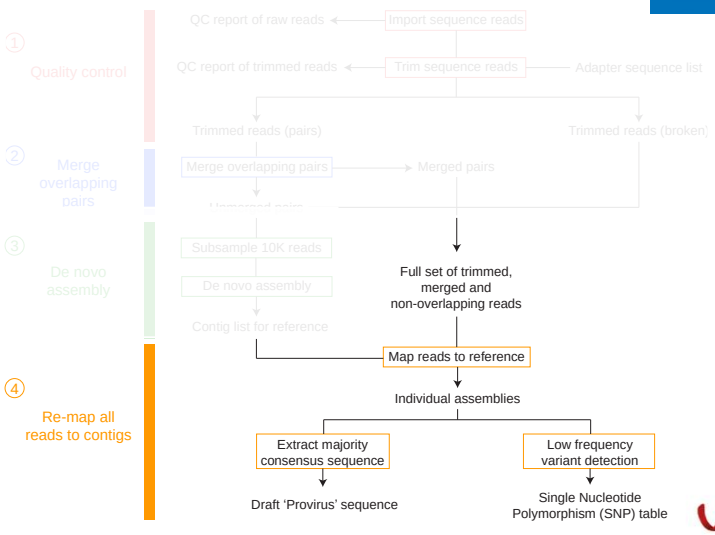
- Cuts reads into smaller bits (“words”), then looks for overlap and then extends to make longer “contigs”
- Each set should contain amplified target (i.e. provirus)
- Downsample reads to 10K to **run on typical desktop**
  - Full set of 300,000 reads = slow, needs larger memory
- Provirus should be an assembled contig
  - Most abundant & matched to PCR product size

Open source = Trinity, Spades, Velvet

The Westmead Institute FOR MEDICAL RESEARCH

Assembly workflow

CLC Genomics Workbench



The Westmead Institute FOR MEDICAL RESEARCH

4. Final provirus sequences

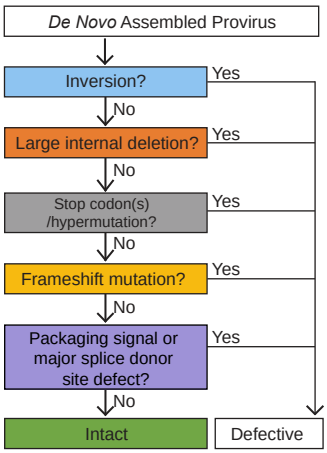
- Identify HIV provirus from list of *de novo* contigs
- Map the full set of reads and take a final “consensus”
- Look for heterogeneity in data
  - One amplicon = One provirus = No SNPs or variation
- The final ‘provirus’ must then be annotated/analysed

Intact or Defective?

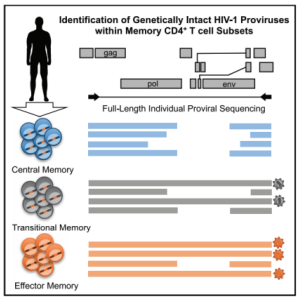
Open source = Bowtie, Samtools, Freebayes

The Westmead Institute FOR MEDICAL RESEARCH

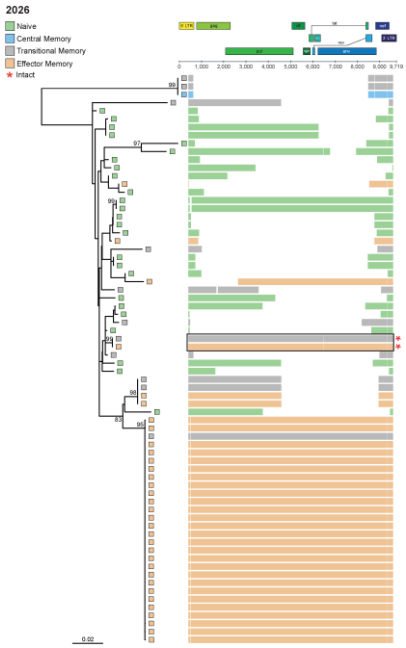
Provirus annotation



Based on notion that the virus needs complete, intact gene sets for functional replication



The Westmead Institute FOR MEDICAL RESEARCH

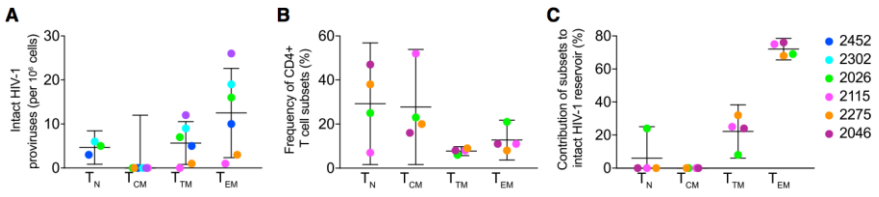


Provirus Phylogeny

- Compare sequences for each patient
- Map proviruses to cell subsets
- Identical sequences suggest clonal expansions
- Where intact genomes?
- Must check non-coding regions!

The Westmead Institute FOR MEDICAL RESEARCH

Where are the intact proviruses?



- Intact proviruses were unequally distributed between T-cell subsets
- Effector memory cells containing the largest proportion of genetically intact HIV-1 proviruses

The Westmead Institute FOR MEDICAL RESEARCH

Pros and Cons

- Previous methods targeting *env* or *gag-pol* overestimated frequency of intact proviruses
- High-throughput sequencing screen more provirus
- The context of provirus integration cannot be determined using FLIPS
- Single copy assays prone to contamination
- PacBio single molecule seq (i.e. 1 provirus = 1 read)

The Westmead Institute FOR MEDICAL RESEARCH

Summary

- FLIPS utilizes NGS to sequence and characterize HIV-1 proviruses
- FLIPS can identify genetically intact and likely replication competent HIV-1 proviruses
- Identical HIV-1 proviruses suggest maintenance of reservoir by cellular proliferation & importance of reservoir in *Effector Memory T Cells*

The Westmead Institute FOR MEDICAL RESEARCH

Acknowledgments

Palmer lab WIMR: Centre for Virus Research
A/Prof S. Palmer
B. Hiener
B. Horsburgh
E. Lee
V. Morcilla
K. Fisher
M-A. De Scheerder
C. Wang
Z. Boyer
A. Winckelmann

WIMR: Centre for Virus Research
Prof T. Cunningham
J. Lai

University of Montreal
N.Chomont
R.Fromentin

Immunology Laboratory, Vaccine Research Centre, NIH
E. Boritz
D. Douek

Ramaciotti Centre UNSW

University of Sydney
T. Schlub

Department of Medicine UCSF
F. M. Hecht
S. G. Deeks
J. Milush
T. Liegler
M. Somsouk
P. Hunt
E. Sinclair
P. Lewis
H. Hatano
L. Epling
M. Kilian
T. Ho
A. Tan
J. Custer
L. Loeb
R. Hoh
L. Poole

We acknowledge with gratitude the participants of this study