The External RNA Controls Consortium: a progress report

The External RNA Controls Consortium*

Standard controls and best practice guidelines advance acceptance of data from research, preclinical and clinical laboratories by providing a means for evaluating data quality. The External RNA Controls Consortium (ERCC) is developing commonly agreed-upon and tested controls for use in expression assays, a true industry-wide standard control.

External controls have been used to verify technical performance and interpret quality of data from microarray and quantitative real-time reverse transcriptase polymerase chain reaction (QRT-PCR) experiments since the techniques were first reported in the scientific literature[1–7]. The external (spike-in) controls used by most laboratories are custom controls developed and optimized for specific platforms and assays, and are therefore of limited utility as a well-characterized reference or traceable control material of known performance. In May 2003, researchers formed the ERCC after participating in a meeting held at Stanford University—Metrology and Standards Needs for Gene Expression Technologies: Universal RNA Standards, sponsored by the US National Institute of Standards and Technology (NIST), Genomic Health Inc., Agilent Inc. and Affymetrix Inc.[8]. Acknowledging the need for controls for many aspects of the expression assay process, there was consensus that it would be of significant scientific and practical value to first collaborate on a reference set for evaluating technical performance. The ERCC is developing external RNA controls useful for evaluating technical performance in gene expression assays performed by microarray or QRT-PCR analysis, and of particular utility to research and clinical laboratories, regulatory agencies, accrediting agencies, reference laboratories, test, microarray and reagent manufacturers (Fig. 1). Today scientists representing over 50 international pharmaceutical, diagnostic, biotechnology, academic, clinical and government organizations are participating with the goal of delivering: (i) a certified reference material (CRM) consisting of a reference set of approximately 100 well-characterized clones comprising RNA transcripts from random unique sequences as determined by sequence comparison to mouse, rat, human, drosophila, bacteria,and mosquito sequence databases, as well as inclusion of other nonhuman sequences; (ii) access to clones of the reference set deposited at a public repository; (iii) publication of all sequence information and test data;. (iv) protocols for the preparation and use of the controls; (v) suitable algorithms and bioinformatics tools for quantitative assessment and evaluation (Box 1). The consortium is open to all persons interested in participating and contributing to the effort (Box 2).

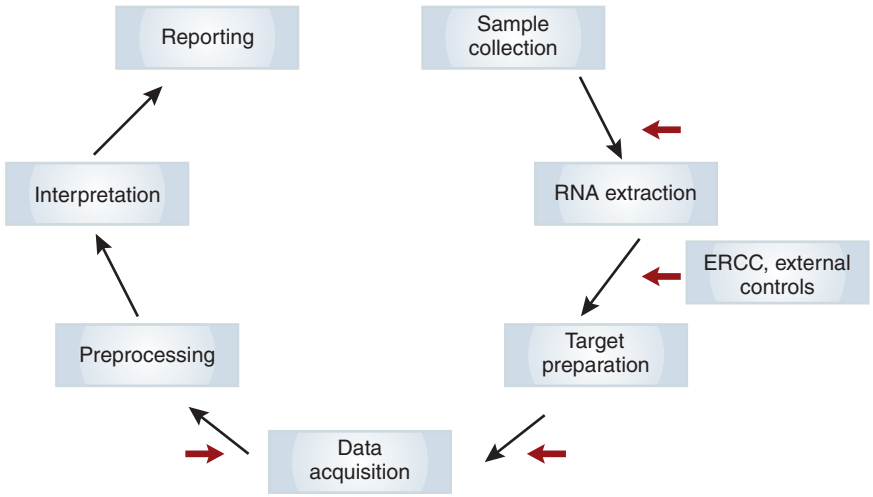

Figure 1 | The expression assay process. Sources of variability can be measured at various steps in the expression assay process indicated by the red arrows. Present work of the ERCC is addressing technical performance through the creation of control material introduced to total RNA before the labeling step.

Specification document

A specification document, External RNA Controls: Proposed Design and Input Requirements, was created in 2003 by the ERCC, reviewed and modified in an open forum held at NIST in December 2003. The document is available online (http://www.

*Authors and their affiliations are listed at the end of the paper. Correspondence should be addressed to Janet A. Warrington (janet_warrington@affymetrix.com).

NIST.gov). The purpose of the document is to provide an initial guide to control design, development and testing of the external RNA controls as well as a means of communicating the scope and goals of the ERCC. The preparation of this document involved five months of discussion and consensus building. Discussions centered on scope, access, utility, controls description, characterization, potential configurations, shelf life, storage, protocols, testing, definitions, assumptions for measuring acceptable performance and quality control. Over the past 18 months, the product concept has been refined to include; a reference set of ~100 well-characterized, tested clones, published sequence data, test performance data, acceptance criteria, protocols developed via an accredited process for use in research and clinical laboratories, analytical and bioinformatics tools and protocols.

The original specification document called for production of individual transcripts as well as a variety of pools of different quantities and configurations. Although pools will be evaluated during testing it is not planned to prepare defined pools as the reference material produced by the ERCC. It is anticipated that market opportunity and innovation will encourage the development of CRMs configured as pools by reagent kit manufacturers using the reference set created by the ERCC. It is anticipated that once produced and available the CRMs will be used in many ways by the scientific community to further expand and accelerate expression technology. Applications discussed include encapsulated controls for measuring nucleic acid extraction efficiency, optimization of labeling methods, optimization of normalization methods and improved ability to measure sources of quantification bias.

Sequence selection

Several organizations participating in the ERCC have contributed sequences to be tested (Table 1). Sequences are accepted with the condition that the ERCC can use them unencumbered by licensing fees or intellectual property issues. Many of these sequences have been previously used on one or more platforms. Some of the controls are well-characterized non-human sequences, for example, *Bacillus subtilis*. Others have been created as randomly generated unique sequences. All sequence information will be published upon completion of testing. Fundamental questions of sequence uniqueness, secondary structure and sequence length are being addressed. Sequence collection and analysis of secondary structure in randomly generated sequences have been completed.

BOX 1 ORGANIZATIONS PARTICIPATING IN THE ERCC

Affymetrix Inc., Santa Clara, California, USA
Agilent Technologies, Inc., Santa Clara, California, USA
Ambion, Inc., Austin, Texas, USA
Applied Biosystems, Foster City, California, USA
American Type Culture Collection, Manassas, Virginia, USA
Atactic Technology, Inc., Houston, USA
Biomerieux, Marcy l'Etoile, France
Bristol-Myers Squibb, New York, USA
Cambridge University, Cambridge, UK
CapitalBio Corporation, Beijing, China
Celera Diagnostics, Alameda, California, USA
Cenetron Diagnostics Ltd., Austin, Texas, USA
Clinical & Laboratory Standards Institute
Clinical Hospital Center Zagreb, Zagreb, Croatia
Combimatrix Corporation, Mukilteo, Washington, USA
Eli Lilly & Company, Indianapolis, USA
Enzo Life Sciences, Farmingdale, New York, USA
Eppendorf Array Division, Namur, Belgium
Expression Analysis, Durham, North Carolina, USA
GE Healthcare, Chandler, Arizona, USA
Genentech, South San Francisco, USA
Genetics Society of Vietnam, Ho Chi Minh City, Vietnam
Illumina, Inc., San Diego, USA
Informax, Inc., Bethesda, Maryland, USA
International Federation of Clinical Chemistry & Laboratory Medicine, Milano, Italy
Invitrogen Corporation, Carlsbad, California, USA
Johns Hopkins School of Public Health, Baltimore, USA
Lawrence Livermore Laboratory, Livermore, California, USA LGC, Teddington, UK
Maine Molecular Quality Controls, Scarborough, Maine, USA
Mayo Clinic, Rochester, Minnesota, USA
Merck & Co., West Point, Pennsylvania, USA
National Institute of Standards & Technology, Gaithersburg, Maryland, USA
Northwestern University, Evanston, Illinois, USA
NuGEN Technologies, Inc., San Carlos, California, USA
Qiagen, Hilden, Germany
Queens University Hospital, Ontario, Canada
Roche Molecular Systems, Pleasanton, California, USA
Stanford University, Stanford, California, USA
Stratagene Corporation, La Jolla, California, USA
Tokyo University, Tokyo, Japan
University of California Los Angeles, Los Angeles, USA
University Health Network, Toronto, Canada
US Centers for Disease Control, Atlanta, USA
US Centers for Medicare & Medicaid Services, Baltimore, USA
US Department of Agriculture, Peoria, Illinois, USA
US Food and Drug Administration, Center for Biologics Evaluation and Research, Rockville, MD
US Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring and Rockville, Maryland, USA US Food and Drug Administration, Center for Drug Evaluation Research, Rockville, Maryland, USA US Food and Drug Administration, National Center for Toxicological Research, Jefferson, Arizona, USA US Food and Drug Administration, Office of In Vitro Diagnostic Device Evaluation and Safety, Rockville, Maryland, USA US National Institutes of Health, National Cancer Institute, Bethesda, Maryland, USA
Veridex, Johnson & Johnson, Warren, New Jersey, USA
ViaLogy Corporation, Altadena, California, USA
VigeneTech, Inc., North Billerica Massachusetts, USA

BOX 2 ERCC WAYS OF WORKING

The ERCC
- is a volunteer organization
- is open to anyone with an interest in working together to achieve our goals
- makes decisions by consensus based on data, when needed, members will vote on issues affecting the group and influencing work in progress
- intends to produce controls, protocols and analysis tools accessible to the entire community
- will publish final results as a group, by the group

Clone construction

To determine if a particular poly(A) tail length might bias labeling efficiency for any one of the labeling methods being used by the ERCC, studies were carried out by participating laboratories to evaluate poly(A) tail lengths and an acceptable length was identified, 20-mer (data available on request). Data supporting selection of the 20-mer poly(A) tail length has been reported at ERCC open meetings and at the Johns Hopkins University/FDA/PhRMA meeting held July 20, 2005 in Rockville, Maryland, USA. All control clones will be prepared in a public-domain vector suitable for *in vitro* transcription. The ERCC will publish a comprehensive summary of the development work and deposit relevant data into one or more stabile, publicly available data repositories upon completion of the project: for example, GEO at NCBI in the United States, ArrayExpress at EBI in Europe, CIBEX at NIG in Japan.

Protocol development

In conjunction with the activities of the ERCC, a guideline has been developed for use of the controls in research and clinical laboratories via the accredited process of the Clinical Laboratory Standards Institute (CLSI, formerly NCCLS). The document, MM16-P, Use of External RNA Controls in Gene Expression Assays; Proposed Guideline, includes chapters on gene expression profiling in clinical applications, use of control materials to assess technical and clinical performance, limitations of external controls, technical performance metrics, protocols for preparation and assessment of external RNA controls and protocols for using external RNA controls for microarrays and for QRT-PCR, and is planned for publication in January 2006.

Testing and establishment of performance criteria

A proposed plan for testing the controls has been posted online (http://www.NIST.gov) for public review and comment before an open meeting scheduled for October 4 and 5, 2005 at the Lister Hill Auditorium in Bethesda, Maryland, USA. Information about the meeting is available online (http://www.fda.gov/cder/genomics/default.htm and http://www.NIST.gov). The public is invited to attend and participate in the meeting. For those unable to attend, the ERCC is soliciting feedback via e-mail. Generally the test plan consists of testing the controls across multiple platforms at several member sites including microarray and reagent manufacturers, contract research laboratories and government laboratories including, the US National Cancer Institute, the US Department of Agriculture, the US Food and Drug Administration and NIST. All sites will run agreed-upon protocols and analytical methods. The type of information collected during testing includes: crosshybridization, quantitative performance titration experiments, performance in complex samples and independent confirmation of performance.

Evaluation of performance

After the initial testing, acceptance criteria will be established, for example, an acceptable range of deviation within the quantitative information obtained from known input concentrations. These criteria will be applied to the second round of testing and the controls selected for the final reference set will be those that meet the performance requirement across platforms and are independent of assay type. Initial testing will focus on a small number of complex backgrounds including human, rat and mouse. More extensive testing is planned as a follow-up activity as described below.

A related and complementary project, the MicroArray Quality Control (MAQC) project, led by US FDA's National Center for Toxicological Research, is now underway characterizing two high-complexity samples across microarray platforms (http://edkb.fda.gov/MAQC/). The ERCC will use the complex samples produced as part of the MAQC project for testing controls in complex samples, and information regarding sources of variability will be useful. The ERCC is identifying and establishing metrics for the technical performance of control transcripts that perform consistently. Methods of data preprocessing and normalization will be carried out in a consistent manner across all platforms using predetermined algorithms as specified in the test plan. Tools, methods and relevant information will be made available in an open source manner, for example, R code for a bioconductor package and sequence links to public annotation databases. The analytical work is a topic for discussion at the Test Plan Workshop. After testing and preliminary analysis, the ERCC will hold an analysis jamboree to provide participants the opportunity to compare assumptions,

Table 1 | Sources of the 140 control sequences that are being tested

Number	Affiliation	Source	Length (bp)
1–28	Affymetrix	*B. subtlis*	700–2,000
29–40	Affymetrix	Artificial sequences	500–1,900
41–43	USDA-ARS-NCAUR	*Bos taurus*	500
44–46	USDA-ARS-NCAUR	*Glycine max*	500
47	Ambion	Lamda phage	2,002
48–53	Ambion	Artificial sequences	1,000
54–61	Ambion	*E. coli*	750–2,000
62–82	Stanford University	*Methanococcus*	500–750
83–85	Agilent Technologies	Artificial sequences	500
86–90	GE HealthCare	*E. coli*	1,000
91–140	Ambion/Atactic	Artificial sequences	1,000

address open questions and confirm that analysis is being carried out in a manner consistent with the objectives of the ERCC. After the final analysis, manuscript(s) will be prepared and submitted for publication, open-source informatics tools and data will be made publicly available.

Approval and acceptance

A priority for the ERCC membership was creating a process for establishment of commonly agreed-upon controls without discouraging ongoing innovation in the expression community. Working closely with NIST, a process is being developed that will make available a laboratory-ready ERCC reference set as CRMs while allowing others to develop new control materials and demonstrate equivalent performance if they choose not to use the ERCC reference set. Once the reference set is on deposit in a public repository, the ERCC will encourage additional sites to perform testing in a variety of sample types. Ultimately the ERCC plans to host a controls symposium to provide a forum for presentation of those results.

Milestones past, present and future

The ERCC has achieved key milestones in the past two years including voluntary participation of key stakeholders, public and private, building widespread international participation with more than 100 scientists from organizations in 11 countries, agreement on common goals, scope and work plan, consensus on reference set specifications, approval of CLSI MM16-P, Use of External RNA Controls: Proposed Guideline, and development of a Controls Test Plan to be implemented at the end of 2005, a path forward for the development of the CRM. There is tremendous momentum for advancing this work, and there is much work to be done. The Test Plan has been publicly posted but not yet accepted. Clone construction is underway, with work and reagents contributed and/or paid for by government agencies and private industry. Microarrays for control testing are being designed, manufactured and donated by multiple array manufacturers as well as academic microarray laboratories. Reagent manufacturers are producing and donating materials required for the QRT-PCR work. The testing itself is planned to begin in the fourth quarter of 2005 with analysis work continuing into the first part of 2006. ERCC members are working together, contributing expertise to complete the testing. Development of tools required for analytical work to aid in evaluating performance is underway and will continue into 2006. Consensus on the criteria to be used to define the reference set and subsequent identification of that set is planned for the latter half of 2006. The ERCC welcomes the participation of anyone in the community with an interest in helping to drive these efforts forward.

1. Fodor, S.P.A. *et al. Science* **251**, 767–773 (1991).
2. Fodor, S.P.A. *et al. Nature* **364**, 555–556 (1993).
3. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. *Science* **270**, 467–470 (1995).
4. Lockhart, D.J. *et al. Nat. Biotechnol.* **14**, 1675–1680 (1996).
5. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. *Biotechnology* **11**, 1026–1030 (1993).
6. Heid, C.A. *et al. Genome Res.* **6**, 986–984 (1996).
7. Wittwer, C.T. *et al. Biotechniques* **22**, 130–138 (1997).
8. Cronin, M. *et al. Clin. Chem.* **50**, 1464–1471 (2004).

The authors are: Shawn C Baker[1], Steven R Bauer[2], Richard P Beyer[3], James D Brenton[4], Bud Bromley[5], John Burrill[6], Helen Causton[7], Michael P Conley[8], Rosalie Elespuru[9], Michael Fero[10], Carole Foy[11], James Fuscoe[12], Xiaolian Gao[13], David Lee Gerhold[14], Patrick Gilles[15], Federico Goodsaid[16], Xu Guo[17], Joe Hackett[18], Richard D Hockett[19], Pranvera Ikonomi[20], Rafael A Irizarry[21], Ernest S Kawasaki[22], Tamma Kaysser-Kranich[23], Kathleen Kerr[3], Gretchen Kiser[23], Walter H Koch[24], Kathy Y Lee[24], Chunmei Liu[17], Z Lewis Liu[25], Anne Lucas[26], Chitra F Manohar[27], Garry Miyada[17], Zora Modrusan[28], Helen Parkes[11], Raj K Puri[29], Laura Reid[30], Thomas B Ryder[17], Marc Salit[31], Raymond R Samaha[6], Uwe Scherf[32], Timothy J Sendera[6], Robert A Setterquist[33], Leming Shi[12], Richard Shippy[23], Jesus V Soriano[34], Elizabeth A Wagar[35], Janet A Warrington[17], Mickey Williams[24], Frederike Wilmer[36], Mike Wilson[33], Paul K Wolber[26], Xiaoning Wu[24] & Renata Zadro[37]

[1]Illumina, Inc., 9885 Towne Centre Drive, San Diego, California 92121, USA. [2]US Food and Drug Administration, Center for Biologics Evaluation and Research, Building 29B/2NN12, HFM-521, 1401 Rockville Pike, Rockville, Maryland 20852, USA. [3]University of Washington, 4225 Roosevelt Way NE, Seattle, Washington 98105, USA. [4]Cambridge University, Cambridge Institute for Medical Research (CIMR), Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK. [5]ViaLogy Corporation, 2400 Lincoln Avenue, Altadena, California 91001, USA. [6]Applied Biosystems, 850 Lincoln Center Drive, Foster City, California 94404, USA. [7]University of London, Imperial College of London, London SW7 2AZ, UK. [8]Enzo Life Sciences, Inc., 60 Executive Boulevard, Farmingdale, New York 11735, USA. [9]US Food and Drug Administration, Center for Devices and Radiological Health, Federal Laboratories at White Oak, Life Sciences Building 64, Room 3020, 10903 New Hampshire Avenue, Silver Spring, Maryland 20903, USA. [10]Stanford University School of Medicine, CCSR 4256, 269 Campus Drive, Stanford, California 94305, USA. [11]LGC, Bio-Molecular Innovation, Queens Road, Teddington TW11OLY, UK. [12]US Food and Drug Administration, National Center for Toxicological Research, 3900 NCTR Road, HFT-020, Jefferson, Arkansas 72079, USA. [13]Atactic Technology, Inc., 2575 W. Bellfort, Suite 270, Houston, Texas 77054, USA. [14]Merck Research Laboratories, Department of Molecular Profiling, West Point, Pennsylvania 19486, USA. [15]Invitrogen Corporation, 1600 Faraday Avenue, P.O. Box 6482, Carlsbad, California 92008, USA. [16]US Food and Drug Administration, Center for Drug Evaluation and Research, 1451 Rockville Pike (HFD-860), Woodmont II Office 2078, Rockville, Maryland 20852, USA. [17]Affymetrix Inc., 3380 Central Expressway, Santa Clara, California 95051, USA. [18]US Food and Drug Administration, Center for Devices and Radiological Health, HFZ-440, 2098 Oak Grove, Rockville, Maryland 20850, USA. [19]Eli Lilly and Company, Lilly Research Lab, Lilly Corporate Center, Indianapolis, Indiana 46285, USA. [20]American Type Culture Collection, 10801 University Boulevard, Manassas, Virginia 20110, USA. [21]Johns Hopkins School of Public Health, 615 N. Wolfe St. E3132, Baltimore, Maryland 21205, USA. [22]National Cancer Institute Advanced Technology Center, National Institutes of Health, 8717 Grovemont Circle, Bethesda, Maryland 20892, USA. [23]GE Healthcare, 3200 W. Germann Road, Chandler, Arizona 85248, USA. [24]Roche Molecular Systems, 4300 Hacienda Dr., Pleasanton, California 94588, USA. [25]US Department of Ariculture, 1815 N. University Street, Peoria, Illinois 61604, USA. [26]Agilent Technologies, Inc., 5301 Stevens Creek Blvd. MS 55L WM, Santa Clara, California 95051, USA. [27]Lawrence Livermore National Laboratory Microarray Center, 7000 East Avenue, Livermore, California 94550, USA. [28]Genentech, 1 DNA Way, South San Francisco, California 94080, USA. [29]US Food and Drug Administration, NIH Building 29B, Room 2NN10, 29 Lincoln Drive MSC 4555, Bethesda, Maryland 20892, USA. [30]Expression Analysis, 2605 Meridian Parkway, Durham, North Carolina 27713, USA. [31]National Institute of Standards and Technology, Chemical Science and Technology Laboratory, 100 Bureau Drive, Stop 8391, Gaithersburg, Maryland 20899, USA. [32]US Food and Drug Administration, Office of In Vitro Diagnostic Device Evaluation and Safety, OAK8, RM366, HFZ-440, 2098 Gaither Road, Rockville, Maryland 20850, USA. [33]Ambion, Inc., 2130 Woodward St., Austin, Texas 78744, USA. [34]American Type Culture Collection, 10801 University Boulevard, Manassas, Virgnia 20110, USA. [35]UCLA Clinical Laboratories, 10833 Le Conte Avenue, Los Angeles, California 90095, USA. [36]Qiagen Inc., 40727 Hilden, Germany. [37]Zagreb University School of Medicine, Medical School, 10 000 Zagreb, Salata 3 Croatia.