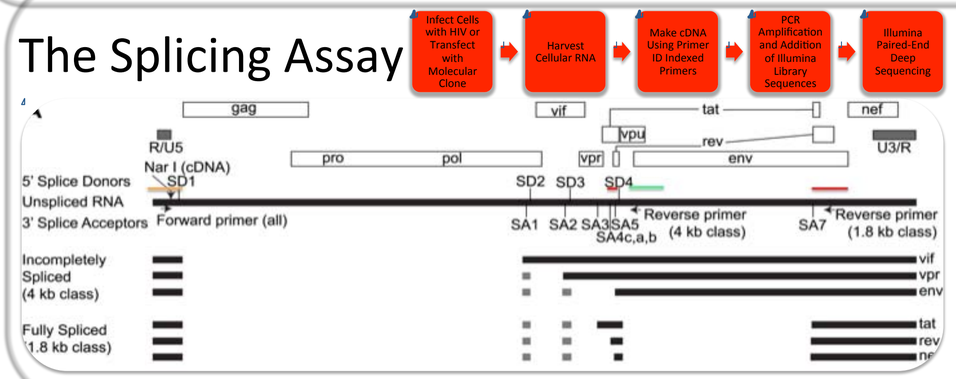
Quantification of HIV-1 Splicing Using a PrimerID-Tagged Deep Sequencing Assay

A Emery[1] and R Swanstrom[2,3,4]

Abstract

HIV-1 RNA undergoes complex splicing to produce more than 40 different transcripts. Previous research on HIV-1 splicing used artificial constructs and single exon systems, and quantification was done by size and relative abundance of PCR products. It is now possible to quantify HIV-1 RNA splicing patterns in the context of viral infection using Illumina paired end deep sequencing and PrimerID. This quantitative assay can also be used to explore the role of splice regulatory sequences, using splicing patterns as a phenotype. This splicing assay was used to observe APOBEC3G induced mutations to the region between splice acceptor 5 (A5) and splice donor 4 (D4). This region contains a regulatory element, ESE(GAR), thought to control splicing at both A5 and D4. Data from the splicing assay can be used to correlate a panel of mutations in the ESE(GAR) control region to both upstream and downstream splicing events.

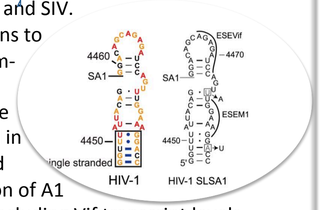
The Splicing Assay



The single HIV-1 transcript made is a full length copy of the RNA form of the genome that can serve as either genomic RNA or as a template for the Gag/Pro/Pol proteins. About half of transcripts are spliced. All spliced RNAs use D1 and one of the 8 downstream acceptor sites. About half of spliced transcripts have the D4-A7 exon removed (1.8kb size class). The 4kb size class retains the D4-A7 exon. The reverse primer used to exclusively capture the 4 kb class of viral RNAs is at the 5' end of the D4-A7 exon (shown in green). The reverse primer for the 1.8 kb class is shown in red. This primer spans the D4-A7 splice junction, which is specific to the 1.8 kb RNAs. The same forward primer, shown in orange, is used in the subsequent PCR amplification of both size classes.

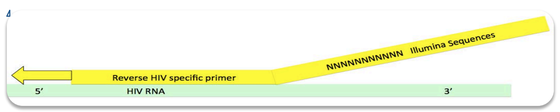
Background- Structure & Splicing

The SLSA1 stem loop is highly conserved[2] in HIV-1 and SIV. Mutations to this stem-loop structure resulted in a 10 fold reduction of A1 usage, including Vif transcript levels.



What is PrimerID?

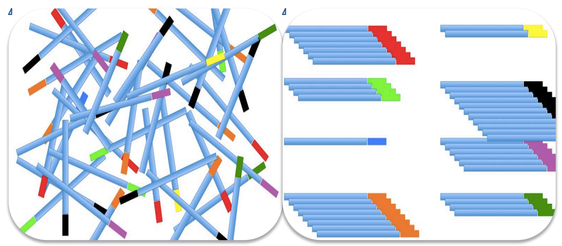
PrimerID[1] is a random sequence tag associated with each oligo primer, which is incorporated into a cDNA and all PCR products subsequently made from that tagged cDNA.



The random PrimerID sequence is made by including all four bases in 14 successive synthesis steps of the reverse primer. This creates an embedded random NNNNNNNNNNNNNN sequence between the specific HIV-1 and the adaptive Illumina sequences. This is the PrimerID. The reverse primer is therefore a mix of about 64 million different primers, each with the same HIV-1 and Illumina specific sequences, but with different PrimerID sequences.



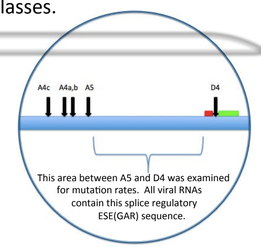
For example, starting with 8 different viral mRNAs, each of them gets made into a cDNA with a unique primerID tag. If the number of random primerIDs is in sufficient excess, each cDNA will be tagged by a different primerID.



The tagged cDNAs get amplified in subsequent PCR reactions and all the descendants of a parent cDNA inherit this unique PrimerID tag.
After sequencing, the PCR products can be sorted by their primerIDs. Some get amplified many times, some not at all. Since each unique primerID represents one specific viral transcript, each occurrence of a primerID is counted exactly once, and skewing in the PCR steps is filtered out. Additionally, a consensus sequence can be made from multiple reads of the same specific sequence. This makes it possible to filter out errors introduced by PCR.

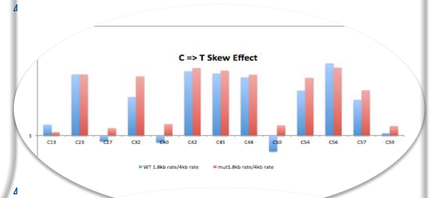
Mutation Screen Results

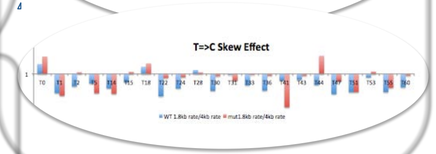
As this panel of selected mutations in the ESE(GAR) region illustrates, C=>T and T=>C transitions can affect upstream (A5) splicing and/or downstream (D4-A7) splicing.



consensus sequence	1.8:4kb proportion	fraction A5 in 1.8kb	fraction A5 in 4kb	A5 1.8kb:4kb ratio
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	0.992720104	0.638856523	0.863414678	0.739918534
CTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	1.281435144	0.691358025	0.855855856	0.807797271
TCTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	0.708577847	0.565217391	0.850877193	0.664276109
TTCGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	0.796722193	0.632653061	0.888888889	0.711734694
TTTGATAGAGAAGCTCGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	0.812713076	0.571428571	0.859504132	0.664835165
TTTGATAGAGAAGCTTGATGAGTCTGACTGCTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	0.708763436	0.671641791	0.855421687	0.785158713
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCTGCT	2.132335199	0.617647059	0.835714286	0.739064857
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTTCGCT	4.745046288	0.25984252	0.531914894	0.488503937
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTTGCTGTCTCCGCT	2.747073652	0.493670886	0.821782178	0.600732042
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTCTTTGTCGCTGTCTCCGCT	3.24701875	0.493197279	0.748427673	0.658977877
TTTGATAGAGAAGCTTGATGAGTCTGACTGTTCTGATGAGCTTTCGTCGCTGTCTCCGCT	3.433786694	0.501628664	0.834394904	0.601188552
TTTGATAGAGAAGCTTGATGAGTTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	3.206683113	0.59047619	0.765217391	0.771645022
TTTGATAGAGAAGTTTGATGAGTCTGACTGTTCTGATGAGCTCTTCGTCGCTGTCTCCGCT	1.313824664	0.708333333	0.923076923	0.767361111

Reduced vif transcripts led to increased C=>T mutations which were unevenly skewed towards the 1.8kb splice class. T=>C mutations skewed towards the 4kb splice class





References

1. Jabara CB, Jones CD, Roach J, Anderson JA, and Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci U S A. 2011 Dec 13;108(50):20166-71.
2. Pollom E, Dang KK, Potter EL, Gorelick RJ, Burch CL, Weeks KM, and Swanstrom R. Comparison of SIV and HIV-1 Genomic RNA Structures Reveals Impact of Sequence Evolution on Conserved and Non-Conserved Structural Motifs. PLoS Pathog, Apr 2013;9:4.

1. Curriculum in Genetics and Molecular Biology, University of North Carolina Chapel Hill
2. UNC Center for AIDS Research, University of North Carolina Chapel Hill
3. Lineberger Comprehensive Cancer Center, University of North Carolina Chapel Hill
4. Department of Biochemistry and Biophysics, University of North Carolina Chapel Hill