ABRF GVRG 2011 Research Poster

Development of ERCC RNA Spike-in Control Mixes

Luming Qu, Andrew Lemire, Kristi Lea, Diana Batten, Sheila Heater, Jian Gu, Penn Whitley, Joel Brockman, Jeff Schageman, Kelli Bramlett.
Life Technology. 2130 Woodward St., Austin, TX 78744.

Abstract

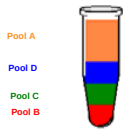
One outcome of the External RNA Controls Consortium (ERCC) effort is a National Institute of Science and Technology (NIST) candidate standard reference material (SRM) containing 96 control sequences. The SRM contains 96 DNA plasmids containing certified sequence. The plasmids inserts generate a set of transcripts, from 250nt to 2000nt in length and containing poly(A) tails, that mimic natural eukaryotic mRNAs. Two complex ERCC RNA control mixes (ERCC RNA Spike-In control Mix 1 and Mix 2) are generated using those transcripts. Each mix contains 92 RNA transcripts and spans over 6 logs of dynamic range with balanced GC content and length. Each mix is further divided into 4 sub-pools that can be utilized to evaluate fold-change measurements of gene expression between the two mixes. In addition, TaqMan® assays for the 92 transcripts have been designed to serve as tools to QC the ERCC mixtures themselves as well as monitor the controls through library preparation steps. These control mixes can be spiked-in to total or poly(A) RNA prior to processing for gene expression studies. We have used the ERCC mixes to spike-in to RNA that is used to create a cDNA library for transcriptome sequencing. We also demonstrate the use of TaqMan® assays to eight of the ERCC controls to inform library quality prior to continuing with the expensive and time-consuming next generation sequencing experiment. Overall our data demonstrates the potential of these ERCC control mixes to evaluate gene expression measurements using multiple systems. In addition, the controls are invaluable as an in-process quality control check for library amplification processes prior to next generation sequencing experiments.

Introduction

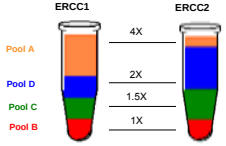
Existing RNA controls are platform specific, there is strong need to develop "positive controls" that can be universally applied across multiple gene expression measurements to help monitor platform performance. The ERCC controls have been designed to mimic natural eukaryotic mRNA sequences with the ability to be used across multiple measurement platforms. Our main focus is to create mixtures utilizing the NIST reference material to measure reproducibility, sensitivity, and robustness in gene expression experiments. The major challenge is robust manufacturing of these controls into easy to use formats with minimal lot to lot variability due to the large number of RNA transcripts in the mixtures.

Figure 1. ERCC Control Mix Products

ERCC one tube sku



ERCC two tube sku



Ratio between ERCC 1 and ERCC2

Subgroup	Mix1:Mix2
A	4.0
B	1.0
C	0.67
D	0.5

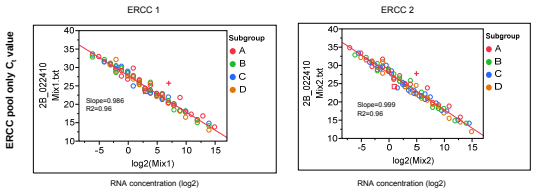
The 92 transcripts in each mix are equally divided into 4 sub-pools. Pool A (orange), pool B (red), pool C (green), and pool D (blue). Each subpool contains 23 transcripts covering a million fold dynamic range that are balanced for RNA transcript size and GC content.

Materials and Methods

Approximately 1ug of individual SRM plasmid DNA from NIST were linearized by restriction enzyme digestion, *in vitro* translated using MEGAscript® T7 Kit, followed by bead purification using MagMAX™-96 Total RNA Isolation Kit to generate 92 individual RNA transcr ipts. RNA concentration was measured using Quant-iT RNA Broad Range kit. RNA purity and integrity were verified by running 48 well E-gels. The 92 high-quality RNA transcripts were pooled using a Tecan instrument program to make the final mixtures robotically. TaqMan® Gene Expression assays targeting the 92 transcripts were utilized to quality control the manufacturing process of the ERCC mixtures. These same TaqMan® assays have also been used to monitor progress through the transcriptome cDNA library preparation steps required to create a next generation sequencing (NGS) library from RNA. Libraries shown here were generated from Hela poly(A) or total RNA containing ERCC Mix 1 or ERCC Mix 2 using the SOLiD™Total RNA-Seq Kit (STaR-Seq) and sequenced using the SOLiD™ system.

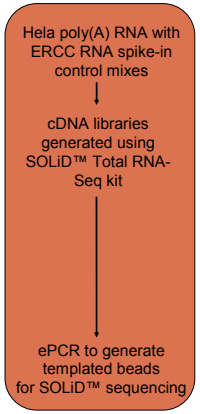
Result

Figure 2. TaqMan® Analysis of ERCC Control Mixes Illustrating Dynamic Range



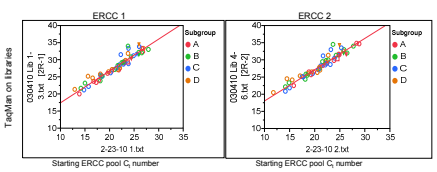
Shown are expected concentrations from ERCC mix formulations on the X-axis plotted against Ct values generated by TaqMan™ analysis of the ERCC mixes on the Y-axis.

Figure 3. Experimental Workflow



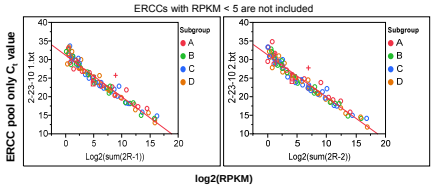
Graphs show Ct values generated from the initial ERCC pools (see Fig. 2) on the Y-axis plotted against log2 RPKM values for the ERCC transcripts after SOLiD™sequencing on the X-axis.

Figure 4. Transcriptome Library QC



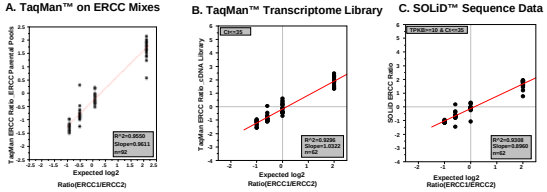
Libraries were generated with STaR-Seq kit from HeLa Poly(A) RNA. Data below includes 56 of 92 ERCCs that are detected at Ct < 35 in all libraries tested. Graphs show Ct values generated from ERCC mixes (see Fig. 2) on the Y-axis plotted against Ct values from cDNA transcriptome libraries on the X-axis.

Figure 5. TaqMan™ and SOLiD™ Concordance



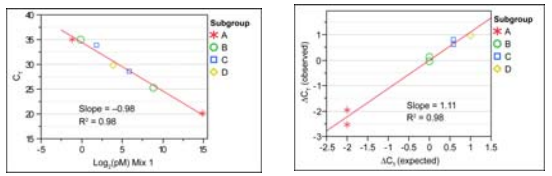
For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.

Figure 6. Expected Concentration Ratios Between ERCC Mixes vs Measured Ratio Change



Graph A shows the expected Log2 transformed concentration ratios between ERCC transcripts in Mix 1/Mix 2 on the X-axis compared to the actual ratios generated from TaqMan™ Ct values on the Y-axis. **Graph B** The same comparison in generated utilizing the TaqMan™ assays on the STaR-Seq libraries (Y-axis) vs expected ratios (X-axis) in; **Graph C** SOLiD™ sequence data (Y-axis) vs expected ratios (X-axis) in

Figure 7. Recommended Sequencing library QC linear regression and fold change



We have identified a set of eight ERCC controls and their corresponding TaqMan™ assays that are recommended for library quality control (QC) measurements that can inform a researcher about the library quality prior to costly downstream gene expression experimentation. The graph on the left shows the expected concentration of these 8 transcripts in ERCC Mix 1 (X-axis) plotted against the actual Cts (Y-axis) generated from a cDNA transcriptome library prior to next generation sequencing. The graph on the right shows the expected fold change ratios between the two ERCC Mixes of these selected 8 control transcripts expressed as ΔCt (X-axis) plotted against the actual observed ΔCt values from the transcriptome libraries.

Conclusions

- Our real time qRT-PCR, PCR and SOLiD™ data show the ERCC controls mixes can serve as a very powerful tool for gene expression analysis across multiple platforms
- The ERCC controls provide an alternative for normalization in relative quantification analysis, assessment of technical performance of cDNA generation, and deliver certified controls to monitor RNA based work flows and data analysis for next generation sequencing systems
- This initial product from the NIST traceable ERCC plasmids has great potential in the hands of researchers to open-up new capabilities in understanding variability in RNA library preparation, determining detection limits of measurement systems, and informing downstream analysis

Reference

1. The External RNA Controls Consortium: a progress report. *Nature Methods* - 2, 731 – 734 (2005).

Acknowledgements

I would like to thank all the members of the Transcriptome Analysis Group and the Bioinformatic team from the Austin R&D group for their help and suggestions during the development of these control products.