Section 2 Methods

2.1 Platform Assessment - Background

To fully characterize the performance of any gene expression platform requires a comprehensive set of data with engineered truth that can be summarized to provide meaningful performance indicators.

It is useful to examine the evidence that Illumina presents to support its RNA-Seq platform. On the Illumina web site there is a page titled **Advantages of RNA-Seq technology** with a section on Benefits of RNA-Seq vs. Microarray Technology. The benefits are described as falling in the following categories:

- **Ability to detect novel transcripts**: Unlike arrays, RNA-Seq technology does not require species- or transcript-specific probes. It can detect novel transcripts, gene fusions, single nucleotide variants, indels (small insertions and deletions), and other previously unknown changes that arrays cannot detect.[4,5]
- **Wider dynamic range**: With array hybridization technology, gene expression measurement is limited by background at the low end and signal saturation at the high end. RNA-Seq technology produces discrete, digital sequencing read counts, and can quantify expression across a larger dynamic range (>105 for RNA-Seq vs. 103 for arrays).[4–6]
- **Higher specificity and sensitivity**: Compared to microarrays, RNA-Seq technology can detect a higher percentage of differentially expressed genes, especially genes with low expression.[7–9]
- **Simple detection of rare and low-abundance transcripts**: Sequencing coverage depth can easily be increased to detect rare transcripts, single transcripts per cell, or weakly expressed genes.

See Illimuna sequencing vs micro-array platform summary for a brief review of the cited papers.

In evaluating the performance of the qNPA TTx panel, we will focus on **sensitivity and specificity of differential expression assessment**. The reason for this focus is that accurately in detecting differential expression is at the core of the most common genomic scale gene expression analysis tasks, be it the biomarker discovery, outcome prediction or sub-group discovery. We will also explore dynamic range properties of the platform which is of interest only to the extent that it affects differential expression the limits of detection differential expression analysis. The detection of novel transcripts is not an application that the TTx panel can be used for. The detection of rare transcripts is also not an application that is being targeted by the current qNPA TTx panel as it requires expensive, eep sequencing datasets for its implementation and assessment.

The role of RNA-Seq Data

In evaluating the qNPA TTx panel platform it will be helpful to compare performance metrics with RNA-Seq as this provides a benchmark against which we can assess performance. We should bear in mind that the RNA-Seq platform performance is strongly dependent on context - **the quality of the input RNA**, for instance - and sequencing parameters, such as depth of sequencing. As the **qNPA assay requires small amounts of RNA and is resistant to RNA degradation**, this context is the one in which the qNPA TTx panel platform is expected to shine. To fully assess the qNPA TTx panel platform performance thus requires a dataset that is rich enough to enable performance assessment on good quality RNA, in which context RNA-Seq can be considered a gold-standard, as well as degraded RNA in limited supply, in which context the qNPA TTx panel platform is expected to gain an edge. When using FFPE blocks of varying ages, storage conditions, and handling of source of RNA, some level of degradation is expected. That factor alone does not provide a means of assessing the effect of RNA quality on relative platform performance. In order to assess the effect of RNA quality on relative platform performance one needs to control that factor in the design of the experiment, either by using RNA of known and differing quality (getting RNA from cell cultures vs FFPE blocks with verified RNA degradation), or by manipulating the RNA quality by artificial means, as was done in the experiment reported in Holik Holik et al. (2017) [10].

Our assessment of differential gene expression assessment results will be based on data collected in the context of properly designed experiments and controls including:

- Spike-in RNA Controls
- Mixture Experiment Data
- Well Characterized Samples

2.2 Spike-in RNA Controls

Metrics derived from analysis of external spike-in RNA control mixtures can be used to assess technical performance. Spike-in cocktails can be customized and matched to samples to produce specified ratios between samples and enable assessment of performance in 1) differentially expressed transcript identification, 2) ratio limit of detection, and 3) estimated ratio variability and bias (Munro et al. (2014) [11]).

A library of 96 external RNA spike-in controls developed by the External RNA Controls Consortium (ERCC) [12] and distributed by NIST as Standard Reference Material 2374 [13] can act as a source of technology-independent controls for differential expression experiments [11]. In any differential expression experiment, a pair of ERCC control ratio mixtures can be added ('spiked') into total RNA samples such that for each ERCC control the relative abundance of the control between samples (ratio) is either of known difference (a true-positive control) or the same (a true-negative control).

Figure 1 and 2 illustrate a typical spike-in design.

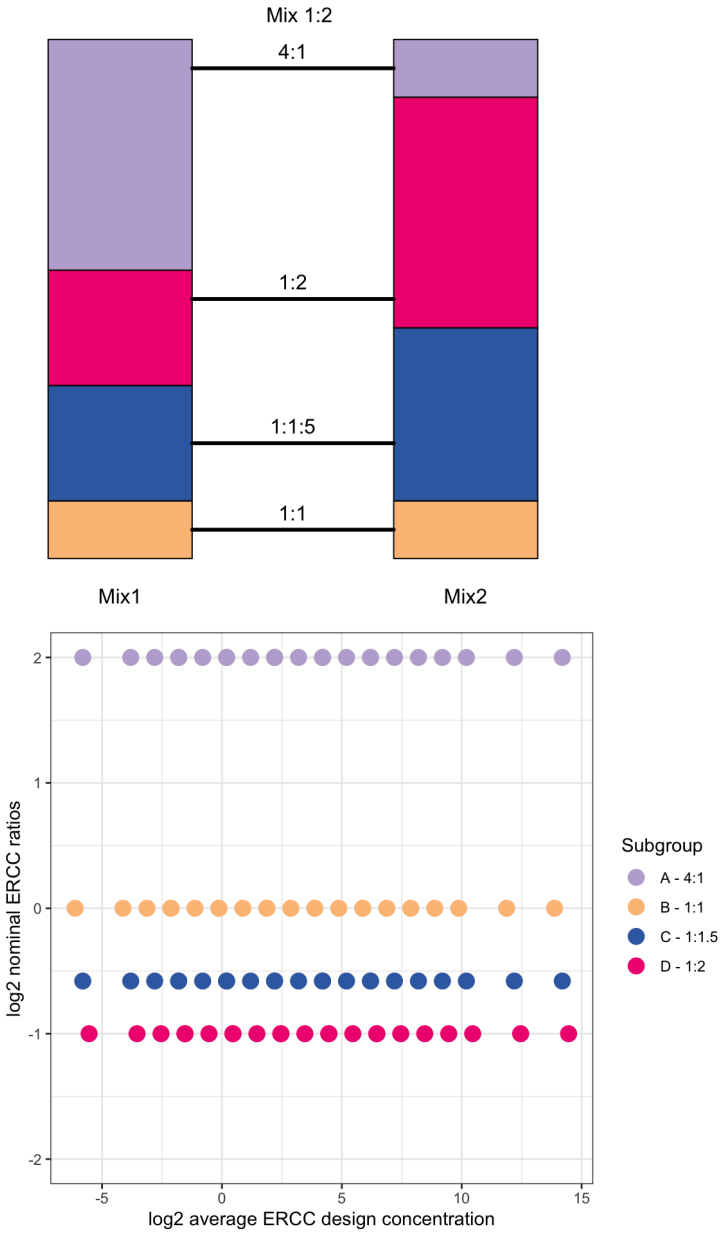


Figure 1. Two mixtures of the same 92 ERCC RNA transcripts are prepared such that four subpools with 23 transcripts per subpool are in four defined abundance ratios between the two mixtures. Within each subpool the 23 controls (several points overlap) span a broad dynamic range of transcript concentrations.

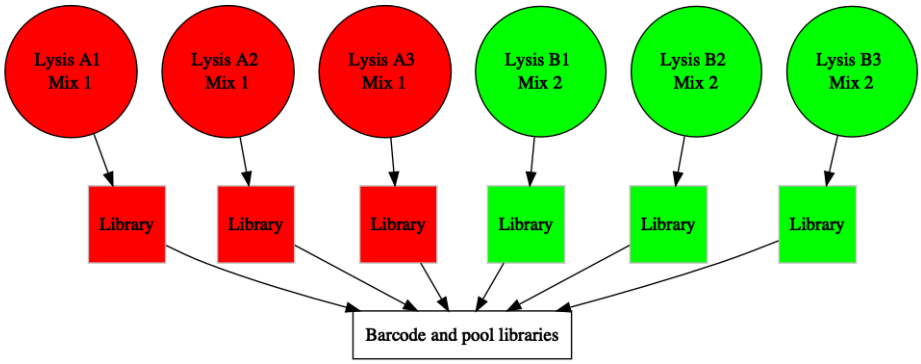


Figure 2. The biological replicates are prepared from two sources of RNA, A and B. In order to minimize biological variability while maintaining sample processing variability, we extract 3 lyses from the same biological specimen and spike-in different ERCC mixes in the different lysis sets.

Munro et al. [11] illustrated the use of the spike-ins in two differential expression experiments: a rat toxicogenomics experiment (Fig 1.c) and a Universal Human Reference RNA (UHRR) vs Human Brain Reference RNA (HBRR) experiment (Fig 1.d). In our assessment, we will use samples from the mixture experiment described below to provide the differential experiment context for the spike-in data assessment. A differential expression context is required to properly assess the significance of the observed spike-in ratios as the significance of the spike-in ratios needs to be assessed in relation to a *representative* distribution of “background” ratios - the gray dots in the following figure also take from Munro et al. [11] (Figure 5 in the article, reproduced using R erccdashboard package):

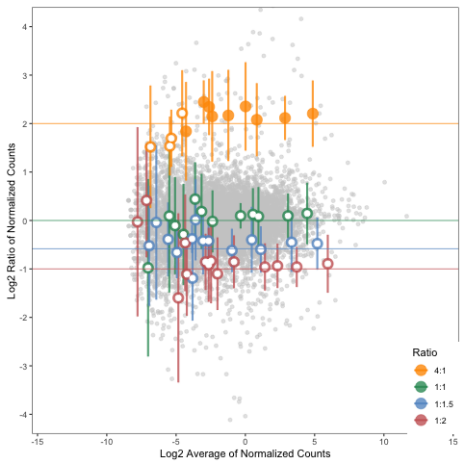


Figure 4. Shown for biological replicates (n=3) of control (CTL) and methimazole treated (MET) from a rat toxicogenomics experiment. ERCC data points (colored by ratio) represent the mean ratio measurements per ERCC. Error bars represent the standard deviation of replicate ratios. Filled circles indicate ERCC ratios above the LODR estimate for 4:1, 1:1.5 and 1:2 ratios. Endogenous transcript ratio measurements are shown as gray points. The nominal ratios are annotated with solid colored lines for each ratio subpool.

The significance of the spike-in ratios depend on the variability in the distribution of all assessed ratios in a non-trivial manner. See Storey and Tibshirani (2003) [14] for a theoretical foundation for measuring statistical significance in genomewide studies. What constitutes a *representative* distribution of “background” ratios is not immediately obvious. The two examples described in the Munro paper, provide contrasting levels of variability, with many genes being differentially expressed between UHRR and HBRR samples, but few between treated and untreated rats. The mixture experiment, which we describe next, will provide varying levels of differential expression for the assessment of spike-in significance.

2.3 Mixture Experiment

Mixture experiments (sometimes called titrations) entail mixing different sources of RNA in various ratios to produce predictable ratios for all transcripts that are differentially expressed in the RNA sources. Holik et al. (2017) [10] illustrate the use of mixture experiments to compare alternative algorithms for analyzing RNA-Seq data. We will apply these methods to the problem of platform performance assessment.

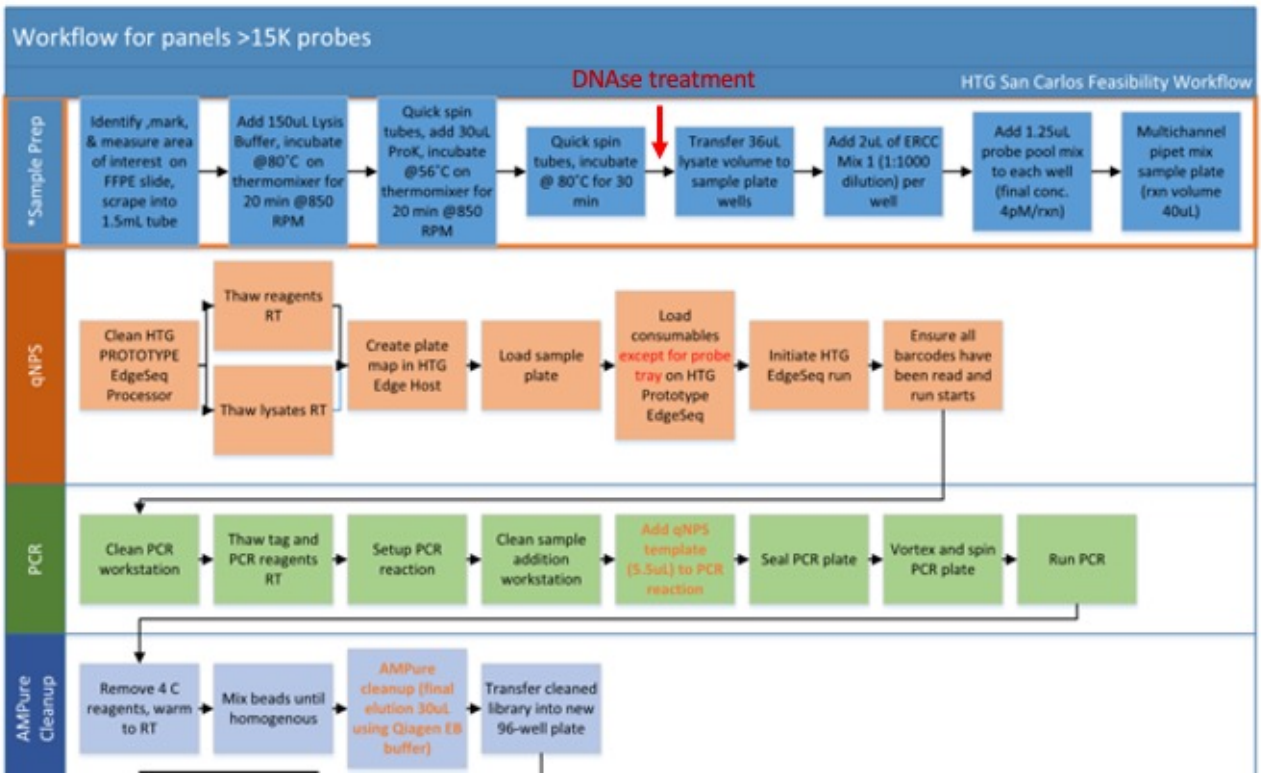
In a mixture experiment, two sources of RNA, A and B, are mixed in specified ratios, 75:25, 50:50 and 25:75, for example. In one analysis, gene identified as differentially expressed in the comparison between the pure samples, A and B, are used as **truth** in evaluating the genes identified in the comparison between the mixed samples. In another analysis, the data from the entire mixture series are used to “predict” the log-ratios when the mixture samples are compared. The observed log-ratios can then be compared with the predicted log-ratios to get a sense of platform accuracy.

2.4 Well Characterized Samples

Spike-in and mixture experiments provide a rich dataset for platform performance assessment but have some limitations and leave some questions unanswered. Spike-in transcripts are limited by their number and the fact that synthetic DNA may not capture all sources of variation that endogenous sequences are subject to. Targets available for differential expression assessment in mixture experiments are greater in number but may not reflect biological systems of interest both in terms of the magnitude and the number of differences in expression between groups. There is also the issue of platform fidelity in the context of small amounts and degraded input material which cannot be unambiguously resolved by these data. To address questions that remain unaddressed by the spike-in and mixture data, we will examine differential expression in well-characterized samples that represent realistic and representative biology. This experiment should include enough samples to enable us to assess the full range of between sample variability and should be designed to enable us to assess the effect of RNA quality on performance.

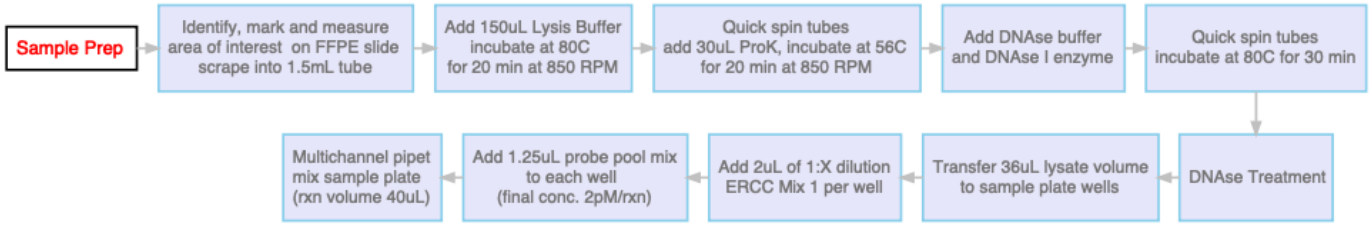
This arm of the study still needs to be designed.

2.5 Laboratory Methods



Placeholder Figure - Workflow for panels > 15K probes.

2.5.1 Sample Prep



Placeholder Figure - Sample Preparation