Pool design for phase IV ERCC RNAs available for RNA-Seq experiments.

Stocks of individual RNAs were purified and quantified with UV absorption. The stocks were adjusted in concentration such that they each had the same molarity. These adjusted stocks were mixed into five subpools that each contained about 20 transcripts, covering a 10^6 dynamic range of abundance. This dynamic range was achieved by preparation of sub-subpools covering a range of 64-fold or 128-fold abundance, reasonable sets of dilutions using standard pipetting technique. These sub-subpools were mixed to create the subpools, which were then mixed in proportions to create pools 12-15 (see table S1), which were arranged such that they had a subpool in constant relative abundance, and four subpools that changed in abundance relative to each other in a Latin square design.

Subpool/Pool	12	13	14	15
A	10%	10%	10%	10%
B	10%	15%	25%	40%
C	15%	25%	40%	10%
D	25%	40%	10%	15%
E	40%	10%	15%	25%

Estimates of required read depth:

$$n = \frac{k}{p}$$

$$k = \frac{C * t}{r}$$

$$p = \frac{t * \frac{330}{6.02e23} * m}{y}$$

n is total reads number

k is minimum number of reads mapped to transcript A to achieve annotation required coverage

p is probability of a random reads sampling from transcript A

C is the minimum coverage required to cover 99% of transcript A, in our analysis, on average 8X coverage is the minimum sequencing depth to cover 99% of a transcript

t is the length of transcript A
330 is average molecule weight for one nucleotide
6.02e23 is the approximate of Avogadro constant
r is read length
m is abundance of transcript A in a cell
y is yield of total mRNA in a single cell

Based on the above equation, we can get a formula to calculate total reads number for annotating a transcript at given abundance as

$$n = \frac{\frac{C * t}{r}}{\frac{t * 330}{6.02e23} * m}{y}$$

Simplified:

$$n = 1.46e22 * \frac{y}{r * m}$$

In a S2 cell with 0.17pg mRNA per cell, the minimal required read number (36bp single end read) to cover 99% of a transcript with an average abundance of 1 copy per cell is 68 million.