Investigator: Dominic D. LaRoche
Proposal Title: Novel methodology for evaluation of next-generation sequencing measurements.

1 Significance

The accuracy and precision of many RNA measurement systems is largely unknown despite the importance of these measurements in modern biological research. We define an RNA measurement system as any sequence of procedures designed to quantify RNA from a biological sample and provide data for analysis. Our proposed research will provide a methodology for studying measurement error in NGS-based RNA measurement systems. Specifically, we will provide a method to directly compare the precision of any two RNA measurement systems using only technical replicates from each system. Even though RNA is widely studied, the precision of measurements may limit the quality of inference and slow the rate of scientific progress. Continued progress in fields that rely on NGS-based RNA measurement systems will depend on understanding how and why errors occur.

There is currently no robust methodology for comparing measurement precision between NGS-based RNA measurement systems. The current approach to comparing measurement systems is to indirectly compare each system's reproducibility and repeatability, an approach with limited utility (Lovell et al. 2015). Alternatively, others have used the final outcome measure, i.e. differential expression of genes, as a measure of performance. However, the outcome of an analysis is only a crude measure of measurement precision at best, even when using simulated data where the truth is known. Our method will allow direct comparison of the precision of any two measurement systems. This will allow for: 1) evaluating measurement systems for diagnostic procedures; 2) improving a single measurement system through experimentation to determine where errors arise; 3) evaluating the relative impact on precision from the numerous data normalization methods currently in widespread use; and 4) selecting the best measurement system and normalization method for any given experiment. Finally, we will provide all of our methodology as an freely accessible web application.

2 Innovation

We will apply a little known mathematical framework, Relative Sensitivity, originally developed by John Mandel in 1985 for evaluating measurement in analytical chemistry. This method does not require knowledge of the true amount of the sequences being measured, a quantity that is often difficult or impossible to know in RNA measurement problems. Relative sensitivity provides the framework for making comparisons among normally distributed measurements. Our approach is conceptually innovative because we will need to derive the formulas necessary to apply relative sensitivity to less well-understood count data, often with multiplicative errors, arising naturally from NGS-based measurement systems.

A new methodology is only useful if it is accessible to a wide range of researchers and scientists. However, many advances in statistical methodology take considerable time for adoption by the general scientific community because of the barriers to implementation. Our approach is technically innovative because We will provide a software suite for implementing our methods through the recently developed web platform for R programs, shinyapps.io. This platform allows users to interact with statistical analyses pre-programmed in R through a web-based graphical user interface (GUI). The platform greatly reduces the barriers to method implementation because it does not require anything to be installed on user computers and can provide elaborate point-and-click documentation to guide the user through the analyses. We will provide our suite of methods as a dynamic GUI available for free to researchers who wish to compare measurement or normalization systems on their own data. This platform has never been used for widespread implementation of a new method but holds great potential as a translational statistics tool.

3 Specific Aim 1 Research Plan

3.1 Background and Rationale

There is no established method for directly comparing the precision and accuracy of NGS-based RNA measurement systems. The rapid adoption of NGS-based RNA measurement systems across a wide range of biological disciplines necessitates improved understanding of the precision of these systems. For manufacturers of these systems, understanding the precision of a measurement enables improvements. The ability to compare measurement systems would also enable scientists to select the optimal system for their experiment. Currently, measurement systems are evaluated on the basis of repeat-ability and reproducibility but this provides no information on the actual accuracy and precision of the measurement system. Moreover, there is no way to directly compare existing metrics of repeat-ability and reproducibility among competing measurement systems due to the differences in scales and underlying distributions of the measurements.

The objective of this aim is to establish a methodology to directly compare two measurement systems. We hypothesize that the method of relative sensitivity will provide a powerful and useful framework for making these comparisons. The theory of relative sensitivity was originally developed by John Mandel in 1984 as an extension of his original work on the sensitivity of analytical chemistry measurements (Mandel 1957). Relative sensitivity has several key properties which make it suitable for this application. For example, relative sensitivity is not affected by the scale of the measurement, which can vary substantially between NGS-based measurement systems. Moreover, most measurement systems include some monotone transformation of the final result for analysis, of which there are many choices, and relative sensitivity is invariant to these transformations.

We will use both simulated and real data from 4 measurement systems to evaluate the estimates of precision and accuracy we generate from the relative sensitivity framework. Simulated data is necessary as the truth of real-world measurements can never be known. We will focus on simulating the measurement process, rather than the final result as is typically done, so that we can understand the impacts of each step in the measurement process. We will confirm the utility of our method by evaluating the accuracy and precision of four competing measurement systems: HTG EdgeSeq, NanoString nCounter, Illumina RNASeq, and TaqMan Gene Expression assays. We will use technical replicate samples and quantify each sample replicate 24 times on each platform. We have already completed sample quantification on HTG EdgeSeq and NanoString nCounter.

A metric for comparing NGS-based measurement systems will have immediate utility in identifying optimal measurement systems and sources of measurement error. This method will provide a critical window into the performance of measurement systems for both consumers of these systems and the manufacturers of these systems. Due to the diverse nature of the technology of systems currently in use, it is likely that the performance of these systems will correlate with elements of the technology utilized. We believe understanding these differences in performance will be important for mitigating problems at the development level and selecting the appropriate method for a given experiment.

At the completion of this aim we expect that we will provide an implemented methodology for comparison of any 2 NGS-based measurement systems such that a scientist with little understanding of the underlying theory of relative sensitivity will be able to implement the methodology and interpret the results.

3.2 Experimental Plan

The theory of relative sensitivity was initially presented in John Mandel's book *The Statistical Analysis of Experimental Data* (1984) but was never widely adopted in the analysis of measurement systems outside of analytical chemistry (with few references even within analytical chemistry). However, we believe the theory of relative sensitivity provides a simple, yet powerful, statistical framework for the evaluation of complicated NGS-based measurement systems.

The theory of relative sensitivity is an extension of Mandel's work on estimating sensitivity curves (1957) which removes the need to know the actual analyte concentration in order to evaluate the precision of a measurement. The sensitivity of a measurement is the slope of the functional relationship between a property of interest and its measurement, $Y = aX + b$, where Y is the measured value, X is the true value, and a represents some unknown relationship between these values. The function aX can be non-linear in practice. Steep slopes (a) correspond to greater sensitivity because a steep slope results in large differences in the measured value for small differences in the corresponding property being measured (fig. 1). The error around a sensitivity curve for a measurement process also affects the utility of the measurement. For example, if the slope is small but there is little error then the measurement can still discriminate between different states of the property whereas large error can "swamp out" a steeper slope (fig. 2).

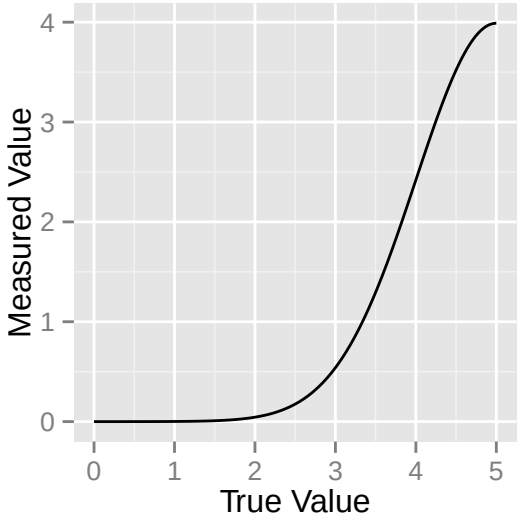

Figure 1: A single sensitivity curve for a measurement system representing the relationship $Y = aX + b$, where aX is some unknown non-linear function of X. The sensitivity of the measurement is greatest between the true values of 3 and 4.5 and smallest at 0 and 5. This measurement system would not be a good choice for those interested in measuring true values of less than 2 since the method is unable to differentiate between these values.

In order to construct and evaluate a sensitivity curve as described above one must know the true value of the property being measured. For many NGS-based measurement systems this property is either unknowable or is exceedingly difficult to know. However, relative sensitivity can be used to compare two measurement systems without knowing the true state of the underlying property being measured. Mandel (1984) constructed relative sensitivity a way to compare two measurement systems:

$$RS(Y_1/Y_2) = \frac{|dY_1/dY_2|}{\sigma_{Y_1}/\sigma_{Y_2}}.$$

In this formulation, dY_1 and dY_2 represent the slopes of the respective measurements while σ_{Y_1} and σ_{Y_2} represent the standard deviations of the slopes. This formulation does not involve the true value of the property being evaluated while providing a simple metric of relative utility for two measurements. Relative sensitivity can also be formulated as a function of some unknown X, in which case the relative sensitivity can vary over some measured range.

We will use a Bayesian approach the estimate the parameters of the relative sensitivity curve. The original formulation by Mandel suggested directly estimating the slope of the relative sensitivity curve by regressing y_1 onto y_2 and estimating standard deviations from replicated measurements. Unfortunately, this method assumes no measurement error in y_1 which is clearly not plausible. We will, instead, use a Bayesian approach to simultaneously estimate the slope and measurement errors.

We will test the utility of our method by simulating the measurement processes under consideration. Frequently, investigators will simply simulate the final measurement using assumptions which favor the method under investigation (probably need a citation here). We will simulate the measurement process for each of the 4 measurement systems under consideration. This approach will have 2 benefits: 1) we believe the final simulated data set will more closely match real data likely to be encountered by end users, and 2) we will be able to investigate how perturbations for a given step within a measurement system affect the final utility of the measurement. We will also test our method using real data collected from all four measurement systems under review. We have already collected this data for identical samples on the HTG EdgeSeq and the NanoString Ncounter platforms.

We will implement the method in the open source statistical package R and provide a open source package of functions to import data and compare measurement systems. We will also provide a graphical user interface to this package by creating an interactive web page that allows users to point and click through an analysis. We will host the application through the free shiny.io service provided by Rstudio.

3.3 Expected Outcomes, Potential Problems, and Alternative Strategies

At the completion of this aim we expect to have an implemented methodology which can compare any two NGS-based measurement systems in terms of precision and accuracy. We expect to have our method freely available for use by scientists through a user-friendly graphical user interface available to anyone with an internet connection. We also believe the FDA will benefit from our methodology due to the increasing use of NGS-based measurement systems to make clinical decisions.

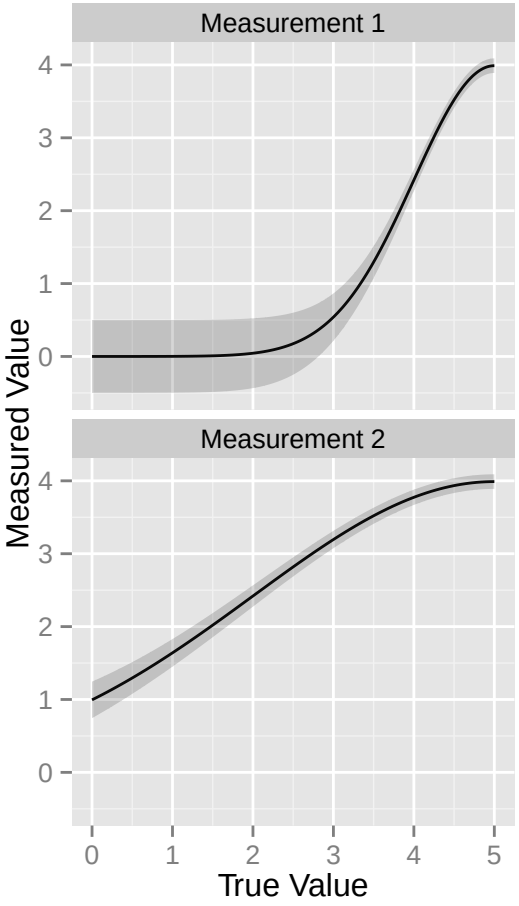

Figure 2: A comparison of sensitivity curves for 2 hypothetical measurement systems. Here measurement 2 is clearly superior to measurement 1 for true values less than 3, whereas measurement 1 is clearly superior to measurement 2 for values greater than 3.5. The choice of the optimal measurement method is dependent on the expected range of the true value being measured.

The utility of relative sensitivity to evaluate subtle differences in the performance of NGS-based measurement systems is obviously unknown. It is possible that we will be unable to differentiate between existing systems using our proposed estimation of the relative sensitivity curve. Given this scenario we will still be guaranteed to identify a upper bound on the difference in precision between any two measurements using information obtained through simulations.

If we are unable to detect a difference in precision between measurement systems using our proposed estimation scheme we will re-formulate our estimators based on the maximum-likelihood. These estimators will require more assumptions about the data but will likely yield smaller confidence intervals around our estimates of precision due to these assumptions. Although we believe our proposed method with fewer assumptions is more desirable, the assumptions made by the maximum likelihood estimates are common when dealing with NGS-based data so we feel this method would also provide utility to scientists and manufacturers.

4 Aim 2 Evaluation of frequently used normalization procedures.

4.1 Background and Rationale

Data from NGS-based RNA measurement systems must be normalized prior to most subsequent data analyses to remove systematic technical effects. There are currently over 7 methods to normalize NGS-based measurement data in widespread use and no clear guidance exists for selecting a method. A recent comparison of 7

popular methods by Dillies et al. (2012) was ambiguous about the optimal normalization method and under what conditions optimality would hold. Two years after the comparison was published we still see widespread use of nearly al of the methods compared, indicating uncertainty among scientists and statisticians about the optimal method. At best, researchers are selecting methods based on limited, and sometimes conflicting, information. At worst, researchers are selecting a normalization method which they find to be optimal in terms of the results they expect (or would like) to see.

Our aim is to create an objective measure of the relative efficacy of the 7 most popular normalization methods so that researchers can select the optimal method for their data *without* basing this decision on the final outcome. The normalization procedure can be thought of as part of the measurement system. As such, we hypothesize that relative sensitivity can be adapted to provide a relative measure of performance for any pair of normalization methods. By comparing the impact of each type of normalization on the relative precision of the measurement we will provide necessary information for researchers when selecting a normalization method.

The current practice, unintentionally suggested by Dillies et al., of selecting a normalization method based on 'optimal' final results could lead to an increase of false discoveries due to the circular nature of the decision. Unfortunately, without some objective measure or clear guidance this practice will likely continue. Our research will provide an objective measure and guidance about the impacts of all 7 commonly used normalization methods. We believe this methodology will improve the quality of research across the wide range of research areas which utilize this type of data.

4.2 Experimental Plan

Existing normalization methods are based on different assumptions about the underlying nature of the observed data (lots of refs). We will simulate observed data under a variety of these assumptions in order to compare normalization methods under conditions when assumptions are and are not met. We will use relative sensitivity to evaluate each normalization method against each simulated data type to determine when each method is either appropriate or inappropriate and optimal or sub-optimal. We will define a method as being appropriate when the inferences that would be made from normalized data (e.g. differential expression results) are consistent with the truth. We will define an optimal method as the method normalization which produces final measurements with the highest precision across the entire range of measurement.

Some methods may outperform others in only a particular range of measurement. We will divide the range of measurement into high, medium, and low expressed probes by dividing the simulated measurements into quadrilles. Low expressed probes will be defined as those less than the first quartile, medium expressed as those between the 1st and 3rd quartiles, and high expressed as those above the 3rd quartiles. Since different normalization methods are likely to place different sets of probes into each of these categories we will identify which probe is assigned to each category based on the random variable, λ, associated with the expected count for that simulated probe.

Simulations provide important information about the performance of normalization under tightly controlled conditions where the truth is known. However, real data often do not conform with the simplified and controlled nature of simulations. Therefore, we will also compare normalizations on real data using the technical replicate samples quantified on each of the 4 platforms identified in aim 1. We will use relative sensitivity to evaluate the effect of each normalization method for the data generated by each platform.

4.3 Expected Outcomes, Potential Problems, and Alternative Strategies

Previous research has shown many normalization methods to perform similarly on real observed data (Dillies et al. 2012; Rapaport et al. 2013). Therefore, we expect that the performance, as measured by relative sensitivity, of each normalization method to also be quite similar. However, we believe relative sensitivity will provide much

more information about the precise nature of the differences in normalization methods than previous comparisons. Previous comparisons of normalization methods relied on using the outcome, e.g. the number of differentially expressed probes, as the measure of the effect. The outcome is a very crude measure of the effect of a normalization procedure and has been unable to differentiate among competing procedures in previous research. Relative sensitivity has the potential to directly compare the effects of each normalization method allowing the direct comparison of each method with respect to any, or all, measurement range(s). We believe this level of detail will enable us to differentiate the impacts of closely related normalization methods such as TMM (Robinson and Oshlack 2010) and Quantile (Amaratunga and Cabera 2001) methods. The nature of these differences will highlight when each of these methods is optimal and where each method performs poorly.