Supplementary Data

```
____  _____    _    ____
|  _\| ____|  / \  | _  \
| |_)| _|    / _ \ | |_) |
| __/| |___ / ___ \| _  <
|_|  |_____/_/   \_\_| \_\
PEAR v0.9.5 [September 12, 2014]
Citation - PEAR: a fast and accurate Illumina Paired-End reAd mergeR
Zhang et al (2014) Bioinformatics 30(5): 614-620 | doi:10.1093/bioinformatics/btt593
Forward reads file.................: (Forward FastQ File)
Reverse reads file.................: (Reverse FastQ File)
PHRED..............................: 33
Using empirical frequencies........: YES
Statistical method.................: OES
Maximum assembly length............: 999999
Minimum assembly length............: 50
p-value............................: 0.010000
Quality score threshold (trimming).: 0
Minimum read size after trimming...: 1
Maximal ratio of uncalled bases....: 1.000000
Minimum overlap....................: 10
Scoring method.....................: Scaled score
Threads............................: 1
Allocating memory..................: 200,000,000 bytes
Computing empirical frequencies....: DONE
................................................................................
Lentiviral Merged Reads
Assembled reads ...................: 1,137,595 / 1,198,876 (94.888%)
Discarded reads ...................: 0 / 1,198,876 (0.000%)
Not assembled reads ...............: 61,281 / 1,198,876 (5.112%)

Gammaretroviral Merged Reads
Assembled reads ...................: 3,267,104 / 3,558,352 (91.815%)
Discarded reads ...................: 0 / 3,558,352 (0.000%)
Not assembled reads ...............: 291,248 / 3,558,352 (8.185%)

Foamy virus Vector merged Reads
Assembled reads ...................: 4,006,264 / 4,263,921 (93.957%)
Discarded reads ...................: 0 / 4,263,921 (0.000%)
Not assembled reads ...............: 257,657 / 4,263,921 (6.043%)
```

Supplementary Figure S1. Paired-End reAd mergeR (PEAR) settings and representative output. PEAR is an efficient and highly accurate Illumina pair-end read merger. PEAR evaluates all possible paired-end read overlaps and implements a statistically observed expected alignment score (OES) test for minimizing false-positive results. Utilizing PEAR with the default parameters, we observed that >90% of paired sequences were able to be merged for retroviral integration site analysis, and no sequences were discarded because of poor sequence quality.