National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2374

DNA Sequence Library for External RNA Controls

This Standard Reference Material (SRM) is intended for use as a template for ribonucleic acid (RNA) control synthesis using in vitro transcription (IVT). These RNA controls are designed to be used as external, or "spike-in", controls to support confidence in gene expression assays by providing quantitative assessment of the technical performance of a gene expression measurement. A unit of the SRM contains 96 different 0.5 mL polypropylene tubes, with approximately 10 µg, as measured by absorbance (A_{260}), of dehydrated plasmid deoxyribonucleic acid (DNA) in each. Each tube contains plasmid DNA with a unique template sequence for a different external RNA control. These controls were developed in cooperation with the External RNA Controls Consortium (ERCC).

These templates can be readily used to make RNA controls (see Figure 1). Depending on the strand transcribed, the controls will mimic either "sense" or "anti-sense" eukaryotic messenger RNA (mRNA). When used to make sense control RNA, each control will have a nominal 24 nucleotide (nt) polyadenylated (polyA) segment at the 3′ end. The DNA templates and RNA transcripts produced from them are annotated and described in Figures 2 and 3. The control sequences range from 273 nt to 2022 nt, with two population distributions of GC fraction[(1)], one centered at 35 % GC and another at 47 %.

Certified Properties: The certified properties of SRM 2374 are the DNA sequences of the ERCC control inserts. The nucleotide identities are encoded using International Union of Pure and Applied Chemistry (IUPAC) nucleotide symbols, including the ambiguity codes for several bases in the sequences. Table 2 includes, for each component, a reference to a record containing the certified sequences in the National Center for Biological Information (NCBI) GenBank database. The certified sequences are available from NIST in a FASTA-formatted data file [1], associated with this certificate.

Since the measurand is the value of a nominal property (a sequence of nucleotides), a conventional evaluation and expression of measurement uncertainty conforming to the JCGM GUM [2] cannot be used. Instead, the uncertainty associated with each nucleotide is expressed in an ordinal scale that represents the strength of the belief in the assigned value (0 = Most Confident, 1 = Very Confident, 2 = Confident, 3 = Ambiguous). Characteristics of sequence data associated with the levels of the ordinal scale are described in Table 1 (see "Certified Properties Confidence Estimates"). These confidence estimates are available in a companion FASTA-formatted file [1].

In the absence of a fully developed metrology for identity (the current state of affairs), a pragmatic way forward is to consider these DNA sequences as the source of "comparability of identity" for RNA controls transcribed from the library.

Expiration of Certification: The certification of **SRM 2374** is valid, within the specified confidence levels, until **01 October 2027**, provided the SRM is handled and stored in accordance with the instructions given in this certificate (see "Instructions for Handling, Storage, and Use"). The certification is nullified if the SRM is damaged, contaminated, or otherwise modified.

Coordination of the technical measurements and analysis leading to the certification was under the direction of M. Salit of the NIST Material Measurement Laboratory Office.

Michael J. Fasolka, Acting Chief
Material Measurement Laboratory Office

Gaithersburg, MD 20899
Certificate Issue Date: 06 December 2017
Certificate Revision History on Last Page.

Steven J. Choquette, Director
Office of Reference Materials

(1) GC fraction is the ratio of guanine + cytosine (GC) to adenine + thymine (AT).

Information Properties: Table 2 contains various useful characteristics of the template sequences. Additional companion data files contain supporting data to enable simple, accurate use of the material and primary data when performing sequence bioinformatics [1]. The secondary data include a simplified representation of the sequence data (with IUPAC ambiguity codes resolved using a best likelihood estimate of the nucleotide); a Gene Transfer File (GTF) describing the structures of the RNA controls as if they were genes; a sequence reference file and GTF with the polyA tail excluded; and a sequence reference file of the plasmid vector as described in reference 1.

Maintenance of SRM Certification: NIST will monitor this SRM over the period of its certification. If substantive technical changes occur that affect the certification before the expiration of this certificate, NIST will notify the purchaser. Registration (see attached sheet or register online) will facilitate notification.

J. McDaniel of the NIST Material Measurement Laboratory Office led the certification and supporting study measurement processes and oversaw materials preparation and packaging. M. Roesslein of Empa in Switzerland and J. McDaniel of NIST analyzed the data and established the confidence estimates. J. Zook, S. Munro, P.S. Pine, M. Munson, and J. Kralj of the NIST Material Measurement Laboratory Office and M. Kline of the NIST Biomolecular Measurement Division contributed to the measurements and analysis. A. Young and R. Blakeslee of the National Institute of Health (NIH) Intramural Sequencing Center and K. Becker of the National Institute of Aging hosted NIST staff to perform measurements at their laboratories. The many members of the External RNA Control Consortium (ERCC) from government, academic, and private laboratories contributed intellectual content, measurements, and analysis to the development of this SRM [3][(2)].

Statistical consultation for this SRM was provided by A. Possolo and W. Liggett of the NIST Statistical Engineering Division.

Support aspects involved in the issuance of this SRM were coordinated through the NIST Office of Reference Materials.

Certified Properties Confidence Estimates: A set of heuristic, experience-based, rules (see Table 1) were used to establish confidence estimates for the DNA sequence in the ERCC control inserts. These rules were independently applied by two different analysts who hand-curated the replicate two-stranded Sanger sequence data, collated the data and then further curated with alternative sequencing platforms.

Table 1. Definitions of Heuristic Rules

Confidence Level	*Heuristic Definition*
Most Confident	Have good answers (fully reliable, unambiguous base calls) on both strands; all data from multiple reads of both strands agree.
Very Confident	Have good answer on one strand; poor answer (less than fully reliable, potentially ambiguous base call) on the second/alternate strand; base calls from both strands typically agree, and there is biochemical context that explains the anomalous sequence data.
Confident	Have good answer on one strand; anomalous sequence data that may give rise to a conflicting base call on the alternate strand; judgment required to resolve anomaly.
Ambiguous	No clear mutually supporting results; unambiguous base calls disagree; or — no unambiguous base calls on either strand; data from the two opposing strands could not be authoritatively reconciled.

INSTRUCTIONS FOR HANDLING, STORAGE, AND USE

Handling: SRM 2374 is a BACTERIAL SOURCE MATERIAL derived from a well-characterized strain of *Escherichia coli* (*E. coli*), not known to consistently cause disease in immunocompetent adult humans, and presents minimal potential hazard to laboratory personnel and the environment. Handle the components as Biosafety Level 1 [4]. SRM 2374 components and derived solutions should be disposed of in accordance with local, state, and federal regulations.

Storage: SRM 2374 will arrive on gel freezer packs to maintain sub-ambient conditions. Upon receipt, SRM 2374 should be kept in the dark at –20 °C for long-term storage, or in the dark at 4 °C for short-term storage (if use is imminent). Following rehydration, it is recommended that the plasmid DNA components be used in their entirety.

[(2)] Certain commercial equipment, instrumentation, or materials are identified in this certificate to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

Use: Figure 1 is a flow chart of the use process. Table 2 annotates the appropriate restriction enzymes to use to linearize each control. Figure 2 diagrams the five different classes of restriction enzyme pairs used for different components of the SRM and the base counts of flanking sequence, which depend on the DNA template used. Figure 3 diagrams the IVT process for the various DNA templates. The process selected will include different configurations (and lengths) of flanking sequence, as detailed in Figure 2. Nominal RNA control transcript lengths can be calculated using the "Insert Length" from Table 2, added to the flanking sequence lengths derived from Figures 2 and 3.

- Rehydration is confirmed by determining the concentration of the DNA prior to use in IVT. If rehydration cannot be confirmed, ensure that the inner walls of the vial have been wetted with rehydration solution and repeat vortex, centrifuge, and concentration determination. A small number of empty vials have been observed. If presence of DNA cannot be confirmed and it is determined the vial is empty, please email srminfo@nist.gov to report the issue and receive a replacement.

- Linearization of the plasmid prior to IVT is the recommended method to ensure that the RNA polymerase ceases transcription at the desired sequence location, yielding controls of proper length that contain the control sequence with minimal excess flanking sequence.

RNA Pooling: Individual RNA controls are typically pooled to assess assay technical performance. Excellent performance of gene expression assays typically covers a dynamic range of $\approx 10^4$ (microarrays) to 10^6 (RNA-Seq and extended-range microarrays). Assessment of technical performance relies on adding the RNA controls in mixed pools; typical pools may be configured to have a 2^{20} (1048576:1) dynamic range between the most- and least-abundant RNA species. This permits assessment of the dynamic range of signal from the measurement system.

Pools are often formulated to be used in sets to permit assessment of the "fold-change" ratio performance of the measurement system. So-called "differential expression" or enrichment of gene expression between pairs of biological conditions is a critical performance property that can be assessed using these "ERCC" controls. The canonical gene expression experiment evaluates enrichment between a "case" and "control". In these typical applications, pools might practically be created by mixing sub-pools (each with its own large dynamic range of abundance), with the ratios between pools for individual controls established by the mixture fraction of the subpools within the pools [5].

Derivation and Selection: The sequences for SRM 2374 were derived largely from a library of sequences gathered by NIST from members of the ERCC [3]. These sequences included a variety of "anti-genomic" synthetic sequences that were designed to have no significant homology with known genomes. Along with these anti-genomic sequences, sequences derived from the organism *Methanocaldococcus jannaschii* (*M. jannaschii*) were submitted by Stanford University (Stanford, CA), and several controls derived from *Bacillus subtilis* (*B. subtilis*) were submitted by Affymetrix, Inc (Santa Clara, CA). Contributors of antigenomic sequences included Affymetrix, Inc., Invitrogen, Inc. (now part of Life Technologies, Grand Island, NY), and Atactic, Inc (Houston, TX). NIST contracted with DNA 2.0 (Menlo Park, CA) for the synthesis of 48 sequences from a random antigenomic library developed in consultation with R. Setterquist of Life Technologies.

All sequences were submitted to the public domain at the same time as a material transfer of some embodiment of the sequence was delivered to NIST. There are 176 controls in the NIST library in total, of which 96 sequences were selected for inclusion in SRM 2374. Selection was done through a collaborative study with laboratories of the ERCC (including participation by J. Warrington and G. Tanimoto of Affymetrix, Inc.; A. Bergstrom-Lucas of Agilent, Inc. [Santa Clara, CA]; J. Lozach of Illumina, Inc. [San Diego, CA]; and a core laboratory operated by T. Myers of the National Institute of Allergies and Infectious Diseases).

Construction: The plasmid construct (derived from the pUC18 vector) was engineered to have IVT promoter sequences (for T7 and T3 RNA polymerases) and common sequencing primer sites (M13F and M13R) around the insert region (which is flanked by restriction enzyme sites). A 24 base (nominal length) polyT tail at the 5′ end of the insert is included so the synthetic RNA controls mimic eukaryotic messenger RNA (mRNA) with a 3′ polyA tail. Figure 2 depicts the vector features of the common plasmid useful in preparing RNA controls and diagrams the restriction site map. The basic vector structure was provided free of license for this application by R. Setterquist of Life Technologies, Inc.

The ERCC sequence library and vector were provided to Commonwealth Biotechnology Inc. (now AIBioTech, Inc., Richmond, VA) for synthesis of RNA that was used in ERCC evaluation and testing. Commonwealth Biotechnology Inc. (CBI) finalized engineering of the vector, cloned the control sequences, and verified the functional cloning with Sanger sequencing. CBI prepared large-scale cultures of *E. coli* containing the plasmid DNA constructs and then isolated, dispensed and dehydrated the plasmid DNA in the tubes as packaged in the SRM.

Note: It is to be expected that some small fraction of the DNA in the SRM components may be *E. coli* genomic DNA. This should not affect the synthesis of RNA controls.

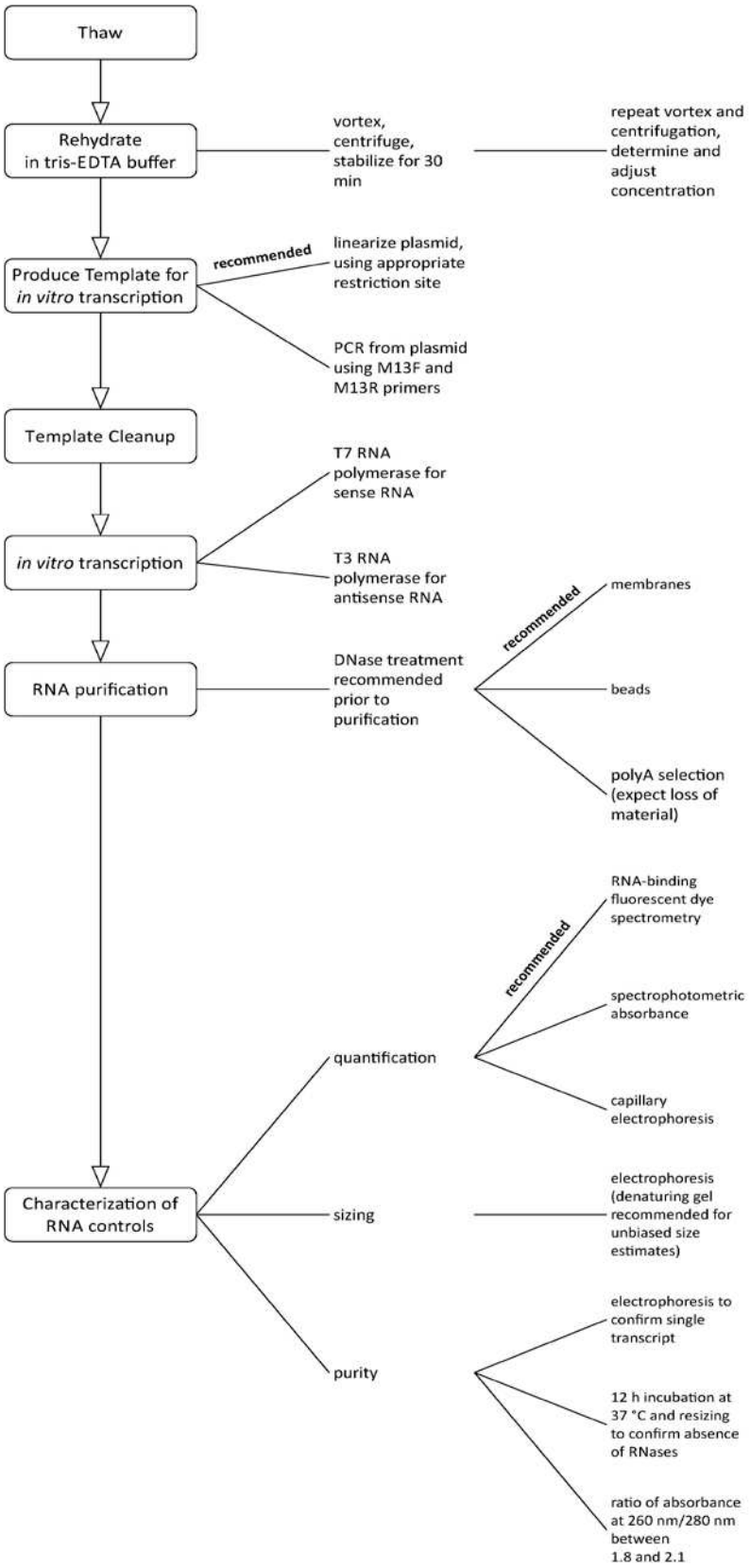

Figure 1. Flowchart for RNA synthesis from DNA template library.

Measurement and Analysis: Sanger sequencing was used to determine the sequence for each component in SRM 2374. Complete sequencing of both strands of DNA in the control insert region was performed on replicate samples, and the sequence reads were independently hand-curated and evaluated for both sequence and confidence estimation by two analysts. Two independent ultra-high throughput short-read sequencing experiments were performed on two different platforms. These data were used to resolve confidence estimates for the full data set, and were also used to identify eight polymorphic bases in the sequences (described in Table 2 and noted in the reference sequence data with IUPAC ambiguity codes).

Homogeneity and Stability: Evaluation of the SRM material stability and homogeneity was performed by subjecting multiple vials of six different controls (spanning a range of GC fraction, length, and source) to an accelerated aging process using elevated temperature over a timecourse. The complete set of materials was fully sequenced to establish the stability. Functional testing was performed using IVT with subsequent evaluation of the RNA. No effects of aging or vial-to-vial differences were observed.

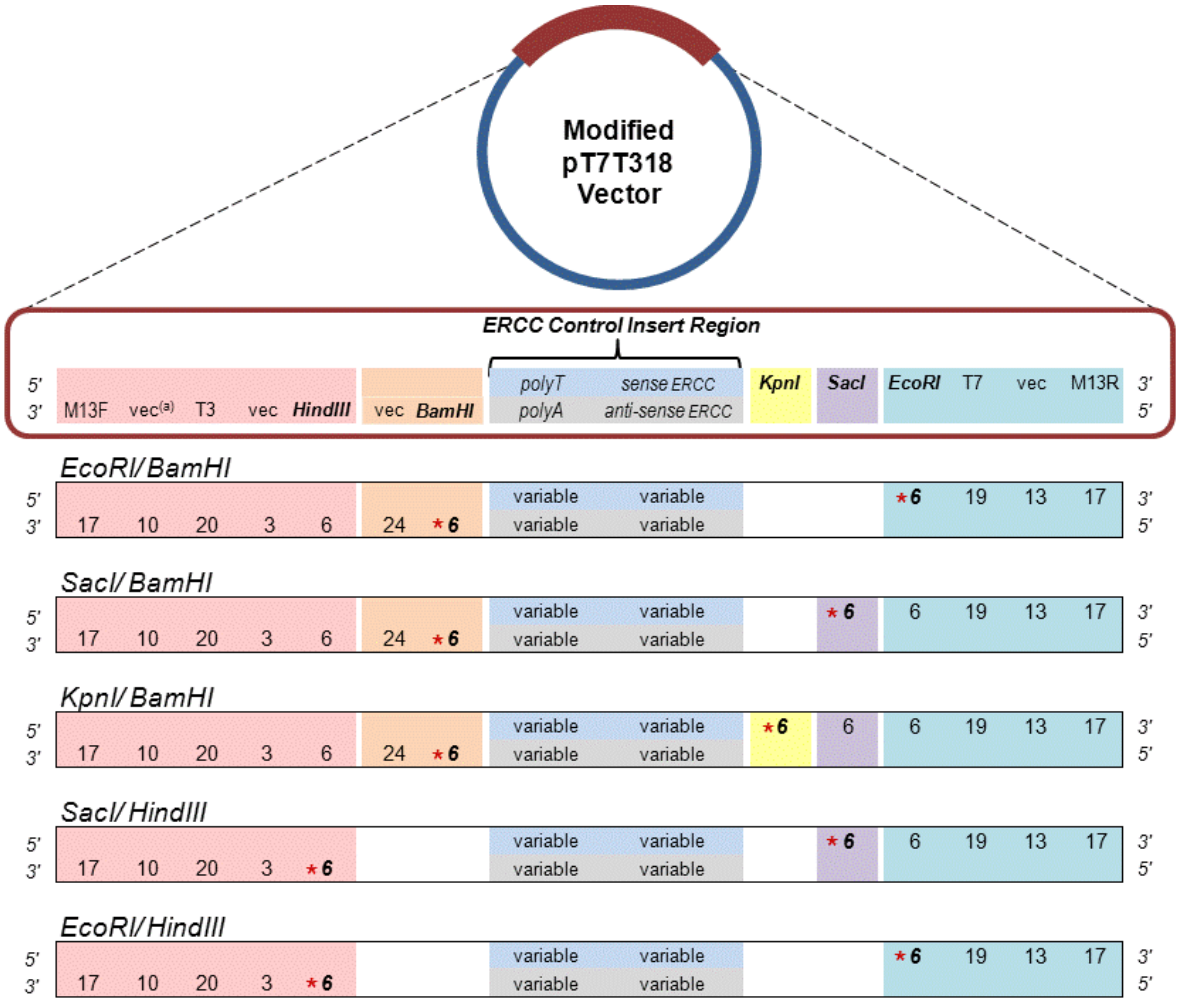

Figure 2. Restriction enzyme recognition site diagram for plasmid DNA template library of SRM 2374. Each component in this library is in one of the five classes of enzyme restriction site pairs. Each class is depicted, with base counts for each flanking feature, which includes, in addition to the restriction enzyme recognition sites, the M13F and M13R sequencing promoter sites (convenient for polymerase chain reaction (PCR) priming); various vector sequence and T3 and T7 IVT promoter sites. Restriction enzyme cut sites are shown in bold, italic-font text with a red asterisk.

[a] vec: vector sequence

Figure 3a. In Vitro Transcription From Linearized Plasmid DNA Template.

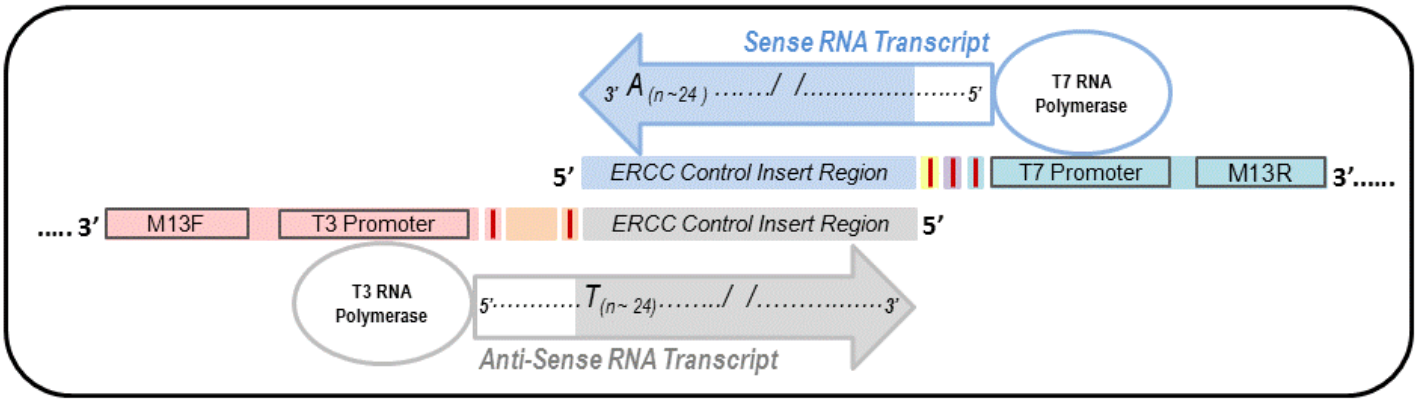

Figure 3b. In Vitro Transcription from M13 Produced PCR Product DNA Template.

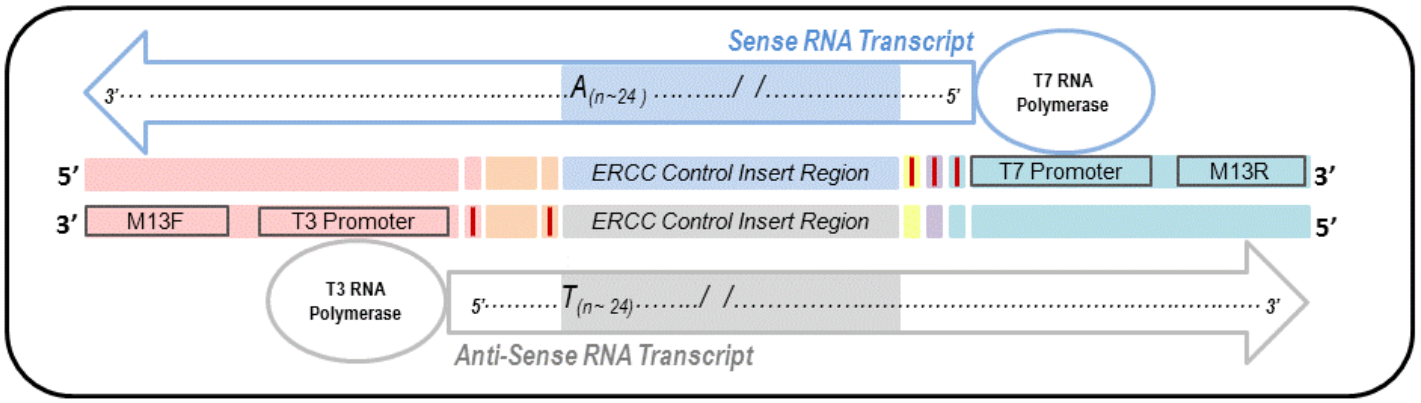

Figure 3. Plasmid vector feature diagram, annotated to illustrate RNA synthesis using in vitro transcription from the various DNA templates, with either T7 or T3 RNA polymerase. Red bars denote restriction enzyme recognition sites.

Table 2. Characteristics of the Components of SRM 2374

Spot Label in Box	Certified Component Plasmid Control[a]	Restriction Sites for Linearization (Antisense/Sense)	Information Values			GenBank Accession Number	Contributor – Source
			Insert Length[b,c]	polyA Tail Length	GC Fraction of Control Insert		
A1	ERCC-00002	EcoRI/BamHI	1061	24	0.53	KC702164	Invitrogen – Synthetic construct microarray control MC28 mRNA, complete sequence
A2	ERCC-00003	EcoRI/BamHI	1023	24	0.33	KC702165	Stanford – *M. jannaschii* spike-in control MJ-1000-67 genomic sequence
A3	ERCC-00004[d]	EcoRI/BamHI	523	24	0.36	KC702166	Stanford – *M. jannaschii* spike-in control MJ-500-35 genomic sequence
A4	ERCC-00007	EcoRI/BamHI	1135	24	0.47	KC702167	Affymetrix – Synthetic construct spike-in microarray control hypothetical protein (ysdC) gene, complete cds
A5	ERCC-00009	EcoRI/BamHI	984	24	0.48	KC702168	Affymetrix – Synthetic construct clone TagJ microarray control
A6	ERCC-00012	EcoRI/BamHI	994	24	0.52	DQ883670	NIST-DNA20 – Synthetic construct clone NISTag29 external RNA control sequence
A7	ERCC-00013	EcoRI/BamHI	808	24	0.44	KC702169	Affymetrix – Synthetic construct spike-in microarray control methionine aminopeptidase (map) gene, partial cds
A8	ERCC-00014	KpnI/BamHI	1957	20	0.45	KC702170	Affymetrix – Synthetic construct microarray control DAP gene, complete sequence
A9	ERCC-00016	SacI/BamHI	844	24	0.50	DQ883664	NIST-DNA20 – Synthetic construct clone NISTag23 external RNA control sequence
A10	ERCC-00017	EcoRI/BamHI	1136	23	0.52	KC702171	Invitrogen – Synthetic construct microarray control MC09 mRNA, complete sequence
A11	ERCC-00018[e]	SacI/BamHI	1026	22	0.44	KC702172	Affymetrix – Synthetic construct spike-in microarray control hypothetical protein (yurP) gene, complete cds
A12	ERCC-00019	EcoRI/BamHI	644	25	0.51	DQ883651	NIST-DNA20 – Synthetic construct clone NISTag10 external RNA control sequence
B1	ERCC-00022	EcoRI/BamHI	751	24	0.49	KC702173	Atactic – Synthetic construct clone AG019.0111 external RNA control sequence
B2	ERCC-00023	EcoRI/BamHI	273	24	0.34	KC702174	Stanford – *M. jannaschii* spike-in control MJ-250-27 genomic sequence
B3	ERCC-00024	EcoRI/BamHI	536	25	0.49	KC702175	Atactic – Synthetic construct clone AG005.0110 external RNA control sequence
B4	ERCC-00025	EcoRI/BamHI	1994	24	0.51	DQ883689	NIST-DNA20 – Synthetic construct clone NISTag48 external RNA control sequence
B5	ERCC-00028	SacI/BamHI	1130	24	0.52	KC702176	Invitrogen – Synthetic construct microarray control MC08 mRNA, complete sequence
B6	ERCC-00031	EcoRI/BamHI	1138	24	0.49	KC702177	Invitrogen – Synthetic construct microarray control MC30 mRNA, complete sequence
B7	ERCC-00033	SacI/BamHI	2022	22	0.33	KC702178	Stanford – *M. jannaschii* spike-in control MJ-2000-79 genomic sequence
B8	ERCC-00034	EcoRI/BamHI	1019	24	0.50	KC702179	Atactic – Synthetic construct clone AG012.1111 external RNA control sequence
B9	ERCC-00035	EcoRI/BamHI	1130	24	0.52	KC702180	Invitrogen – Synthetic construct microarray control MC02 mRNA, complete sequence
B10	ERCC-00039	EcoRI/BamHI	740	24	0.51	DQ883656	NIST-DNA20 – Synthetic construct clone NISTag15 external RNA control sequence
B11	ERCC-00040	EcoRI/BamHI	744	24	0.54	DQ883661	NIST-DNA20 – Synthetic construct clone NISTag20 external RNA control sequence
B12	ERCC-00041	KpnI/BamHI	1123	22	0.46	KC702181	Affymetri – Synthetic construct spike-in microarray control O-sialoglycoprotein endopeptidase (gcp) gene, complete cds
C1	ERCC-00042	EcoRI/BamHI	1023	24	0.40	KC702182	Stanford – *M. jannaschii* spike-in control MJ-1000-66 genomic sequence
C2	ERCC-00043[f]	EcoRI/BamHI	1023	24	0.34	KC702183	Stanford – *M. jannaschii* spike-in control MJ-1000-70 genomic sequence

Table 2. Characteristics of the Components of SRM 2374

Spot Label in Box	Certified Component Plasmid Control[a]	Restriction Sites for Linearization (Antisense/Sense)	Information Values			GenBank Accession Number	Contributor – Source
			Insert Length[b,c]	polyA Tail Length	GC Fraction of Control Insert		
C3	ERCC-00044	SacI/HindIII	1156	21	0.51	KC702184	Invitrogen – Synthetic construct microarray control MC19 mRNA, complete sequence
C4	ERCC-00046	EcoRI/BamHI	522	24	0.37	KC702185	Stanford – *M. jannaschii* spike-in control MJ-500-31 genomic sequence
C5	ERCC-00048	EcoRI/BamHI	992	24	0.49	DQ883671	NIST-DNA20 – Synthetic construct clone NISTag30 external RNA control sequence
C6	ERCC-00051	EcoRI/BamHI	274	24	0.38	KC702186	Stanford – *M. jannaschii* spike-in control MJ-250-23 genomic sequence
C7	ERCC-00053	EcoRI/BamHI	1023	24	0.32	KC702187	Stanford – *M. jannaschii* spike-in control MJ-1000-68 genomic sequence
C8	ERCC-00054	EcoRI/BamHI	274	26	0.41	KC702188	Stanford – *M. jannaschii* spike-in control MJ-250-14 genomic sequence
C9	ERCC-00057	SacI/BamHI	1021	21	0.51	KC702189	Affymetrix – Synthetic construct clone TagQ microarray control
C10	ERCC-00058	EcoRI/BamHI	1136	24	0.51	KC702190	Invitrogen – Synthetic construct microarray control MC07 mRNA, complete sequence
C11	ERCC-00059	EcoRI/BamHI	525	24	0.50	KC702191	Affymetrix – Synthetic construct clone TagA microarray control
C12	ERCC-00060	EcoRI/BamHI	523	25	0.33	KC702192	Stanford – *M. jannaschii* spike-in control MJ-500-46 genomic sequence
D1	ERCC-00061	EcoRI/BamHI	1136	24	0.51	KC702193	Invitrogen – Synthetic construct microarray control MC22 mRNA, complete sequence
D2	ERCC-00062	EcoRI/BamHI	1023	24	0.32	KC702194	Stanford – *M. jannaschii* spike-in control MJ-1000-69 genomic sequence
D3	ERCC-00067	EcoRI/BamHI	644	24	0.49	DQ883653	NIST-DNA20 – Synthetic construct clone NISTag12 external RNA control sequence
D4	ERCC-00069	EcoRI/BamHI	1137	24	0.51	KC702195	Invitrogen – Synthetic construct microarray control MC10 mRNA, complete sequence
D5	ERCC-00071	EcoRI/BamHI	642	24	0.50	DQ883654	NIST-DNA20 – Synthetic construct clone NISTag13 external RNA control sequence
D6	ERCC-00073	EcoRI/BamHI	603	24	0.49	KC702196	Affymetrix – Synthetic construct clone TagC microarray control
D7	ERCC-00074	EcoRI/BamHI	522	24	0.36	KC702197	Stanford – *M. jannaschii* spike-in control MJ-500-37 genomic sequence
D8	ERCC-00075	EcoRI/BamHI	1023	24	0.37	KC702198	Stanford – *M. jannaschii* spike-in control MJ-1000-61 genomic sequence
D9	ERCC-00076	EcoRI/BamHI	642	24	0.52	DQ883650	NIST-DNA20 – Synthetic construct clone NISTag9 external RNA control sequence
D10	ERCC-00077	EcoRI/BamHI	273	23	0.36	KC702199	Stanford – *M. jannaschii* spike-in control MJ-250-25 genomic sequence
D11	ERCC-00078	EcoRI/BamHI	993	24	0.51	DQ883673	NIST-DNA20 – Synthetic construct clone NISTag32 external RNA control sequence
D12	ERCC-00079	EcoRI/BamHI	644	24	0.51	DQ883652	NIST-DNA20 – Synthetic construct clone NISTag11 external RNA control sequence
E1	ERCC-00081	EcoRI/BamHI	534	25	0.51	KC702200	Atactic – Synthetic construct clone AG002.0011 external RNA control sequence
E2	ERCC-00083[g,h]	SacI/BamHI	1023	23	0.35	KC702201	Stanford – *M. jannaschii* spike-in control MJ-1000-63 genomic sequence
E3	ERCC-00084	EcoRI/BamHI	994	24	0.52	DQ883682	NIST-DNA20 – Synthetic construct clone NISTag41 external RNA control sequence
E4	ERCC-00085	EcoRI/BamHI	844	24	0.50	DQ883669	NIST-DNA20 – Synthetic construct clone NISTag28 external RNA control sequence
E5	ERCC-00086	SacI/BamHI	1020	22	0.33	KC702202	Stanford – *M. jannaschii* spike-in control MJ-1000-74 genomic sequence
E6	ERCC-00092	EcoRI/BamHI	1124	24	0.51	KC702203	Invitrogen – Synthetic construct microarray control MC20 mRNA, complete sequence

Table 2. Characteristics of the Components of SRM 2374

Spot Label in Box	Certified Component Plasmid Control[a]	Restriction Sites for Linearization (Antisense/Sense)	Information Values			GenBank Accession Number	Contributor – Source
			Insert Length[b,c]	polyA Tail Length	GC Fraction of Control Insert		
E7	ERCC-00095	SacI/BamHI	521	22	0.39	KC702204	Stanford – *M. jannaschii* spike-in control MJ-500-42 genomic sequence
E8	ERCC-00096	EcoRI/BamHI	1107	24	0.52	KC702205	Invitrogen – Synthetic construct microarray control MC27 mRNA, complete sequence
E9	ERCC-00097	EcoRI/BamHI	523	25	0.38	KC702206	Stanford – *M. jannaschii* spike-in control MJ-500-41 genomic sequence
E10	ERCC-00098	EcoRI/BamHI	1143	25	0.53	KC702207	Invitrogen – Synthetic construct microarray control MC04 mRNA, complete sequence
E11	ERCC-00099	KpnI/BamHI	1350	26	0.42	KC702208	Affymetrix – Synthetic construct microarray control PHE gene, complete sequence
E12	ERCC-00104[h]	SacI/BamHI	2022	22	0.33	KC702209	Stanford – *M. jannaschii* spike-in control MJ-2000-98 genomic sequence
F1	ERCC-00108	EcoRI/BamHI	1022	25	0.50	KC702210	Affymetrix – Synthetic construct clone TagO microarray control
F2	ERCC-00109	EcoRI/BamHI	536	24	0.48	KC702211	Atactic – Synthetic construct clone AG009.1100 external RNA control sequence
F3	ERCC-00111	EcoRI/BamHI	994	24	0.48	DQ883685	NIST-DNA20 – Synthetic construct clone NISTag44 external RNA control sequence
F4	ERCC-00112	EcoRI/BamHI	1136	24	0.48	KC702212	Invitrogen – Synthetic construct microarray control MC14 mRNA, complete sequence
F5	ERCC-00113	EcoRI/BamHI	840	23	0.52	DQ883663	NIST-DNA20 – Synthetic construct clone NISTag22 external RNA control sequence
F6	ERCC-00116	EcoRI/HindIII	1991	22	0.51	KC702213	Affymetrix – Synthetic construct clone TagIN microarray control
F7	ERCC-00117	EcoRI/BamHI	1136	24	0.52	KC702214	Invitrogen – Synthetic construct microarray control MC01 mRNA, complete sequence
F8	ERCC-00120	EcoRI/BamHI	536	24	0.51	KC702215	Atactic – Synthetic construct clone AG003.0011 external RNA control sequence
F9	ERCC-00123	EcoRI/BamHI	1022	24	0.36	KC702216	Stanford – *M. jannaschii* spike-in control MJ-1000-65 genomic sequence
F10	ERCC-00126[h]	SacI/BamHI	1119	21	0.52	KC702217	Invitrogen – Synthetic construct microarray control MC25 mRNA, complete sequence
F11	ERCC-00128	EcoRI/BamHI	1133	24	0.49	KC702218	Invitrogen – Synthetic construct microarray control MC26 mRNA, complete sequence
F12	ERCC-00130	SacI/BamHI	1059	22	0.47	KC702219	Affymetrix – Synthetic construct spike-in microarray control hypothetical protein (ynbA) gene, partial cds
G1	ERCC-00131[i]	EcoRI/BamHI	771	24	0.49	KC702220	Atactic – Synthetic construct clone AG016.1110 external RNA control sequence
G2	ERCC-00134[j]	EcoRI/BamHI	274	25	0.35	KC702221	Stanford – *M. jannaschii* spike-in control MJ-250-22 genomic sequence
G3	ERCC-00136	SacI/BamHI	1033	22	0.42	KC702222	Affymetrix – Synthetic construct spike-in microarray control DNA-directed RNA polymerase alpha subunit (rpoA) gene, complete cds
G4	ERCC-00137	EcoRI/BamHI	537	24	0.52	KC702223	Atactic – Synthetic construct clone AG011.0011 external RNA control sequence
G5	ERCC-00138[k]	EcoRI/BamHI	1022	25	0.34	KC702224	Stanford – *M. jannaschii* spike-in control MJ-1000-60 genomic sequence
G6	ERCC-00142	EcoRI/BamHI	493	23	0.52	DQ883646	NIST-DNA20 – Synthetic construct clone NISTag5 external RNA control sequence
G7	ERCC-00143	EcoRI/BamHI	784	25	0.50	KC702225	Affymetrix – Synthetic construct clone TagG microarray control

Table 2. Characteristics of the Components of SRM 2374

Spot Label in Box	Certified Component Plasmid Control[a]	Restriction Sites for Linearization (Antisense/Sense)	Information Values			GenBank Accession Number	Contributor – Source
			Insert Length[b,c]	polyA Tail Length	GC Fraction of Control Insert		
G8	ERCC-00144	EcoRI/BamHI	538	25	0.48	KC702226	Atactic – Synthetic construct clone AG006.1100 external RNA control sequence
G9	ERCC-00145[l]	EcoRI/BamHI	1042	26	0.45	KC702227	Affymetrix – Synthetic construct microarray control LYS gene, complete sequence
G10	ERCC-00147	EcoRI/BamHI	1023	24	0.37	KC702228	Stanford – *M. jannaschii* spike-in control MJ-1000-73 genomic sequence
G11	ERCC-00148	EcoRI/BamHI	494	24	0.51	DQ883642	NIST-DNA20 – Synthetic construct clone NISTag1 external RNA control sequence
G12	ERCC-00150	EcoRI/BamHI	743	24	0.49	DQ883659	NIST-DNA20 – Synthetic construct clone NISTag18 external RNA control sequence
H1	ERCC-00154	EcoRI/BamHI	537	24	0.52	KC702229	Atactic – Synthetic construct clone AG009.0011 external RNA control sequence
H2	ERCC-00156	EcoRI/BamHI	494	24	0.51	DQ883643	NIST-DNA20 – Synthetic construct clone NISTag2 external RNA control sequence
H3	ERCC-00157	EcoRI/BamHI	1019	24	0.51	KC702230	Atactic – Synthetic construct clone AG090.1111 external RNA control sequence
H4	ERCC-00158	SacI/BamHI	1021	23	0.35	KC702231	Stanford – *M. jannaschii* spike-in control MJ-1000-78 genomic sequence
H5	ERCC-00160	EcoRI/BamHI	743	24	0.47	DQ883658	NIST-DNA20 – Synthetic construct clone NISTag17 external RNA control sequence
H6	ERCC-00162	EcoRI/BamHI	523	24	0.38	KC702232	Stanford – *M. jannaschii* spike-in control MJ-500-33 genomic sequence
H7	ERCC-00163	EcoRI/BamHI	543	24	0.49	KC702233	Affymetrix – Synthetic construct clone TagD microarray control
H8	ERCC-00164	EcoRI/BamHI	1022	23	0.38	KC702234	Stanford – *M. jannaschii* spike-in control MJ-1000-62 genomic sequence
H9	ERCC-00165	EcoRI/BamHI	872	24	0.51	KC702235	Affymetrix – Synthetic construct clone TagH microarray control
H10	ERCC-00168	EcoRI/BamHI	1024	25	0.35	KC702236	Stanford – *M. jannaschii* spike-in control MJ-1000-59 genomic sequence
H11	ERCC-00170[h]	EcoRI/BamHI	1024	24	0.35	KC702237	Stanford – *M. jannaschii* spike-in control MJ-1000-56 genomic sequence
H12	ERCC-00171	EcoRI/BamHI	505	24	0.50	KC702238	Atactic – Synthetic construct clone AG006.0011 external RNA control sequence

[a] Any uncertainty in the base assignment or note about the specific plasmid is included in a footnote to the plasmid control.
[b] The insert length includes the polyA tail length.
[c] The complete RNA transcript length can be calculated using Figures 2 and 3.
[d] Ambiguity Y at base 153, most likely base call for this position is "C".
[e] Ambiguity R at base 599, most likely base call for this position is "A".
[f] Ambiguity Y at base 584, most likely base call for this position is "C".
[g] ERCC-00083 contains 2 SacI restriction sites, being within the control sequence. SacI is the 5' restriction site and will not be an issue for preparing the linearized plasmid for sense RNA IVT. Anti-sense RNA cannot be produced with the transcript using a linearized plasmid.
[h] At the 3' end of the polyA tail there is a "G".
[i] Ambiguities Y and R respectively at bases 150 and 674, most likely base calls for this positions is "C" and "G".
[j] Ambiguity M at base 29, most likely base call for this position is "A".
[k] Ambiguity Y at base 403, most likely base call for this position is "C". Ambiguities K, K, and R at bases 994, 995 and 997, respectively, as a result of ambiguous primary CE data and lack of coverage for secondary NGS data.
[l] Ambiguity Y at base 749, most likely base call for this position is "T".

REFERENCES

[1] Certified data can be downloaded at https://www-s.nist.gov/srmors/view_detail.cfm?srm=2374 (accessed Dec 2017). The certified sequence data file is: SRM2374_Sequence_v1.FASTA. The certified confidence estimates for every base of sequence data file is: SRM2374_Quality_v1.FASTA. Additional, secondary files are also available which include a simplified representation of the sequence data (with IUPAC ambiguity codes resolved using a best likelihood estimate of the nucleotide):

- SRM2374_ambiguities_resolved_v2.FASTA is the sequence data file with IUPAC ambiguity codes resolved;
- SRM2374_ambiguities_resolved_v1.GTF describes the structures of the RNA controls as if they were genes;
- SRM2374_ambiguities_resolved_NoPolyA_v2.FASTA is the sequence reference file with the polyA tail excluded;
- SRM2374_ambiguities_resolved_NoPolyA_v1.GTF is a GTF with the polyA tail excluded;
- SRM2374_PlasmidVector_by_ERCCNumber_v1.FASTA is the sequence reference file of the plasmid vector.
- SRM2374_putative_T7_products_NoPolyA_v2.FASTA

[2] JCGM 100:2008; *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement* (GUM 1995 with Minor Corrections); Joint Committee for Guides in Metrology (2008); available at http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf (accessed Dec 2017); see also Taylor, B.N.; Kuyatt, C.E.; *Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results*; NIST Technical Note 1297; U.S. Government Printing Office: Washington, DC (1994); available at http://www.nist.gov/pml/pubs/tn1297/index.cfm (accessed Dec 2017).

[3] Baker, S.C.; Bauer, S.R.; Beyer, R.P.; Brenton, J.D; Bromley, B.; Burrill, J; Causton, H.; Conley, M.P.; Elespuru, R.; Fero, M.; Foy, C.; Fuscoe, J.; Gao, X.; Gerhold, D.L.; Gilles, P.; Goodsaid, F.; Guo, X.; Hackett, J.; Hockett, R.D.; Ikonomi, P.; Irizarry, R.A.; Kawasaki, E.S.; Kaysser-Kranich, T.; Kerr, K.; Kiser, G.; Koch, W.H.; Lee, K.Y.; Liu, C.; Liu, Z.L.; Lucas, A.; Manohar, C.F.; Miyada, G.; Modrusan, Z.; Parkes, H.; Puri, R.K.; Reid, L.; Ryder, T.B.; Salit, M.; Samaha, R.R.; Scherf, U.; Sendera, T.J.; Setterquist, R.A.; Shi, L.; Shippy, R.; Soriano, J.V.; Wagar, E.A.; Williams, M.; Wilmer, F.; Wilson, M.; Wolber, P.K.; Wu, X.; Zadro, R.; *The External RNA Controls Consortium: A Progress Report*; Nat. Methods, Vol. 2, pp. 731–734 (2005).

[4] CDC/NIH: *Biosafety in Microbiological and Biomedical Laboratories,* 5th ed.; HHS publication No. (CDC) 21-1112; Chosewood, L.C.; Wilson, D.E.; Eds.; US Government Printing Office: Washington, D.C. (2009); available at http://www.cdc.gov/biosafety/publications/bmbl5/ (accessed Dec 2017).

[5] External RNA Control Consortium, *Proposed Methods for Testing and Selecting the ERCC External RNA Controls*; BMC Genomics, Vol. 6, p. 150 (2005).

[6] Munro, S.A.; Lund, S.P.; Pine, P.S.; Binder, H.; Clevert, D.A.; Conesa, A.; Dopazo, J.; Fasold, M.; Hochreiter, S.; Hong, H.; Jafari, N.; Kreil, D.P.; Labaj, P.P; Li, S.; Liao, Y.; Lin, S.M.; Meehan, J.; Mason, C.E.; Santoyo-Lopez, J.; Setterquist, R.A.; Shi, L.; Shi, W.; Smyth, G.K.; Stralis-Pavese, N.; Su, Z.; Tong, W.; Wang, C.; Wang, J.; Xu, J.; Ye, Z.; Yang, Y.; Yu, Y.; Salit, M.; *Assessing Technical Performance in Differential Gene Expression Experiments With External spike-in RNA Control Ratio Mixtures*; Nature Communications, Vol. 5; pp. 5125 (2014).

[7] Munro. S.A.; Lund, S.; *Bioconductor erccdashboard R Package: Assess Differential Gene Expression Experiments with ERCC Controls*; available at https://bioconductor.org/packages/release/bioc/html/erccdashboard.html (accessed Dec 2017)

[8] Jiang, L.; Schlesinger, F.; Davis, C.A.; Zhang, Y; Li, R.; Salit, M.; Gingeras, T.R.; Oliver, B,; *Synthetic Spike-in Standards for RNA-seq Experiments*; Genome Res., Vol. 9, pp. 1543–51 (2011).

[9] Lee, H.; Pine, P.S.; McDaniel, J; Salit, M.; Oliver, B.; *External RNA Controls Consortium Beta Version Update*; J. Genomics, Vol. 4, pp. 19–22 (2016).

[10] Pine, P.S.; Munro S.A.; Parsons, J.R.; McDaniel, J; Lucas, A.B.; Lozach, J.; Myers, T.G.; Su, Q.; Jacob-Helber, S.M; Salit, M.; *Evaluation of the External RNA Controls Consortium (ERCC) Reference Material Using a Modified Latin Square Design*; BMC Biotechnology, Vol. 16(1), p. 54 (2016).

[11] Information on the ERCC can be found on the ERCC website http://jimb.stanford.edu/ercc/ (accessed Dec 2017)

Certificate Revision History: 06 December 2017 (Change expiration date; Instructions for Use updated; additional FASTA file included; updated references; editorial changes); 20 March 2013 (Revised storage instructions to include unit shipment packaging information; revised data file names in reference 1; editorial revisions); 07 March 2013 (Original certificate date).

Users of this SRM should ensure that the Certificate of Analysis in their possession is current. This can be accomplished by contacting the SRM Program: telephone (301) 975-2200; fax (301) 948-3730; e-mail srminfo@nist.gov; or via the Internet at http://www.nist.gov/srm.