ERCC 2.0 Workshop

July 10 – 11, 2014 Stanford University, California

Hosted by the National Institute of Standards and Technology - Advances in Biological/Medical Measurement Science (NIST-ABMS) Program

Sarah Munro, NIST-ABMS Program
Marc Salit, NIST-ABMS Program

Executive Summary

The External RNA Controls Consortium (ERCC) was reconvened on July 10 – 11, 2014 at Stanford University for the ERCC 2.0 Workshop – hosted by the NIST-ABMS program and led by Marc Salit and Sarah Munro. The ERCC 2.0 effort objective is to develop an updated suite of RNA controls. RNA measurement technologies and our understanding of RNA biology have advanced significantly since the original set of ERCC controls was developed.

The ERCC 2.0 workshop was held immediately following the ENCODE Project meeting to encourage participation from that scientific cohort. There were over 65 participants (including remote attendees) representing industry, academia, government, and other non-profit institutions (for full list of all meeting registrants see Appendix I). Meeting participants presented their experience developing and using the original ERCC controls as well as their proposed designs and current development efforts for building an updated and expanded suite of RNA controls.

The charge to the workshop was to reach consensus on the scope and arc of new ERCC work. A 'strawperson' proposal was presented and refined to have the multiple stakeholder groups working in parallel on design of a variety of different control cohorts including transcript isoforms, new and improved mRNA mimics, miRNA, and cancer fusion transcripts.

This working group structure will enable rapid development of the many new RNA control products that are called for by the ERCC membership and broader scientific community. Working group conference calls are planned to initiate ERCC 2.0 product design work. Following this "divide-and-conquer" approach to design, commonalities between working groups will be established to support efficient ERCC 2.0 product development and analysis efforts.

For information contact:

Sarah Munro (sarah.munro@nist.gov) and Marc Salit (salit@nist.gov)

ERCC 1.0 Success and Motivation for ERCC 2.0

The original ERCC external RNA spike-in controls have been broadly adopted by the academic and industry research communities to support numerous applications including product and method development, assessment of quality control, sample normalization, single-cell measurements, and others. Although these spike-in controls were originally designed and intended for cDNA microarray measurements, the explosion in RNA sequencing (RNA-Seq) capabilities has lead to widespread adoption of the ERCC controls for RNA-Seq measurements. For a brief summary of ERCC 1.0, see Box 1.

At the workshop NIST presented examples of ERCC 1.0 user feedback from scientists at Expression Analysis (Wendell Jones and Erik Aronesty)[†]. Their comments detailed the value they have found in using the ERCC 1.0 controls and their suggestions for improved RNA controls and methods. These suggestions included the development of transcript isoforms (splice variants) and methods for using spike-in controls with FFPE (Formalin-Fixed Paraffin-Embedded) samples. Such input from ERCC 2.0 participants are essential to guide our collective effort to build new ERCC 2.0 controls.

The ERCC 2.0 effort is an excellent opportunity for all interested parties to come together to develop an updated suite of RNA controls. Our understanding of RNA biology and RNA measurement capabilities have advanced significantly since the first set of ERCC controls were developed. We can build new RNA controls that meet the diverse RNA measurement needs that have become available, such as quantitation of transcript isoforms, miRNA, and cancer gene fusion transcripts. With new controls we can also address the recognized limitations of the original ERCC 1.0 controls (e.g. splice structure, poly-A tail length, and GC content, etc.).

Box 1. ERCC 1.0 Summary

The ERCC 1.0 effort was an industry-initiated, NIST-hosted consortium to develop external RNA spike-in controls for genome-scale RNA measurement method validation.

The original ERCC controls are available as a NIST standard reference material (SRM) 2374) – a DNA Sequence Library for External RNA controls: https://www-s.nist.gov/srmors/view_detail.cfm?srm=2374.

SRM 2374 is the DNA source for production of external RNA spike-in control mixtures, which are intended and openly available for use in commercial products by any entity.

In addition to SRM 2374 and derived commercial reference material products, other products have also been developed by the ERCC including a documentary standard (CLSI MM16-A) and the 'erccdashboard' open source software tool to enable reproducibility of gene expression experiments.

Charge to the Workshop

NIST gave a charge to the workshop to approach consensus on the scope and arc of an ERCC 2.0 effort to develop an updated suite of RNA controls. Participants were invited to

[†] This presentation and others from the workshop are briefly summarized in this report, for more details see presentations available at http://www.slideshare.net/ERCC-Workshop/presentations.

consider their shared interests to develop the ERCC 2.0 portfolio, which might include controls, analysis tools, and documentary standards. Participants were also asked to consider consortium operation and structure including formation of working groups and a steering committee.

ERCC 2.0 Principles of Operation

Principles of operation for ERCC 2.0 that were fully consistent with and inspired by ERCC 1.0 were put forth by NIST. As in ERCC 1.0, consensus decision-making will be the standard for ERCC 2.0. ERCC 2.0 will be NIST-hosted, but leadership for the working groups and steering committee will be contributed by ERCC 2.0 participants. ERCC 2.0 will be a data-driven and technology-independent effort that is pre-competitive. There will be no formal documentation for consortium participation – it is expected that all interested parties will participate openly and voluntarily in ERCC 2.0.

Scope and Arc of ERCC 2.0

The proposed scope for ERCC 2.0 is to create an updated suite of standard controls for RNA measurements. Three working groups were proposed to address the design, development, and analysis of ERCC 2.0 controls. Objectives for each working group are shown in Box 2.

Box 2. Proposed ERCC 2.0 Working Groups

Design	Selection of the types of RNA controls Selection of sequences
Development	Building controls Developing & testing control mixtures
Analysis	Determine standard performance metrics Create tools as needed to support design and development

Products from ERCC 2.0 may include control sequences representing different types of RNA such as transcript isoforms, miRNA, and new mRNA mimics, documentary standards for using such controls, and performance metrics for assessing technical performance of experiments.

Logistical considerations for the arc of ERCC 2.0 will include decisions about the workshop number and frequency, communication methods such as use of teleconference calls, mailing list, wiki, etc. As work progresses development schedules and dissemination plans will need to be established. Development schedules will include the necessary steps to achieve ERCC 2.0 products with the intent to not only have clear 'mile markers', but also an appropriate 'finish line' for the ERCC 2.0 effort. A steering committee is needed to address business models for dissemination of ERCC 2.0 products (either from NIST or commercial entities) such as physical reference materials, documentary standards, and performance metric analysis tools.

ERCC 2.0 Process Discussion

ERCC 2.0 members can collectively define ERCC 2.0 activities and process for decision-making and operations. Beginning with this initial workshop, ERCC 2.0 participants are expected to reach consensus on the following questions:

1. What will we do together?
2. How will we do it?

NIST is committed to hosting the consortium and supporting product development. The consortium is expected to define the consortium mission (e.g. 'providing infrastructure to discern signal from artifact' and 'providing confidence in RNA measurement results') and the ERCC 2.0 product portfolio. Examples of potential ERCC 2.0 products include reference materials, reference data, analysis methods, analysis tools, and documentary standards.

Participant Presentations

There was an open invitation for meeting attendees to present their experience with the original ERCC controls and design ideas for a new cohort of RNA controls. See Box 3 for speaker names, affiliations and presentation titles. In Appendix II brief summaries of each presentation are provided. Publicly available slides are available at http://www.slideshare.net/ERCC-Workshop/presentations.

Box 3. Speakers and presentation titles

Speaker	Affiliation	Title
Bob Setterquist	Thermo Fisher Scientific	Ambion ERCC Spike-in Mix
Lukas Paul	Lexogen GmbH	Spike-in transcripts for the quantification of mRNA isoforms in Next Generation Sequencing
Anne Bergstrom Lucas	Agilent Technologies	ERCC Transcripts Give Confidence to the Performance of One-Color and Two-Color Microarray Experiments
Karol Thompson	U.S. Food and Drug Administration	Defined mixes of synthetic microRNAs: prototypes for controls for microRNA profiling assays
Christopher Mason	Weill Cornell Medical College	Epitranscriptome Standards
James Willey	University of Toledo	Use of ERCC synthetic spike in standards to validate competitive multiplex PCR amplicon library method for targeted NGS analysis
Tyson Clark	Pacific Biosciences	Single Molecule, Real-Time Sequencing of Full-length cDNA Transcripts
Elizabeth Tseng	Pacific Biosciences	Read count information and technical variability in PacBio full-length cDNA sequencing
Stephanie Pond	Illumina	Synthetic Spike-in mRNA-Seq data for Cancer Gene Fusion Detection
Sarah Munro	National Institute of Standards and Technology	Standard Performance Metrics for Gene Expression Experiments with the "erccdashboard"
Joel Myerson	Agilent Technologies	Efficient Chemical Syntheses of Long and Modified RNA Oligonucleotides

Working Group Scope Discussion

NIST presented a 'strawperson' proposal of three parallel technical working groups for ERCC 2.0, on the topics of 'Design', 'Development', and 'Analysis' (Box 2) and invited workshop participants to openly discuss this proposal. The conversation began with discussion of designing new controls and participants quickly reached consensus that multiple design conversations were needed to address development of different types of controls. Interest was expressed in controls for miRNA, transcript isoforms, non-coding RNA, cancer fusion transcripts, new mRNA mimics to address shortcomings of ERCC 1.0 controls, and refined formulations of ERCC 1.0 controls and corresponding method development to support the use spike-in controls in single-cell and FFPE analysis protocols. Comments from the working group scope discussion are in Appendix III.

Proposed Work Plan

The consensus at the end of the working group scope discussion, which primarily focused on design principles, was that the consortium might be most effective by initiating multiple design discussions prior to launching working groups to develop and analyze ERCC 2.0 products. The initial proposal of three parallel working groups for 'Design', 'Development' and 'Analysis', was revised in favor of beginning with multiple parallel ad hoc 'Design' working groups before initiating the other two working groups. Brief descriptions of candidate Design Working Group Topics are listed here in Box 4. Suggestions for additional design working group topic areas are welcome; this list of topics may be expanded.

Box 4. ERCC 2.0 Design Topics

Transcript Isoforms

- Applications, need, and use scenario for transcript isoform controls
- Lexogen transcript isoform controls (non-cognate)
- *Schizosaccaromyces pombe* controls for biologically sourced (cognate) sequence, but without homology to the human transcriptome
- Proposals for additional controls
- Study design

New and Improved mRNA mimics

- Applications, need, and use scenario for mRNA mimics, including single cell & FFPE samples, RNA-Seq, qRT- and digital PCR
- Improvements relative to ERCC 1.0 controls, e.g. GC content, length, poly-A tails
- Reformulation of ERCC 1.0 mixes for different applications
- Proposals for new mRNA mimic sequence library

miRNA

- Applications, need, and use scenario for miRNA controls
- Proposal and design principles for candidate controls, e.g. cognate or non-cognate, include pre-miRNA
  Proposals for additional small RNA controls (< 200 bp)

Cancer fusions

- Applications, need, and use scenario for cancer fusion controls
- Illumina/TGen cognate cancer fusion controls
- Proposals for additional cancer fusion controls

Conference calls will be organized for each of these design topic areas for participants to establish the measurement problem(s) that will be addressed by the proposed controls and identify the appropriate design criteria for candidate controls. Once these design principles are established we can begin initiating plans for development and analysis of each type of ERCC 2.0 RNA control product.

We invite all interested parties to join the ERCC 2.0 effort and welcome any additional feedback on this workshop report. We thank all workshop presenters and participants for their contributions. Jennifer McDaniel and Matthew Munson provided additional notes for this workshop report.

Appendix I: ERCC 2.0 Workshop Registrants

Name		Institution
Sean	Allen	TGen
Steven	Bauer	FDA/CBER/OCTGT
Adam	Bemis	Bio-Rad: DBC
Anne	Bergstrom Lucas	Agilent Technologies
Darren	Bertagnolli	Allen Institute for Brain Science
Nathan	Boley	UC Berkeley
Magnolia	Bostick	Clontech Laboratories
Stephane	Boutet	Fluidigm Corporation
Sean	Boyle	Personalis
Laurakay	Bruhn	Agilent Technologies
Condie	Carmack	Vela Diagnostics
Ken	Chang	Merck, Sharp & Dohme
Patty	Chiang	AcroMetrix
Tyson	Clark	Pacific Biosciences, Inc.
Michael	Clark	Personalis
John	Coller	Stanford School of Medicine
Amanda	Courtright	TGen
Andrew	Crenshaw	Expression Analysis/Quintiles
Maureen	Cronin	Celgene Corp
Bo	Curry	Agilent Technologies
Joy	Dunkers	BBD
Camila	Egidio	Fluidigm
Richard	Fekete	Fluidigm Corporation
Joshua	Fenrich	Bio-Rad Laboratories
Jeff	Goldy	Allen Institute for Brain Science
Kun	Guo	Bio-Rad Laboratories
Jo Lynne	Harenza	NIST
Ross	Haynes	NIST
Ayal	Hendel	Stanford University
Shawn	Hodges	Bio-Rad: DBC
Kendall	Jensen	TGen
Tomer	Kalisky	Dept of Bioengineering, Bar-Ilan University, Israel
Irina	Khrebtukova	Illumina
Jason	Kralj	NIST
Kathy	Lee	Thermo Fisher Scientific
Sara	Lefort	SPRC
Sharon	Liang	FDA
Shin	Lin	Stanford University
Leonard	Lipovich	Wayne State University
Mei-Lan	Liu	Genomic Health, Inc.
Peng	Liu	University of Wisconsin-Madison
Lewis	Liu	USDA-ARS
Mary	Ma	Bio-Rad

Name		Institution
Ivana	Malenica	TGen
Gary	Mantalas	Stanford University
John	Marino	NIST Materials Measurement Laboratory
Matthew	Marton	Merck, Sharp & Dohme
Christopher	Mason	Weill Cornell Medical College
Adam	McCoy	Bio-Rad Laboratories
Jennifer	McDaniel	NIST
Erin	Mitsunaga	Stanford School of Medicine
Sarah	Munro	NIST/ABMS
Matthew	Munson	NIST/ABMS
Timothy	Myers	NIH
Joel	Myerson	Agilent Laboratories
Norma	Neff	Stanford Bioengineering
Nancy	Ngo	Bio-Rad Laboratories
Kara	Norman	Thermo Fisher Scientific
Steven	Okino	Bio-Rad Laboratories
Jerod	Parsons	NIST
Ben	Passarelli	Stanford
Lukas	Paul	Lexogen GmbH
Mylan	Pho	Genomic Health
P Scott	Pine	NIST/ABMS
Stephanie	Pond	Illumina Inc.
Ronald	Przygodzki	US Dept. Veterans Affairs
Rongsu	Qi	Thermo Fisher Scientific
Deyra	Rodriguez	New England Biolabs
Marc	Salit	NIST/ABMS
Mary	Satterfield	NIST
Uwe	Scherf	FDA/CDRH/OIR/DMD
Aaron	Schetter	FDA
Alexander	Seitz	Lexogen GmbH
Bob	Setterquist	Thermo Fisher Scientific
Cricket	Sloan	Stanford University
Kimberly	Smith	Allen Institute
Michael	Snyder	Stanford University
Paul	Spencer	UC Berkeley
Qin	Su	NIH
Yongming	Sun	Thermo Fisher Scientific, Inc
Andrea	Szakal	NIST Materials Measurement Laboratory
John	Tan	Roche NimbleGen
Waibhav	Tembe	TGen
Zivana	Tezak	FDA
Karol	Thompson	US FDA
Hagen	Tilgner	Stanford University
Elizabeth	Tseng	Pacific Bio

Name		Institution
Lindsay	Vang	NIST
Kamini	Varma	Thermo Fisher Scientific, Inc
Sumathi	Venkatapathy	Affymetrix
Yan	Wang	Bio-Rad Laboratories
Janet	Warrington	Second Genome
Patrick	Weiss	Twist Bioscience
Sherman	Weissman	Yale U School of Medicine
James	Willey	University of Toledo
Angela	Wu	Stanford University
Qi	Zhang	University of Wisconsin-Madison

Appendix II: Participant Presentation Summaries

Ambion ERCC Spike-in Mix
Bob Setterquist (Thermo Fisher Scientific)
This presentation provided an overview of the current Ambion ERCC spike-in mix (derived from NIST SRM 2374), including some lessons learned, and a discussion of customer feedback for ERCC 2.0, such as requests for a simpler formulation with a narrower dynamic range and controls for allele-detection.
Presentation is not available on Slideshare.net

Spike-in transcripts for the quantification of mRNA isoforms in Next Generation Sequencing
Lukas Paul (Lexogen GmbH)
To address the problem of accurate mRNA isoform identification and quantification in RNA-Seq workflows Lexogen GmbH have designed, produced and tested a set of spike-in RNAs representing transcript isoforms. This presentation provided the scope and rationale of the design, and results of these external standards in RNA-Seq data evaluation.

ERCC Transcripts Give Confidence to the Performance of One-Color and Two-Color Microarray Experiments
Anne Bergstrom Lucas (Agilent Technologies)
This presentation described Agilent's experience in determining ERCC RNA spike-in control input levels and assessing ERCC control and microarray assay performance with the ERCC Phase IV spike-in controls.

Defined mixes of synthetic microRNAs: prototypes for controls for microRNA profiling assays
Karol Thompson (U.S. Food and Drug Administration)
Defined mixes of 32 microRNAs in a Latin Square design were tested for potential use as an internal calibration curve. The initial evaluation involved two RT-qPCR platforms and one microarray platform.

Epitranscriptome Standards
Christopher Mason (Weill Cornell Medical College)
Dynamic modifications of RNA have become a key mediator in RNA function, even though they do not change the information content (epitranscriptome). In this presentation several proposed approaches to standardize this nascent field are described. A working group within the ABRF (Association of Biomolecular Resource Facilities) has been established to address this, and others are invited to test the various RNA base modifications.

Use of ERCC synthetic spike in standards to validate competitive multiplex PCR amplicon library method for targeted NGS analysis
Jim Willey (University of Toledo)
A targeted competitive multiplex PCR amplicon library method was developed and applied to analysis of SEQC reference materials containing the ERCC synthetic spike-in controls. This method produced data that were highly reproducible with Taqman PCR, and through convergence of targets during PCR, enabled marked reduction in sequencing costs. The conclusion of this study was that this NGS method is suitable for diagnostic testing of up to hundreds of clinically actionable DNA or RNA targets.

Single Molecule, Real-Time Sequencing of Full-length cDNA Transcripts
Tyson Clark (Pacific Biosciences)
Improved PacBio Iso-Seq library preparation techniques enable sequencing of full-length cDNA molecules up to 10 kb. Size selection technologies and methods for targeted full-length cDNA sequencing were discussed. Proposed design criteria for ERCC 2.0 controls were presented including long transcripts (>10 kb), transcript isoforms that span size bins, complex alternative splicing patterns, and GC content diversity.

Read count information and technical variability in PacBio full-length cDNA sequencing
Elizabeth Tseng (Pacific Biosciences)
The PacBio Iso-Seq method generates high-quality, full-length transcripts up to 10 kb. Analysis was presented on how read count information -- number of reads associated with each distinct full-length isoform -- varies across technical replicates and what factors influence reproducibility across technical replicates. In addition, ideas were presented on the possibility of using normalized read count information for identifying differentially expressed isoforms in long read data.

Synthetic Spike-in mRNA-Seq data for Cancer Gene Fusion Detection
Stephanie Pond (Illumina)
Cancer fusion transcripts developed by Illumina and TGen were tested in a collaborative study. Nine medically-actionable oncogenic fusion targets were evaluated in RNA-Seq experiments and the data was analyzed with three different fusion detection tools and compared to a control method of an alignment to a human reference including the nine targets.

Standard Performance Metrics for Gene Expression Experiments with the "erccdashboard"
Sarah Munro (National Institute of Standards and Technology)
A new open source software tool, the "erccdashboard" R package, is available to assess technical performance of differential gene expression experiments using ERCC control ratio mixtures. This method validation tool provides a standard set of technology-independent performance metrics to evaluate different RNA measurement technologies (e.g. microarray and RNA-Seq) and reproducibility of a gene expression experiments across space and time.

Efficient Chemical Syntheses of Long and Modified RNA Oligonucleotides
Joel Myerson (Agilent Technologies)
Agilent Technologies and collaborators developed a process for direct chemical synthesis of RNA (J. Am. Chem. Soc. 2011, 133, 11540-11556). Using this TC (Thionocarbamate) protecting group chemistry RNA can be synthesized with modified natural or unnatural nucleotides in large well-defined amounts with consistent purity as demonstrated with HPLC-ESI Mass Spectrometry.

Appendix III: Comments from Working Group Scope Discussion

- Ideal to have different types of controls, such as:
    - Transcript isoforms.
    - miRNA,
    - Cancer gene fusions
    - Pseudogenes
    - Epigenetic modifications
- Modularity is important. As with ERCC 1.0 in the NIST SRM 2374 sequence library, ideally every control species developed by ERCC 2.0 will be available as individual components within a set, so that there is complete flexibility mixing control species for different purposes.
- Timeline is important. Start moving forward with existing controls, do not wait for all types of RNA controls to be designed.
- Allele-specific expression is an important application.
- Consider development of controls with and without poly-A tails.
- Controls could be design for absolute calibration, to track efficiencies
- A method for FFPE analysis is an example of a reference method need
- Control mixtures are needed for different applications, e.g. 10 transcripts for qRT-PCR.
- How do applications shape the needs for controls?
    - Will controls be used in every experiment or once per month to validate workflows?
    - What dynamic range is needed for different applications?
- For transcript isoform design, the set of non-cognate ('alien') splice variants developed by Lexogen GmbH are available to the consortium for development as ERCC 2.0 transcript isoform controls.
- Biologically-derived (cognate) transcript isoform controls from a species such as *Schizosaccharomyces pombe* were proposed, which would be selected to have no homology with human and other mammalian transcripts.
- ERCC 2.0 could develop both types of transcript isoforms as standard transcript isoform RNA controls.
- How many isoforms are needed for controls?
    - All of them, <1% or 0.1% detection
    - Don't focus on creating mixtures with lowly expressed isoforms, but make 1 or 2
- There was discussion around RT efficiency
- Make sure that any new ERCC 2.0 products are orthogonal to (don't conflict with) existing ERCC 1.0 control mixtures.
- We can develop different reference annotations ('faux' sequence) and provide 'challenge' data
- 'Analysis' working group considerations:
    - Choose standard performance measures early
    - Enrichment assays and understanding base modification