PE and SE for RNA Seq Xenograft

Introduction

For typical samples, single end (SE) 1x50bp sequencing is typically sufficient. However, one of our users is interested in looking at RNA expression levels for a human mouse Xenograft. This will result in a mixture of Human and Mouse RNA in the sample. We were curious to see what effect the sequencing type can have on the ability to generate useful data. Our hypothesis is that 125 bp paired end (PE) sequencing could reduce cross species alignment rates, resulting in more accurate results.

In this tech note, we discuss the results of an *in-silico* simulation of the two sequencing types in an effort to predict how important the longer reads will be for the Xenograft study.

Methods

To test our hypothesis we took HiSeq 2500 High Output V4, 2x125bp data and trimmed the reads with Cutadapt to simulate 1x 50 bp data. Reads from Human and Mouse samples sequenced on the same run were used for comparison. These reads were then aligned against both Mouse and Human reference genomes (GRCm38 and respectively GRCh37). This was done using the new NGI-RNA-seq BP 2.0 pipeline which uses STAR in two pass mode as the aligner.

Results

Read alignment scores can be seen in Figure 1. We see that considerably more reads in the single-end samples align against the wrong genome (~24%) compared to paired end reads (~8%).

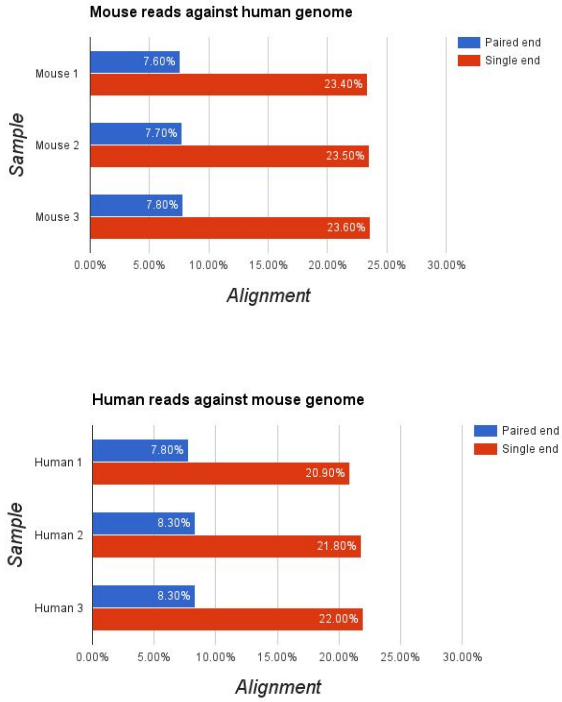

Figure 1: *Mouse reads aligned against the Human reference genome GRCh37 (top) and Human reads aligned against the Mouse reference genome GRCm38 (bottom). Paired end reads in blue and single end reads in red.*

Conclusions

Paired end sequencing greatly outperforms single end sequencing when it comes to minimizing the amount of cross species alignment. We recommend that it be used for all RNA experiments where RNA from multiple species are expected in a sample.