Scotty - Power Analysis for RNA Seq Experiments

[Marth Lab](#)

[Help](#)

Scotty is a tool to assist in the designing of RNA Seq experiments that have adequate power to detect differential expression at the level required to achieve experimental aims.

At the start of every experiment, someone must ask the question, "How many reads do we need to sequence?" The answer to this question depends on how many of the truely differentially expressed genes need to be detected. A greater number of genes will be found with an increase in the number of replicates and an increase in how deeply each existing replicate is sequenced. These parameters are limited by the budget for performing the experiment.

The power that is available using a given number of reads will differ between experiments. Ideally, pilot runs of your experiment (small runs of at least two replicates from one of your conditions) should be used to assess the amount of biological variance that is in the system you are studying, and the amount of sequencing depth that is required to adequately measure the genes. Alternatively, Scotty can be run on data from publicly-available datasets that are very close to your expected experiment (species, library preparation protocol, sequencing technology, and read length).

The Matlab code that runs background calculations is available on [github](#). Please contact us if your require assistance.

Inputs

Pilot Data: Upload your own pilot data or used a stored dataset as a model for your experiment. [(?)](#)

CAUTION

Power analysis results will not be predictive of the actual results unless the power analysis is performed on data that closely matches the experiment. Please read about [generating pilot data](#) and [selecting preloaded datasets](#) before continuing.

○ **Upload Data**

Upload a file containing the number of reads per gene for pilot data as a tab delimitted text file. [See format info.](#)

[Choose file] No file chosen
Number of Replicates in Control: [2 ▾]
Number of Replicates in Test (enter 0 if none): [2 ▾]

◉ **Use a stored dataset(?)**

Choose a model dataset *(Less Accurate)*:

[Human Lymphocyte Cell Line - Cheung   ▼] Dataset
Descriptions

Cost Data (?)

Cost per replicate, excluding reads:
          Control: [10]
          Test: [500]
Cost per million reads sequenced: [45]   (?)
          Alignment Rate (to genes or transcripts): [50]
% (How to calculate?)

Constraints for Power Optimization(?)

Experimental Configurations to Test:

          Maximum number of biological replicates per condition: [10 ▼]
          Assess the power of sequencing depths between [10,000,000]
and [100,000,000]   reads aligned to genes per replicate

Leave the following fields blank to leave parameters unconstrained:

          Detect at least [50] % of expressed genes that are differentially
expressed by a [2] X fold change at p<[0.01]
          Experiment will cost no more than $ [1256320]   (?)
          Limit measurement bias by measuring at least [50] % of genes
with at least [50] % of maximum power (?)

[Optimize Design]

Results processing usually takes about 5 minutes.