Novel methodology for evaluation of next-generation sequencing measurements.

Dominic D LaRoche

September 14, 2015

Hypothesis I hypothesize that relative sensitivity (Mandel 1984) can be adapted to provide a novel method for evaluating the error associated with next-generation sequencing technologies.

1 Significance

- Quantification of RNA expression in a biological sample is a fruitful area of study.
- Many different methods exist to measure expressed RNA such as micro-array, RT-qPCR, and next generation sequencing.
- All existing methods include multiple steps and sources of measurement error.
- Future advancements in the use of measured RNA for understanding biological systems is dependent on reliable measurements.
  - Improved measurements will be needed if we wish to identify important biological processes associated with relatively low expression.
  -
- NGS methods are the newest and least well understood in terms of measurement error.

2 Innovation

- No current methods exist to directly compare the accuracy and precision of NGS based RNA measurement systems.
- Our proposed method will enable researchers to directly compare the precision and accuracy of competing NGS based measurement systems.
  - Our method will allow researchers to evaluate each step of the process for a given measurement system. E.g.:
    - Effect of sample preparation method
    - Effect of normalization method
    - Effect of sample type
  - Our method will enable researchers to compare systems for any given measurement range. E.g. some systems may perform better at lower expression levels whereas others may perform better at higher expression levels.
  - Our method will not require a known analyte concentration.

- We will publish an open-source software suite to implement our method with the open course statistical language R.

3 Methodology

- Relative Sensitivity, as outlined by Mandel in 1984, provides the framework for creating a model to compare measurement systems.
- We will need to re-formulate the model to accomodate the unqiue nature of sequence count data.
  - Several model formulations are possible including negative-binomial, over-dispersed Poisson, and zero-inflated Poisson.
  - We will estimate the parameters of the relative sensitivity models via a Bayesian framework and MCMC sampling.
  - We will examine the properties of the various model formulations via simulated data.
- Estimated relative sensitivity parameters will provide the information necessary to estimate the limit of detection and limit of quantitiation.
  - Can use the standard deviation of the estimated sensitivity ratios to identify the lower *measured* limit of quantitation.
  - Can use community network analysis based on probe-specific relative sensitivity standard deviations to identify non-expressed probes and use the values from these to identify the limit of detection.
- I am still struggling with an appropriate methodological framework to estimate individual calibration curves from multivariate expression data.
- We will provide implementation of our methods by publishing a package for the R programming language.
  - We will provide functions to:
    * estimate relative sensitivity for a given measurement range,
    * estimate relative LOD and LOQ for two measurement systems,
    * optionally interface with the BUGS/WinBUGS software for custom model fitting,
    * and estimate individual calibration curves (maybe).

4 Specific Aims

4.1 Aim 1: Direct Comparisons

Use relative sensitivity to compare measurement methodologies (including normalizations and platforms) without the need for individual calibration curves or known analyte concentrations.

- Relative sensitivity can provide the frame work for comparing the error of competing measurement systems without the need for known analyte concentrations
- Will accomplish this by re-formulating the relative sensitivity model to accomodate negative-binomial distributed count data.
- Methodological advancements for this aim will include characterizing the bias and error around the relative sensitivity measure to get confidence intervals around the estimates.

4.2 Aim 2: Estimate relative LOD and LOQ

Use estimated calibration curve to quantify the relative limit of detection and the limit of quantitation.

- Standard estimates for the limit of detection and limit of quantitation are not applicable to NGS data
  - We will use relative sensititivity to formulate new estimates for the limit of detection and limit of quantitation.
  - Understanding LOD and LOQ will be important for identifing measurement methods capable of reliably measuring low expression.

4.3 Aim 3: Individual Calibration Curve from Multivariate Observations

This may or may not be possible and I really don't have a good handle on an approach yet.

Formulate model for estimation of individual calibration curve from technical replicates on multiple probes.

4.4 Aim 4: Software

Create an R package that provides functions to estimate calibration curves and compare measurement platforms.

- No methodological advancement should occur in a vacuum so it is necessary to create accessible tools for researchers to implement the methodology.
- Create a suite of functions that assist with data importing, model fitting, measurement comparisons, and informative graphical outputs.