Novel methodology for evaluation of next-generation sequencing measurements.

Dominic D LaRoche

September 7, 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 Aim 1: Statistical Methodology

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

- Need some kind of information to constrain the problem and make the true measurement parameter estimable/ identifiable.
  - Estimation is the crux of the problem. Potential paths forward include:
    - Using multivariate information, e.g. from multiple probes on a sample run, to identify the curve without estimating the true counts.
    - Make some assumptions about the data, e.g. that there are "true" non-expressor probes in the assay.
- Need to include measurement error for non-linear model with non-constant (multiplicative) variance
- May need to incorporate multivariate calibration methods from analytic chemistry.
- Use "features" of the individual probes (length, sequence) to estimate likelihood of false positive/ false negative counts.
- Will most likely have to estimate under a Bayesian framework.

Potential Outcomes:

1. Model can be estimated and estimates are consistent with relatively small sample sizes.
   - Great! This is the outcome we are looking for!
2. Model can be estimated but estimates are either not consistent or require very large sample sizes.
   - If estimates are large sample consistent this is still an important advancement and future work may improve small sample properties.
3. There is no way to estimate the parameters of the model without additional information.
   - We know that we can still estimate *relative* sensitivity without knowing the true read count so this wouldn't be a dead end.

1.1 LOD and LOQ

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

- Once the calibration curve has been estimated the limit of quantitation (LOQ; the level at which measurements are reliable) will be the point at which the ratio of the slope of the calibration curve to the standard error of the calibration curve is sufficiently large (investigator must make this decision based on need)
  - If individual calibration curves cannot be reliably estimated we can still compare the relative LOQ between two measurement systems using the relative sensitivity.
  - Would not be able to identify a reliable LOQ for any one measurement system without individual curves.
- Limit of detection is the more difficult problem here. I haven't figured out the path forward for this one.

1.2 Direct Comparisons

Use estimated calibration curves to compare measurement methodologies (including normalizations and platforms).

- If individual calibration curves can be reliably estimated then we can compare different normalization methods for a single measurement system.
- If individual calibration curves cannot be reliably estimated then the relative sensitivity curve can be used to compare two measurement methods (including a given normalization method).

- Methodological advancements for this aim will include characterizing the bias and error around the relative sensitivity measure to get confidence intervals around the estimates.

2 Aim 2: 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.