Improving transcriptome assembly through error correction of high-throughput sequence reads

Matthew D MacManes[1*] and Michael B. Eisen[1,2]

1 UC Berkeley. California Institute of Quantitative Biology, Berkeley, CA, USA
2 Howard Hughes Medical Institute

∗ Corresponding author: macmanes@gmail.com, Twitter: @PeroMHC

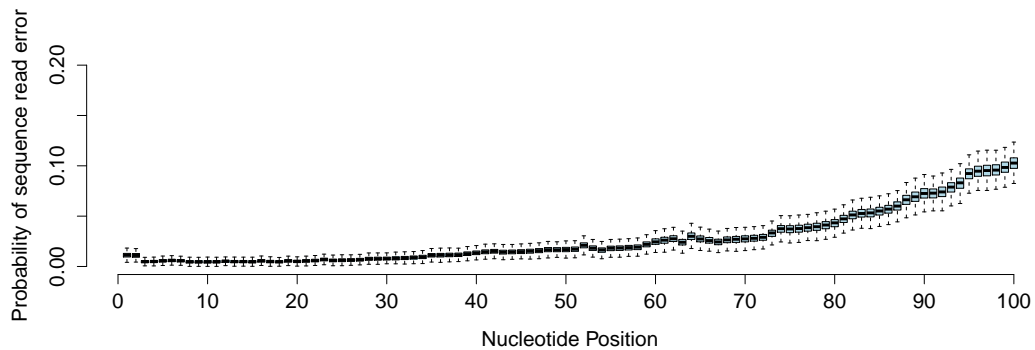## Supplementary Material

Supplementary Figure 1

Figure1. The probability of error increases from the 5' to 3' end of the simulated sequencing reads. This error pattern is typical of Illumina sequencing.
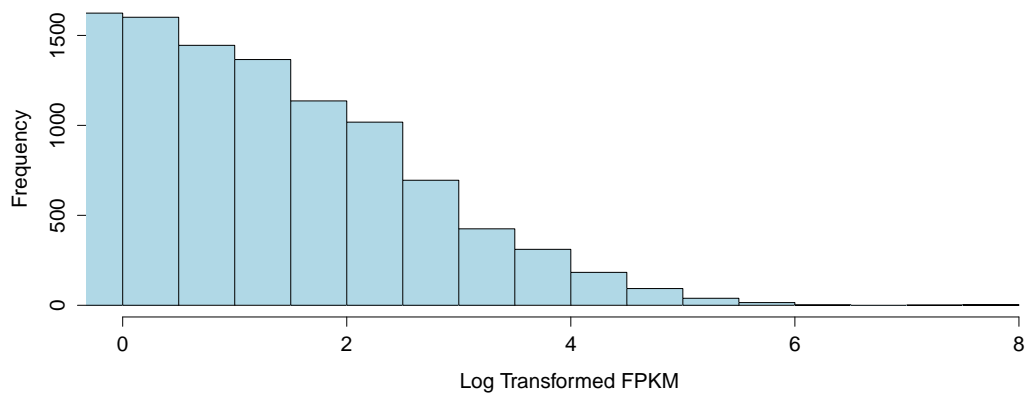
Supplementary Figure 2

Figure 2. The log transformed distribution of gene expression (FPKM) in the simulated dataset. This pattern is typical of many well characterized Eukaryotic tissue types, where the vast majority of genes are expressed moderately, with much fewer being expressed at a very high level.