nature biotechnology

Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study

Sheng Li[1,2,24], Scott W Tighe[3,24], Charles M Nicolet[4], Deborah Grove[5], Shawn Levy[6], William Farmerie[7], Agnes Viale[8], Chris Wright[9], Peter A Schweitzer[10], Yuan Gao[11], Dewey Kim[11], Joe Boland[12], Belynda Hicks[12], Ryan Kim[13,23], Sagar Chhangawala[1,2], Nadereh Jafari[14], Nalini Raghavachari[15], Jorge Gandara[1,2], Natàlia Garcia-Reyero[16], Cynthia Hendrickson[6], David Roberson[12], Jeffrey A Rosenfeld[17], Todd Smith[18], Jason G Underwood[19], May Wang[20], Paul Zumbo[1,2], Don A Baldwin[21], George S Grills[10] & Christopher E Mason[1,2,22]

High-throughput RNA sequencing (RNA-seq) greatly expands the potential for genomics discoveries, but the wide variety of platforms, protocols and performance capabilitites has created the need for comprehensive reference data. Here we describe the Association of Biomolecular Resource Facilities next-generation sequencing (ABRF-NGS) study on RNA-seq. We carried out replicate experiments across 15 laboratory sites using reference RNA standards to test four protocols (poly-A–selected, ribo-depleted, size-selected and degraded) on five sequencing platforms (Illumina HiSeq, Life Technologies PGM and Proton, Pacific Biosciences RS and Roche 454). The results show high intraplatform (Spearman rank $R > 0.86$) and inter-platform ($R > 0.83$) concordance for expression measures across the deep-count platforms, but highly variable efficiency and cost for splice junction and variant detection between all platforms. For intact RNA, gene expression profiles from rRNA-depletion and poly-A enrichment are similar. In addition, rRNA depletion enables effective analysis of degraded RNA samples. This study provides a broad foundation for cross-platform standardization, evaluation and improvement of RNA-seq.

RNA-seq is an important analytical technique that leverages the capacity of high-throughput sequencing instruments to quantitatively sample a population of RNA molecules with a large number of 'reads' or parallel reactions on discrete templates[1,2]. Depending on experimental goals, sample types and read depths, results from RNA-seq data can be similar or superior to those from microarray data[3–5]. However, each sequencing platform has unique aspects of library synthesis, sequencing, alignment and data processing[6–9]. Thus, many questions remain about RNA-seq in regards to interoperability between platforms, cross-site reproducibility, bioinformatics methods and the sources of variance in results with both existing and emerging protocols, such as those for degraded RNA.

Notably, prior work comparing microarray platforms and methods showed high levels of inter-platform concordance for the ability to detect differentially expressed genes. The Microarray Quality Control (MAQC) Consortium landmark study[10] examined the degree of variance within and across many different microarray platforms and found similar coefficients of variation between platforms. The MAQC data also provided an important benchmark for the application of microarray technologies to clinical assays. For high-throughput sequencing platforms, however, very little data exist about cross-site variation of expression measures. Only two inter-site variation studies are publicly available: the Sequencing Qualtiy Control (SEQC)/MAQC-III Consortium[11] study and the GEUVADIS Consortium study[12]. These studies were either limited to one platform or did not assess some newer RNA-seq methods that are now widely used. Moreover, important RNA profiling parameters, such as differential expression and splice variant detection, have not been consistently evaluated. Thus, these studies do not answer key questions about the degree of concordance for RNA-seq across platforms and methods and also about

[1]Department of Physiology and Biophysics, Weill Cornell Medical College, New York, New York, USA. [2]The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, USA. [3]Vermont Cancer Center, University of Vermont, Burlington, Vermont, USA. [4]Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [5]The Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, Pennsylvania, USA. [6]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA. [7]Interdisciplinary Center for Biotechnology Research, University of Florida, Gainesville, Florida, USA. [8]Memorial Sloan-Kettering Cancer Institute, New York, New York, USA. [9]Roy J. Carver Biotechnology Center, University of Illinois, Urbana, Illinois, USA. [10]Biotechnology Resource Center, Institute of Biotechnology, Cornell University, Ithaca, New York, USA. [11]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA. [12]NIH/NCI/SAIC-Frederick, Gaithersburg, Maryland, USA. [13]Genome Center, University of California, Davis, Davis, California, USA. [14]Center for Genetic Medicine, Northwestern University, Chicago, Illinois, USA. [15]NIH/NHLBI, Bethesda, Maryland, USA. [16]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, Mississippi, USA. [17]Division of High Performance and Research Computing, University of Medicine and Dentistry of New Jersey, Newark, New Jersey, USA. [18]PerkinElmer Inc., Seattle, Washington, USA. [19]Department of Genome Sciences, University of Washington. Seattle, Washington, USA. [20]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, USA. [21]Pathonomics LLC, Philadelphia, Pennsylvania, USA. [22]The Feil Family Brain and Mind Research Institute, New York, New York, USA. [23]Present address: Korean Bioinformation Center (KOBIC), National Center for Biological Research Resource Information, Daejeon, South Korea. [24]These authors contributed equally to this work. Correspondence should be addressed to D.A.B. (pathonomics@gmail.com), G.S.G. (gsg34@cornell.edu) or C.E.M. (chm2042@med.cornell.edu).

Received 14 May 2013; accepted 1 July 2014; published online 24 August 2014; corrected after print 10 October 2014; doi:10.1038/nbt.2972

the read depth, type and length of sequence reads required to fully characterize a sample with current techniques. Moreover, RNA-seq is an extremely useful method for exploring the expression of sequence variants, detecting novel RNAs and for discriminating between transcript splicing isoforms[13–20], but there is no gold standard of reference data on the dynamic range of differential expression and splicing that includes different sample preparation protocols, instruments and data analysis strategies.

To address this challenge, members of the Association of Biomolecular Resource Facilities (ABRF)[21] designed and conducted the first phase of a large-scale ABRF-NGS study with a focus on RNA-seq. The goals of the ABRF-NGS study are to evaluate the performance of high-throughput sequencing platforms and to identify optimal methods and best practices. A wide range of variables was evaluated, including library preparation methods (poly-A–enriched and ribo-depleted), size-specific fractionation (1, 2 and 3 kb) and RNA integrity (using heat, RNase A and sonication to degrade the RNA). The latter variable was chosen to mimic some of the damaging effects of tissue fixation with formalin, which is a well-recognized issue for RNA profiling of formalin-fixed, paraffin-embedded (FFPE) clinical specimens[22–24]. Finally, we leveraged a data set of 18,124 PrimePCR reactions, also used by the SEQC Consortium[11]. These primers' data were compared to the exact loci in the transcriptome coordinates of GENCODE v12, which is known to enable a more accurate comparison than the overall gene counts[25]. Both platform-agnostic and platform-specific sequence aligners were also compared to support the validity of our analyses. Taken together, these data represent a broad cross-platform characterization of widely used RNA standards and to our knowledge provide the largest comprehensive comparison of results from degraded, full-length and size-selected RNA across sequencing platforms and protocols.

RESULTS

Platforms, RNA samples and sequencing protocols

Although comparisons of high-throughput sequencing platforms and sample preparation protocols have been reported in past studies[6,26–28], no other study has been conducted using five platforms and two standardized RNA samples replicated at multiple sites (**Fig. 1**). Platforms evaluated included the Illumina HiSeq 2000/2500, Roche 454 GS FLX+, Life Technologies Ion Personal Genome Machine (PGM) and Proton, and the Pacific Biosciences RS (PacBio)[6,8,29]. Data were generated and analyzed by the members of five ABRF Research Groups, including 25 core facilities at 20 different institutions (**Fig. 1** and **Supplementary Table 1**). Additional data from an Illumina MiSeq v2 instrument were used to compare metrics derived from different read lengths from the same Illumina library preparation and sequencing methods. Detection of differential RNA abundance was evaluated using two commercially available and very distinct RNA samples: A, RNA from cancer cell lines; B, RNA from pooled normal human brain tissues; and two predefined mixtures of these samples (C: 75% A + 25% B; D: 25% A + 75% B). All standardized RNA samples also contained synthetic RNA spike-ins from the External RNA Control Consortium (ERCC)[10,30,31]. Results from high-quality RNA on the Illumina HiSeq 2500 platform were compared to results on the same platform from RNAs degraded using three degradation conditions: heat, RNase and sonication. The RNA reference samples were degraded to a RIN (RNA integrity number) of 3 or less. In addition, results from ribosomal RNA-depleted and poly-A–enriched libraries from intact RNA were compared using the Illumina HiSeq 2500 platform.

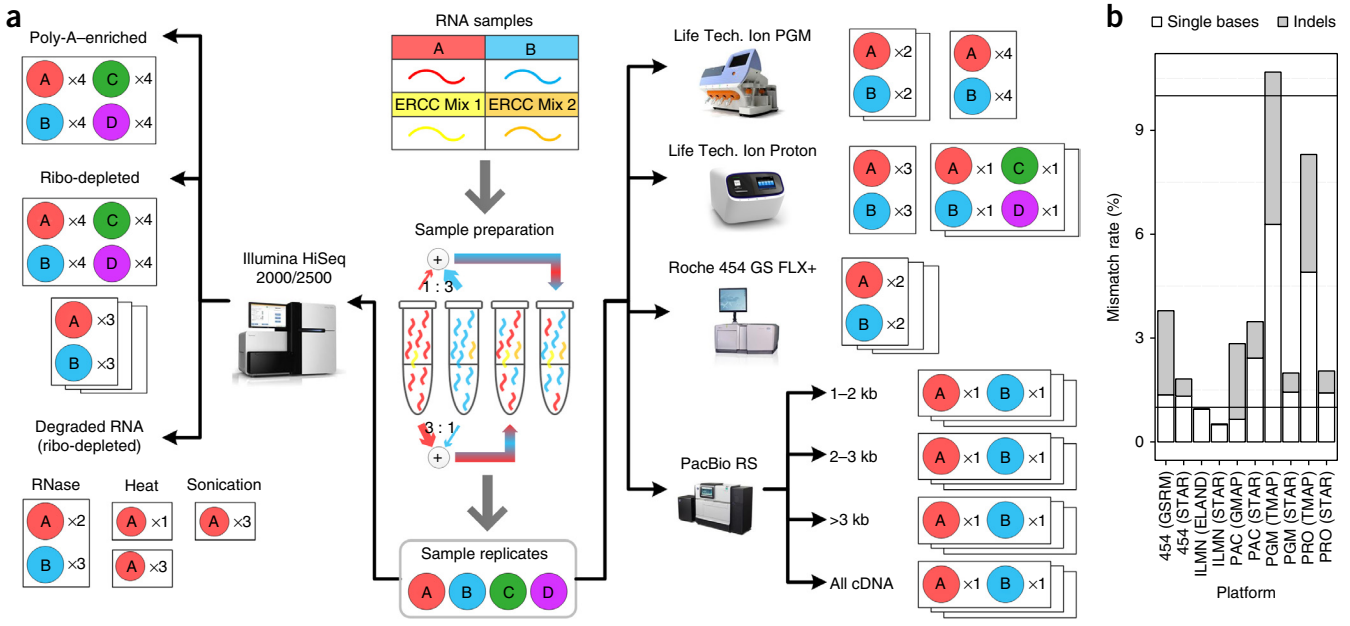

Figure 1 Experimental design and sequencing platforms. (**a**) Two standard RNA samples (A = Universal Human Reference RNA and B = Human Brain Reference RNA) were combined with two sets of synthetic RNAs (ERCCs) to prepare a set of samples to be sequenced on five platforms: Illumina (ILMN) HiSeq 2000/2500, Life Technologies Personal Genome Machine (PGM), Life Technologies Proton (PRO), Pacific Biosciences (PacBio) RS (PAC), and the Roche 454 GS FLX+. Additional RNA samples were also generated: samples C and D were prepared as defined mixtures of A and B, while other aliquots of A and B were degraded by three methods. All these additional samples were ribo-depleted for RNA-seq on the HiSeq platform. The number of technical replicates (×2, ×3 or ×4) of each sample set is indicated for each platform and method. The number of stacked rectangles indicates the number of sites performing the same experiment. (**b**) Stacked bar plots of the sequencing platforms' mismatch rates (*y* axis) for single-base mismatches (white) and insertions/deletions (indels, gray) based on different aligners for each platform (*x* axis). Q10 (90% accuracy) and Q20 (99% accuracy) are shown as the top and bottom line, respectively. *X* axis indicates the platform name, with the aligner name in parentheses.

To map the sequencing reads to the human genome (hg19), we used both vendor-recommended alignment algorithms and 'universal', platform-agnostic aligners. For gene expression quantification, the following aligners were evaluated: STAR[32] (agnostic), ELAND (HiSeq), TMAP (PGM and Proton), GSRM (454) and GMAP (PacBio). With the exception of ELAND, each platform-specific algorithm produced better mapping rates, gene-body coverage evenness and Spearman correlations with PrimePCR quantification (**Supplementary Tables 2–4**) when compared to STAR applied uniformly across all platforms. However, the universal STAR alignments were used as input for shared junction detection (**Supplementary Table 5**), as these alignments always showed the lowest mapping error rate (**Fig. 1**). After mapping, additional processing for quantifying gene counts was performed using the open source r-make package (http://physiology.med.cornell.edu/faculty/mason/lab/r-make/ and Online Methods) to calculate the reads and coverage for each gene feature based on GENCODE (v12) annotation. Quality control data were generated using the fastQC package (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) to calculate a large set of performance metrics for sequence quality, gene coverage, and transcriptome quantification and characterization for all platforms (**Fig. 1** and **Supplementary Figs. 1–23**).

Base qualities, data quality and duplicate rates

Quality values (QV, a per base accuracy estimate) were calculated for all sample runs and for prealignment measures (**Supplementary Figs. 1–6**) and postalignment measures (**Fig. 1b**). Results ranged from Q10 (90% accuracy) to Q60 (99.9999% accuracy) across platforms (**Supplementary Figs. 1–6**) and revealed three notable trends. First, most platforms show a biased QV distribution in the first 1–16 bases, a known effect from the reverse transcriptase (RT) priming step[33]. This RT bias can also affect the observed GC content (**Supplementary Figs. 7–11**) and base-frequency data[11,34,35] (**Supplementary Figs. 12–17**). Second, similar QV profiles were observed for samples A and B, and across different RNA size fractions. Third, although changes in library preparation techniques and sequencing chemistry for various platforms can affect the QVs, the largest increase in QVs came from the circular consensus sequencing for the PacBio data (**Supplementary Fig. 2**), where median QVs near 40 were observed, though with a wide range of variation. Thus, for most platforms, the ends of the reads are where most noise was observed, but lower QVs also occurred at the beginnings of the reads. This results in a source of bias and noise for RNA-seq data that appears in all platforms and is usually addressed by appropriate sequence trimming.

The QVs for each base of a read, as well as the read length, alignment method and reference sequence quality, can all affect mapping accuracy. To estimate the platform-specific and aligner-specific impact of the sequencing error rate on alignment, we calculated the number of mismatches relative to the hg19 human reference genome, normalized by total mapped bases, for two aligners for each platform (**Fig. 1b**). These data showed that a tradeoff between higher mapping rate and accuracy can occur for RNA-seq, such as the increased mapping rate with TMAP and GSRM versus STAR (**Supplementary Table 2**) that led to a higher empirically derived error rate (**Fig. 1b**). The most common type of mismatch was single-base substitutions, with frequencies ranging from 0.6 to 7.1% across all platforms. Insertion/deletion (indel) type mismatch rates were also highly variable between platforms, spanning 0.017–4.4% of all mismatches observed. Moreover, for all platforms, the reported QVs were higher than the empirically derived QVs based on sequence mismatches, similar to the QV-inflation observed for DNA sequencing in the 1000 Genomes Project and GATK[36,37].

Previous work in RNA-seq has found that duplicate reads may be a confounding factor in data analysis because reads with exactly the same start and end may arise from clonal copies produced during library amplification rather than from independently transcribed RNAs in the biological sample[8,34]. However, unlike DNA sequencing of large diploid genomes, RNA-seq is expected to produce some reads from highly expressed transcripts that begin at the same nucleotide and are thus designated "duplicate." An assessment of this question

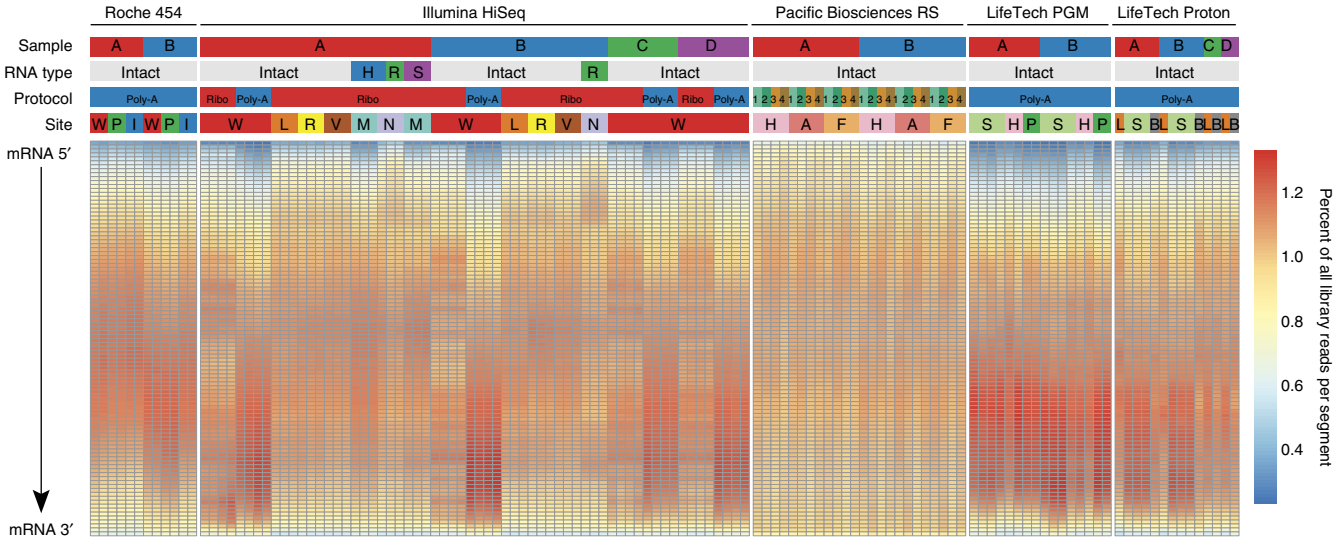

Figure 2 Transcript coverage across all genes detected. Each gene was examined as a set of 100 adjacent segments (percentiles of total transcript length). The relative number of reads that map to each segment was then plotted for each sample, platform and technique (percent of all library reads per segment, see heatmap color key). Samples are categorized by five parameters (top): NGS platforms: Roche 454 GS FLX+, Illumina HiSeq 2000/2500, Pacific Biosciences RS, Life Technologies PGM and Proton; input RNA sample: samples A–D; RNA type: intact or degraded by heat (H, blue), RNase (R, green) or sonication (S, purple); library protocol: poly-A enrichment (poly-A), ribosomal RNA depletion (ribo) or poly-A plus 5′ cap enrichment with (1: 1–2 kb, 2: 2–3 kb, 3: >3 kb) or (4) without size fractionation, see Online Methods for more details; and site: 14 core facility sequencing laboratories. Most platforms showed less coverage at the 5′ and 3′ ends of the transcripts. Details on sequencing platforms, site abbreviations, sample type chemistries, and library preparations are listed in **Table 2**.

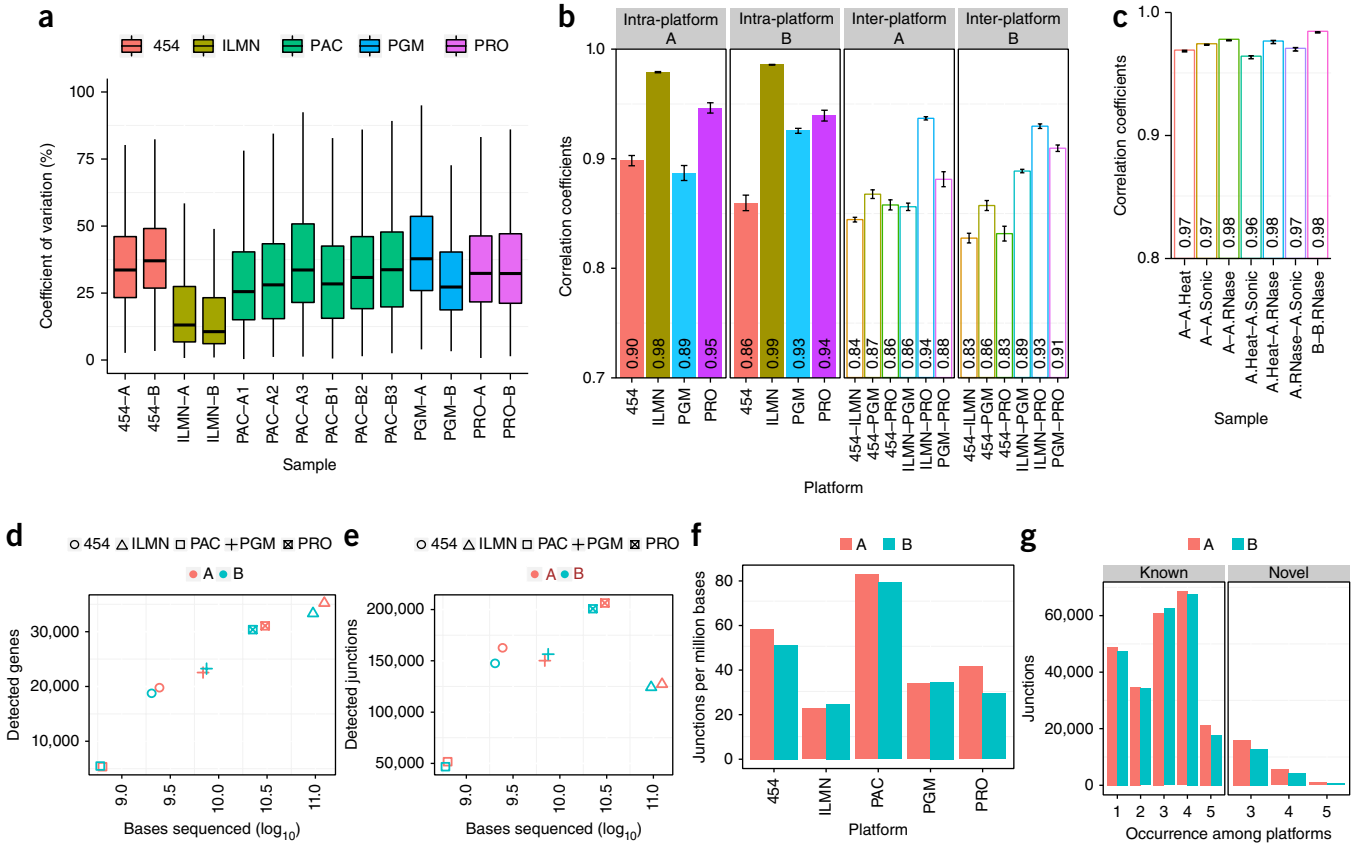

Figure 3 Intra- and inter-platform variation of RNA-seq transcript metrics. The coefficients of variation (CV) of various metrics for transcripts detected across all sites were calculated for the Roche 454 GS FLX+, Illumina HiSeq 2000/2500 (ILMN), Pacific Biosciences RS (PAC), and Life Technologies Personal Genome Machine (PGM) and Proton (PRO). (**a**) Inter-site CV of normalized gene expression for transcripts detected across all sites. The median CV for number of genes detected ranged from 10.70–38.68%, with many outlier genes present for each platform. (**b**) Inter-platform and intra-platform normalized gene expression Spearman correlation coefficients for samples A and B. Error bars, mean ± s.e.m. (**c**) The profiles of degraded RNA match the corresponding profiles of intact RNA on the HiSeq, with very high correlation coefficients (0.975). Error bars, mean ± s.e.m. (**d**,**e**) Sequenced bases were plotted against the number of detected genes or the number of detected splice junctions for known GENCODE junctions. (**f**) More efficient splice junction detection (*y* axis, number of junctions/Mb of sequence) was observed for long read platforms (PAC, 454). Detection efficiencies were calculated at comparable scales by constraining the total number of bases used from each platform to a range of 630–5,451 × 10^6. (**g**) Most known junctions were detected by three or more platforms, indicating concordance among RNA-seq methods (left panel). The novel junctions (right) defined by independent observation on three or more platforms were less numerous than known junctions.

over a range of read lengths has not been previously reported, but is facilitated in this study by RNA-seq of the same samples over a range of varying read lengths (**Supplementary Figs. 19–23**). The read length distributions revealed distinct types for variable-read platforms, including Gaussian (454) and "ski-jump" (Proton and PGM), and the expected uniform lengths for Illumina platforms. Yet, all platforms showed no more than 51% of reads as putative duplicates (**Supplementary Fig. 24**), with the 454 and PacBio platforms showing the fewest duplicates (12–20%). PacBio library construction does not include any amplification step of the final cDNA library, whereas the reduced duplication with 454 is likely because the amplification step takes place after template attachment to single beads, so individual molecules in the library have less chance to spawn multiple reads. For the other platforms, this analysis cannot distinguish whether observed duplicates are due to independent transcripts or are a consequence of library amplification, but future data sets based on these same samples will support investigation of this question.

Coverage of genes

Next we examined the normalized coverage of all GENCODE gene transcripts from 5′ to 3′ termini for any bias in the number of mapped bases originating from different regions of the transcripts. Almost all samples showed a fairly similar distribution of coverage for genes (**Fig. 2**). Notably, the ribo-depleted RNA samples, whether degraded or not, consistently showed more uniform gene coverage than did poly-A–selected libraries. The data also showed "banding" or altered coverage distributions, likely caused by the use of a different library kit version at one of the test sites (W). This indicates that gene coverage can be affected by platform and preparation-dependent factors, but aligners can also play a role (**Supplementary Table 3**). Finally, the highest and most uniform coverage of full-length transcripts came from preparing samples with enrichment for both the 3′ poly-A tail and an antibody for the 5′ methylguanylate cap (5′G cap), combined with long-read technology (see Online Methods for Pacific Biosciences).

Transcriptome profiling and splice junction detection

We investigated the ability of each platform to reproducibly detect and quantify genes and splice junctions across the transcriptome (**Fig. 3**). Data were restricted to genes that were observed at all test sites and in all technical replicates for each platform. The platforms showed a median range of 11–39% inter-site CV (coefficient of variation) in their quantification of detected genes using normalized gene expression

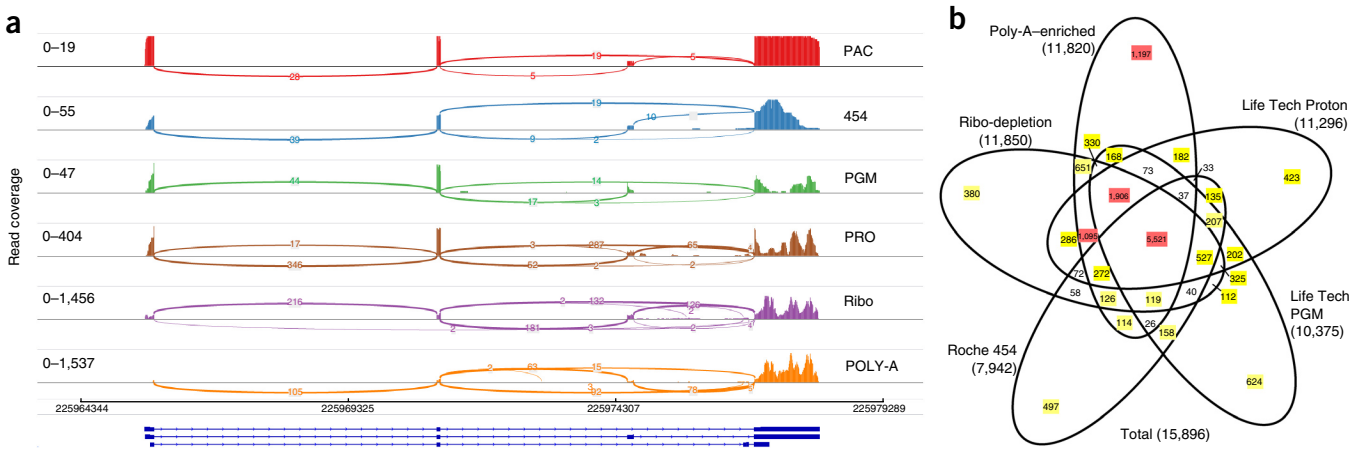

Figure 4 Inter-platform consistency of splicing and differential expression analysis. (**a**) As a representative plot for RNA splicing, transcripts from the SRP9 gene are shown in a sashimi plot across five platforms and two Illumina library protocols. Pacific Biosciences (PAC), Roche 454 (454), and Life Technologies Personal Genome Machine (PGM) detected the two most abundant isoforms. Life Technologies Proton (PRO) and Illumina ribo-depletion (Ribo) or poly-A–enriched (POLY-A) methods also detected a third isoform. PAC showed more uniform sequencing depth across the gene body. The number of reads observed for each junction is indicated within segments, and y-axis ranges for the number of reads per exon base are listed (read coverage, left). (**b**) Starting from the set of genes detected at any expression level on all platforms, the numbers of A versus B differentially expressed genes uniquely or repeatedly detected at statistically significant thresholds (FDR < 0.05 and fold change > 2) are shown; sets of greater than 1,000 genes are indicated in red, 100–999 in yellow.

values (**Fig. 3a** and Online Methods), with HiSeq showing the lowest median CV. The Spearman correlations of normalized transcript levels were measured for samples A and B on different platforms (454, HiSeq, Proton and PGM) across multiple sites for **Fig. 3b**; PacBio was not included because it displayed an (expected) low read count for many genes. The inter-platform correlation was high (correlation coefficient greater than 0.83) for the same samples profiled on different platforms, and the intra-platform correlation was even higher (correlation coefficient greater than 0.86). Each platform was also compared to normalized expression data from an orthogonal quantification method (PrimePCR, **Supplementary Fig. 25**), and the Spearman correlations of the $\log_2$ fold differences were ranked as 454 < PGM < Proton, HiSeq, ranging from 0.83 to 0.89.

Next we examined the impact of read depth and length on transcript identification. A clear log-linear relationship was observed between sequence base depth and gene detection (**Fig. 3d**), showing that increasing the depth of sequencing for any platform is a quick means to find more genes. Characterizations of splice junction detection efficiency and inter-platform agreement have not been previously reported, so to account for each platform's different read lengths, the effect of total sequenced base depth (rather than read count) was examined for previously annotated and new, unannotated splicing. Splice junction profiling showed an early plateau for detection of known junctions (**Fig. 3e**). The Proton, PGM and 454 platforms detected more known junctions despite fewer bases sequenced compared to Illumina HiSeq. However, a follow-up experiment with long-read Illumina MiSeq data (2 × 250 bp paired-end reads) showed a similar boost in junction identification (**Supplementary Fig. 26**), suggesting that splice junction detection is most affected by read length, rather than library preparation or sequencing chemistry. The ratio of the number of junctions detected as a function of total bases sequenced (junctions/Mb) revealed a wide range of values (**Fig. 3f**) but clearly demonstrated that longer reads are a more efficient way to capture junctions. This is reflected in the data from the long-read platforms and also in the comparison of the number of junctions detected in the Illumina HiSeq versus MiSeq data from two aliquots of the same library (22.6 junctions/Mb for HiSeq versus 33.9 junctions/Mb for MiSeq, **Supplementary Fig. 26**).

We also characterized the inter-platform agreement of known and novel junctions. The known GENCODE junctions (v12) showed higher inter-platform agreement, with most of these junctions detected by three or more platforms (**Fig. 3g**, left panel). However, unannotated junctions have lower concordance than known junctions across platforms (**Fig. 3g**). An examination of these rare isoforms revealed that the lower detection agreement is likely due to their lower expression levels (**Supplementary Fig. 27**), but they also may represent platform-specific artifacts. Therefore, only unannotated splice junctions observed on at least three platforms (which still includes >20,000 junctions per sample) are reported in this analysis.

These cross-platform splicing data showed that the types of reads dramatically influenced each platform's measure of low abundance transcripts. This effect was apparent for RNA splice isoforms such as SRP9 (**Fig. 4a**), suggesting that rare-isoform quantification benefits the most from greater read depths (such as from the Illumina HiSeq and Life Technologies Proton). However, uniformity of coverage across exons is improved with long-read technology such as PacBio (**Fig. 4a** and **Supplementary Table 3**), despite less read depth. An examination of the size-selected PacBio circular consensus sequencing libraries demonstrated that the poly-A+5′G cap enrichment method captured the full lengths of expressed transcripts (**Supplementary Fig. 28**), with the majority (90%) showing complete transcript sequences in the 1–2 kb range or even longer. These results indicate that a combination of appropriate sample preparation and long reads can readily create cDNA profiles that approach the full-length sequences of mRNAs from complex samples, underlining the utility of long read platforms, despite the lower read depths they may produce[38].

To examine the ability of each platform to detect differentially expressed genes (DEGs) (**Fig. 4b**, **Supplementary Figs. 29–31**), we used limma-voom[39] to perform DEG analysis on the normalized counts for each platform. Although a majority of DEGs were observed by two or more methods, each produced unique DEGs at all statistical significance and fold-change cutoffs (**Supplementary Figs. 30** and **31**). Thus, although high read–depth platforms showed greater DEG overlap, each platform produced unique subsets (from unique

systematic effects) of statistically significant DEGs (FDR < 0.05, fold change > 2, Online Methods), ranging from 6–11% of all called DEGs detected uniquely by a platform or preparation method (**Fig. 4b**, peripheral sets). These instruments span different chemistries, measurement techniques (optical versus electrical) and base-calling methods, all of which likely play roles in the system-specific noise profiles (**Figs. 1–3** and **Supplementary Figs. 1–24**).

Influence of library preparation on transcriptome profiles

To examine other factors that affect DEG measurements, we prepared libraries using either poly-A enrichment or ribosomal RNA depletion of the standard samples, and then performed sequencing on the same Illumina HiSeq 2500 instrument. Identical aliquots of the standards (A, B, C and D) were separated into quadruplicate sets for library preparation. All replicate libraries were then sequenced in a multiplexed assay on a full Illumina flow cell. The ribo-depletion library method produced a read source distribution very different from the poly-A preparation method (**Fig. 5a**). The ribo-depleted libraries showed 40–47% of the bases mapping to introns versus 7–12% for poly-A RNA from the same sample (lower intronic reads were similarly observed for poly-A RNA on the other platforms, **Supplementary Fig. 32**). Both methods produced fairly consistent measures of RNA abundance (fragments per kilobase of transcript per million mapped reads (FPKM), Online Methods), with a median FPKM difference of only 0.055 between all genes. However, more genes with lower levels of expression were observed with the ribo-depletion method,

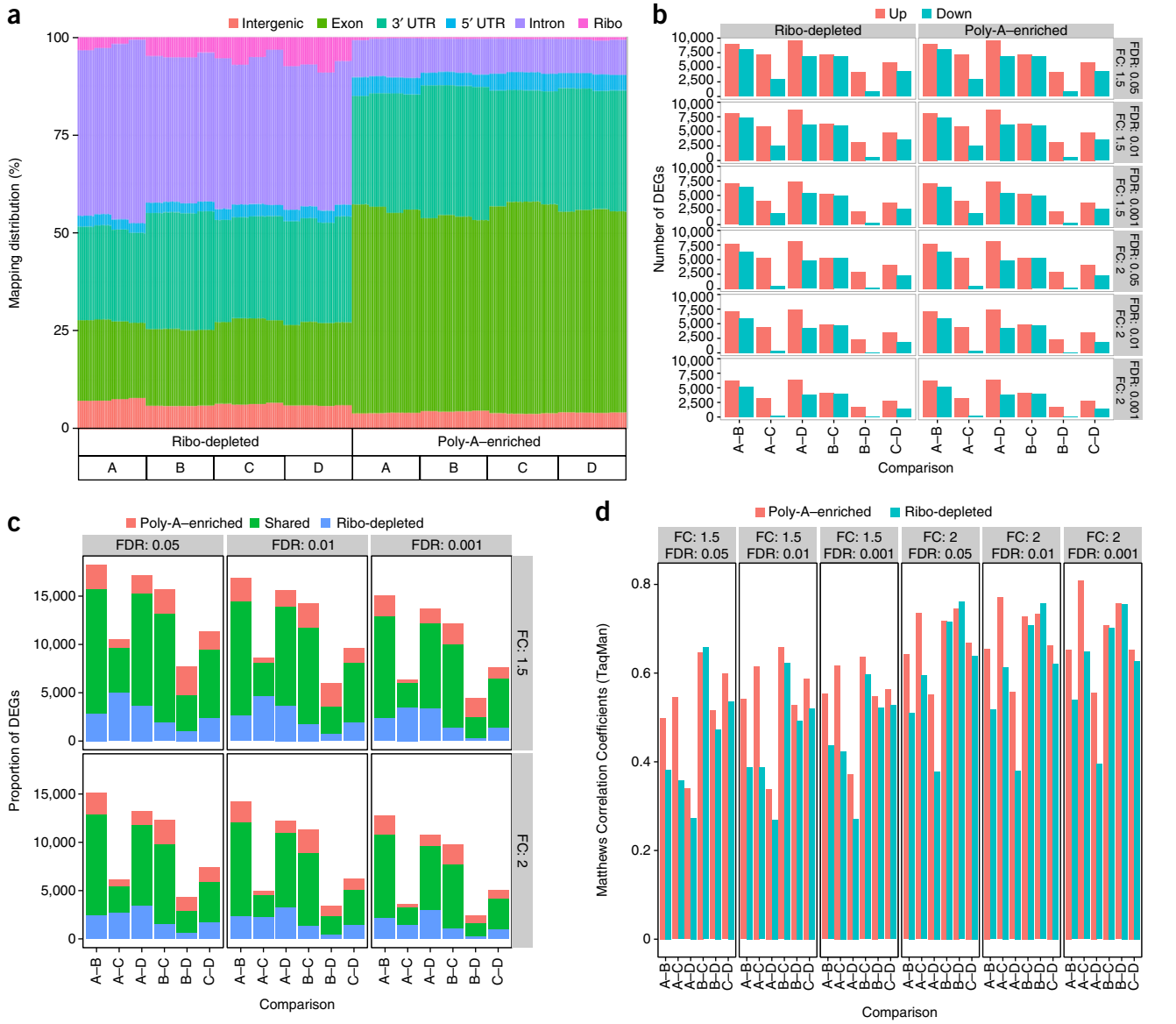

Figure 5 Differentially expressed genes in ribo-depleted and poly-A–enriched libraries. (**a**) The percentage of reads that map to various gene sequence categories. There were more intronic reads from ribo-depleted than poly-A–enriched libraries. The sequence type and read distribution of gene features detected in poly-A–enriched and ribo-depleted libraries from the same sample were examined using GENCODE (v12) annotations. Mitochondrial RNA reads are present at trace levels (<0.1%, data not shown). (**b**) Differentially expressed genes (DEGs) were detected in all pairwise comparisons of the original (A, B) and mixed samples (C, D); (**c**) results were similar for both library types from the common set of detected genes at all fold-change (FC) and false discovery rate (FDR) thresholds tested. (**d**) Both library types show similar accuracy as evidenced by Matthews Correlation Coefficients (MCC) with RT-qPCR assays (see **Supplementary Fig. 29b** for expanded data). A subset of GENCODE mapped reads was used from each library (mean = 37.6×10^6 reads per replicate, s.d. = 2.07×10^6) to ensure the same number of exon-mapped reads per sample was compared between all replicates.

whereas the poly-A libraries contained more highly expressed genes and 3′ untranslated regions (3′ UTR; **Supplementary Fig. 33**). As expected, the ribo-depleted libraries were enriched for noncoding RNAs, such as long noncoding RNAs and small nucleolar RNAs (**Supplementary Table 6**), whereas the poly-A libraries were enriched for protein-coding genes and mitochondrial genes (**Supplementary Tables 6–8**)[40]. Sequence annotations in GENCODE currently labeled as "intron" and other categories are likely to change as new noncoding RNAs (or new transcript classes) are identified.

However, few overall differences were observed between the poly-A and ribo-depleted library preparations in gene quantification and detection of differentially expressed genes. Both data sets were evaluated using alignments from STAR and DEG calculations from limma-voom[39], and surrogate variable analysis (sva) was applied for the detection of latent variables (Online Methods)[41]. A pairwise comparison of the average normalized gene expression across replicates of the two library types for the four standard samples showed high Spearman correlation coefficients (sample A: 0.91, B: 0.93, C: 0.92, D: 0.93). The overall numbers of DEGs detected between the biologically distinct samples (A versus B, A versus D, etc.) were also consistent between library preparation methods (**Fig. 5b**,**c**). These DEG data were then compared to results from 802 TaqMan assays for these same RNA samples (GEO data set GSE5350)[10]. Both library types had similar accuracy as measured by Matthews correlation coefficient (MCC, **Fig. 5d**)[42,43], which is a joint measure of the assay's sensitivity and specificity. The corresponding DEGs without sva analysis show similar but slightly lower overlap percentage and MCC (**Supplementary Fig. 33**). The median MCC is 0.659 before sva and 0.678 after sva, with an average increase of 0.015. Furthermore, the percentage of shared DEGs ranges from 67–81% at FDR < 0.01 and fold change > 2, and similarly ranges from 68–81% after sva. However, the synthetic RNAs spiked into these samples (ERCC controls) performed slightly better in the ribo-depletion protocol than the poly-A–enrichment protocol (mean correlation coefficient = 0.91 and 0.82, respectively), although these ranges of correlation to TaqMan were similar to those observed for ERCCs sequenced on the PGM, where the mean correlation coefficient = 0.78 (**Supplementary Figs. 34** and **35**).

Impact of RNA degradation on transcriptome profiling

As poly-A and ribo-depleted gene quantifications were similar, we sought to test the effect of ribo-depletion on 'low quality' or degraded RNAs. The reference samples A and B were degraded using heat, sonication or RNase-A until all samples showed a high level of degradation when evaluated on the Agilent Bioanlyzer 2100 (RIN < 3.0, Online Methods). Samples were ribo-depleted before library preparation and sequenced on the HiSeq platform at multiple sites. Multiple metrics indicated that the degraded RNA performed as well as the poly-A–enriched or ribo-depleted libraries from intact RNA. First, sequencing of the degraded RNA, after ribo-depletion, fully covered the gene bodies (**Fig. 2**) and, similar to ribo-depleted libraries from intact RNA, more reads mapped to intronic areas of the genome (**Supplementary Fig. 32**). Second, the degraded RNA showed minor differences in gene detection or DEG accuracy, with high Spearman correlation (correlation coefficient > 0.96, **Fig. 3c**) in expression comparisons to intact RNA samples. In addition, a comparison to the orthogonal PrimePCR data set showed that the degraded RNA analysis was highly correlated (Pearson correlation coefficient > 0.83) to the corresponding intact samples (**Supplementary Table 4**). However, the degraded RNA did have a lower Spearman rank-order correlation with quantitative PCR for the expression differences detected between samples A and B. The Spearman correlation was highest for heat degradation (correlation coefficient = 0.83, AH), followed by RNase A (correlation coefficient = 0.79, AR), and then sonication (correlation coefficient = 0.74, AS) (**Supplementary Fig. 36a–c**). Comparison of the results from one degraded sample to the results from one intact sample, repeated at multiple laboratories (sites L, V and R), also produced an overall high average Spearman correlation coefficient (0.80, **Supplementary Fig. 36d**). These data indicate that although appropriate library preparation of degraded RNA can produce accurate expression measurements (**Supplementary Fig. 36**), combining intact and degraded samples (or samples degraded during different types of tissue processing) within an experiment should be avoided.

DISCUSSION

This ABRF-NGS study represents, to our knowledge, the largest reported cross-platform, cross-protocol and cross-site examination of RNA-seq data performed to date. The results provide a unique opportunity to examine various aspects of the transcriptome, including the intra- and inter-site coefficients of variance of gene detection, gene expression quantification and RNA splicing between sequencing platforms, as well as the ability of long read lengths to enable complete isoform characterization. Comparisons of platform-specific aligners with STAR showed that mapping rates, error rates and transcript coverage are all larger concerns than is gene quantification when considering inter-platform data. As such, the use of different alignment algorithms will have different influences on comparisons between experiments depending on the metric studied, and the importance of 'bioinformatics noise' can be placed alongside technical and biological noise as key factors in experimental design. Finally, the results expanded previous work[27] by showing that gene detection and quantification with highly fragmented or degraded RNA samples (from three types of degradation) is similar to data from intact RNA, once ribosomal RNA is removed.

This study found similar RNA-seq results between the various NGS platforms and similar ranges in coefficients of variance across laboratory sites for each platform. These results indicate that both long- and short-read technologies measure gene expression with similar levels of statistical variation, although they show a tenfold variation for error rates in indels. Using normalized gene expression as a comparison measure, we found high intra-platform consistency (correlation coefficient > 0.86) and high inter-platform concordance (correlation coefficient > 0.83) measured by Spearman rank correlation (**Fig. 3b**). However, the results clearly show that deeper sequencing of the transcriptome is needed to reveal low abundance transcripts and splice junctions, indicating that read depth should be a key consideration when experimental goals include rarely expressed genes, coverage of introns and nonpolyadenylated targets. Very deep sampling is not currently cost-effective with long-read platforms such as PacBio or 454 (**Table 1**), and thus the best discovery platforms for low-abundance targets are currently the shorter read platforms, as they can cover a wider dynamic range of RNA molecules (i.e., generate more reads per sample).

Despite lower read depths and higher costs, the longer read NGS technologies have the best ability to efficiently catch the vast majority of known splice junctions (**Fig. 3e–g**), indicating that they can be an effective means to annotate splicing complexity that can reach as high as 10^{108} isoforms[44]. The ABRF-NGS study's results include a wealth of putative novel splice junctions, with more than one million such junctions observed on at least one platform. These putative novel splice junctions displayed greater inter-platform disparity than the known splice junctions (**Fig. 3e**). This difference was likely due to the challenge of correctly predicting novel isoforms and also to the possibly high false-positive rate of such predictions, which is expected

Table 1 Summary of sequencing platforms as deployed in the ABRF-NGS study of RNA-seq

Vendor	Instrument	Version	Run Time (hours)	Read length (mean)	Reads per run (millions)	Yield per run	Cost per run ($)[a]	Cost per Mb ($)	Paired-end	Application	Strengths
Illumina	HiSeq 2000	High Output	132	50	6,000	300 Gb[b]	18,725.00	0.06	Yes	Gene expression; splice junction detection; variant calling; fusions	Deep read counts for transcript quantification
Illumina	HiSeq 2500	High Output	132	50	6,000	300 Gb[b]	18,725.00	0.06	Yes	As above	As above
Illumina	MiSeq	v2 kit	39	250	30	7.5 Gb	982.75	0.13	Yes	Splice junction detection; variant calling	Rapid transcript quantification and variant detection
Life Technologies	PGM	318 chip	7.3	176	6	1.056 Gb	749.00	0.71	No	Splice junction detection; variant calling	Rapid transcript quantification and variant detection
Life Technologies	Proton	Proton I chip	2–4	81	70	5.67 Gb	834.00	0.15	No	Gene expression; variant calling	Good read depth and length for transcript quantification
Pacific Biosciences	RS	RS	0.5–2	1,289	0.03	38.67 Mb	136.38	3.53	No	Splice junction detection; full-length gene coverage	Extended read lengths
Roche	454	GS FLX+	20	686	1	686 Mb	5,985.00	8.72	No	Splice junction detection	Read length

[a]Sequencing reaction reagents, at academic list price; does not include library preparation reagents, labor, data storage or analysis, equipment or maintenance. [b]One sequencing run using two flowcells. Pacific Biosciences is calculated based on CCS (circular consensus sequencing) reads; Gb, billion bases, Mb, million bases.

given their lower expression levels. However, a substantial number of the previously unannotated, predicted junctions are likely genuine, as they were observed using multiple platforms. The resulting data sets nearly double the catalog of splice junctions for these RNA standard samples. The junctions discovered on multiple platforms can be used alone or with previous data for future algorithm design and assay optimization, and as positive controls, to advance splicing isoform characterization by RNA-seq[14,45–47].

Perhaps most notably, the data demonstrated that results from poly-A–enriched and ribo-depleted RNA libraries, and even libraries made from severely degraded RNAs, are comparable. Given sufficient depth of sequencing, results from ribo-depleted libraries can include almost all of the differentially detected genes identified by the poly-A preparation method, without loss of sensitivity or specificity. This was evident not only in the overlap of DEGs, but also in comparisons to TaqMan and PrimePCR data. Furthermore, a near-complete reconstruction of the transcriptome profile was observed when using degraded RNA in the ribo-depletion protocol, with some variation between degradation treatments, as judged by correlations to the expression abundances measured in intact samples A and B by quantitative PCR and by the uniform coverage of full transcript lengths. Similar degraded RNA results were recently reported[27], suggesting that low-quality samples can now be considered for reliable RNA-seq expression profiling. This should support studies using old, degraded or fragmented RNAs, such as those from formalin-fixed, paraffin-embedded (FFPE) tissues in clinical archives. Although the degraded RNA samples were run only on the HiSeq platform, the clear utility of such an approach should spur the development of similar degraded RNA resources for analyses on all sequencing platforms.

However, despite their overall similarities, distinct transcriptomes are represented in libraries prepared by poly-A enrichment, ribo-depletion or combined poly-A and 5′G cap enrichment. The dual enrichment method for PacBio libraries provided superior 5′ to 3′ coverage of the sequenced transcripts, as illustrated by comparisons across platforms for genes consistently detected by PacBio (**Supplementary Fig. 37**). The revised version of the Illumina library kit (v2 versus v3) includes built-in ribo-depletion and tags cDNA strand orientation, and the two protocols produced differences in gene-body coverage. A comparison of poly-A and ribo-depleted libraries showed different detection of nonpolyadenylated transcripts, 3′UTRs and introns. The former is an intentional consequence of the enrichment protocol, but it is not clear if the 3′UTR coverage bias is due to different efficiencies of priming during reverse transcription or to skewed sampling caused by a higher concentration of structural and noncoding RNAs in ribo-depleted libraries. Owing to the higher rate of intron-mapped reads, RNA-seq of total RNA will require greater read depths for ribo-depleted libraries (~2.5×) than for poly-A libraries to achieve equal coverage of exons. Transcriptome measurement variations demonstrated between the reference data sets are easily avoided by consistently using the same protocols, platform and analysis pipeline for all samples in an experiment. Nonetheless, if this is not possible, surrogate variable analysis enabled removal of latent variables from the data for ribo-depleted and poly-A–enriched libraries, producing nearly indistinguishable lists of DEGs and illustrating the utility of surrogate variable analysis as a powerful and strongly recommended method for ameliorating the effects of inter-batch and cross-protocol noise.

The results presented here also highlight additional variables that should be considered when aligning library protocols and platforms with research goals. The reported QV values of all platforms are all higher than empirically derived error rates, indicating that a splicing-aware, base quality score recalibration may be needed for RNA-seq, as is already done for DNA-seq with GATK. Long-read sequencing effectively cataloged splicing isoforms, but had less dynamic range for transcript quantification and discovery due to lower read depths. The use of the ERCCs is generally recommended as a good quality control metric, but these standards performed better in ribo-depleted libraries than in poly-A libraries, and this should also be considered during experimental design. In summary, the priorities for biological interpretation are essential when considering the protocols and methods that will be used in an RNA-seq experiment. Some of these priorities are summarized in **Table 1**, which provides a cross platform

Table 2 Sequencing platforms, chemistries and library preparations

Illumina Hiseq2000/2500 and MiSeq

Labs	Samples	Libraries per sample	Preparation	Read length	Number of reads per library	Output (Mb)
1(L)	MAQC A[a]	3	Ribo-depleted	100 bp (2 × 50)	386,726,967	38,673
	MAQC B[a]	3	Ribo-depleted	100 bp (2 × 50)	251,724,566	25,172
2(R)	MAQC A[a]	3	Ribo-depleted	100 bp (2 × 50)	229,131,233	22,913
	MAQC B[a]	3	Ribo-depleted	100 bp (2 × 50)	229,591,730	22,959
	MAQC B[a]	1	Ribo-depleted	500 bp (2 × 250)	7,848,217	3,924
3(V)	MAQC A[a]	3	Ribo-depleted	100 bp (2 × 50)	207,603,620	20,760
	MAQC B[a]	3	Ribo-depleted	100 bp (2 × 50)	239,930,780	23,993
4(N)	MAQC A[a,c]	2	Ribo-depleted	100 bp (2 × 50)	215,903,801	21,590
	MAQC B[a,c]	3	Ribo-depleted	100 bp (2 × 50)	219,257,190	21,926
	MAQC A[a,d]	1	Ribo-depleted	100 bp (2 × 50)	183,811,383	18,381
5(M)	MAQC A[a,d]	3	Ribo-depleted	100 bp (2 × 50)	386,726,967	38,673
	MAQC A[a,e]	3	Ribo-depleted	100 bp (2 × 50)	181,740,643	18,174
6(W)	MAQC A[b]	4	Ribo-depleted	100 bp (2 × 50)	128,133,887	12,813
	MAQC B[b]	4	Ribo-depleted	100 bp (2 × 50)	137,096,343	13,710
	MAQC C[b]	4	Ribo-depleted	100 bp (2 × 50)	142,135,538	14,214
	MAQC D[b]	4	Ribo-depleted	100 bp (2 × 50)	128,040,437	12,804
	MAQC A[b]	4	Poly-A–enriched	100 bp (2 × 50)	106,762,840	10,676
	MAQC B[b]	4	Poly-A–enriched	100 bp (2 × 50)	111,430,017	11,143
	MAQC C[b]	4	Poly-A–enriched	100 bp (2 × 50)	108,582,900	10,858
	MAQC D[b]	4	Poly-A–enriched	100 bp (2 × 50)	105,978,082	10,598

[a]Illumina RNA TruSeq v2 library kit. [b]Illumina RNA TruSeq v3 library kit. [c]RNaseA degraded. [d]Heat degraded. [e]Sonicated.

Life Technologies Ion Torrent PGM and Proton

Labs	Samples[f]	Libraries per sample	Preparation[g]	Median read length	Mean number of reads	Output (Mb)
1(P) Ion PGM	MAQC A	2	Poly-A–enriched	161	5,323,672	857
	MAQC B	2	Poly-A–enriched	184	5,802,563	107
	ERCC 1	1	Poly-A–enriched	189	4,188,385	792
	ERCC 2	1	Poly-A–enriched	158	3,231,475	511
	ERCC 1	1	Poly-A–enriched	180	4,442,093	800
	ERCC 2	1	Poly-A–enriched	189	4,310,663	815
2(H) Ion PGM	MAQC A	3	Poly-A–enriched	128	3,374,068	445
	MAQC B	3	Poly-A–enriched	129	3,409,662	436
	ERCC 1	2	Poly-A–enriched	152	2,538,594	810
	ERCC 2	2	Poly-A–enriched	112	2,119,884	514
	VL A	1	Poly-A–enriched	187	3,965,022	770
	VL B	1	Poly-A–enriched	162	4,138,326	687
3(S) Ion PGM	MAQC A	3	Poly-A–enriched	198	5,049,998	1,000
	MAQC B	3	Poly-A–enriched	199	5,743,028	1,140
	ERCC 1	1	Poly-A–enriched	206	6,835,287	1,410
	ERCC 2	2	Poly-A–enriched	207	7,119,023	1,480
	ERCC 1	1	Poly-A–enriched	182	6,525,478	1,190
	ERCC 2	1	Poly-A–enriched	191	5,490,495	1,050
4(S) Proton	MAQC A	3	Poly-A–enriched	78	50,063,784	3,900
	MAQC B	3	Poly-A–enriched	85	53,203,028	4,497
5(B) Proton	MAQC A	1	Poly-A–enriched	95	57,701,947	4,864
	MAQC B	1	Poly-A–enriched	75	39,099,605	2,946
	MAQC C	1	Poly-A–enriched	64	41,308,206	2,641
	MAQC D	1	Poly-A–enriched	53	46,665,851	3,160
6(L) Proton	MAQC A	1	Poly-A–enriched	99	60,106,614	5,978
	MAQC B	1	Poly-A–enriched	100	60,769,231	6,085
	MAQC C	1	Poly-A–enriched	107	60,353,696	6,454
	MAQC D	1	Poly-A–enriched	106	69,977,984	7,413

[f]ERCC: synthetic standards only (External RNA Control Consortium); VL: pilot data for sample A or B. [g]Ion Total RNA-Seq v2 library kit.

Pacific Biosciences RS

Labs	Samples	Libraries per sample: size fractionation	Preparation[h]	Avg. read length	Reads/Mb	Output (Mb)
1(A)	MAQC A	1: >3 kb	Poly-A + 5′G cap	3,983	251	663
	MAQC A	1: 2–3 kb	Poly-A + 5′G cap	3,513	284	520
	MAQC A	1: 1–2 kb	Poly-A + 5′G cap	2,811	356	780
	MAQC B	1: >3 kb	Poly-A + 5′G cap	3,467	288	536
	MAQC B	1: 2–3 kb	Poly-A + 5′G cap	3,223	310	459
	MAQC B	1: 1–2 kb	Poly-A + 5′G cap	3,112	321	638

(continued)

Table 2 (continued)

Pacific Biosciences RS

Labs	Samples	Libraries per sample: size fractionation	Preparation[h]	Avg. read length	Reads/Mb	Output (Mb)
2(F)	MAQC A	1: >3 kb	Poly-A + 5′G cap	3,472	288	634
	MAQC A	1: 2–3 kb	Poly-A + 5′G cap	3,644	274	555
	MAQC A	1: 1–2 kb	Poly-A + 5′G cap	2,792	358	927
	MAQC A	1: unfractionated	Poly-A + 5′G cap	2,832	353	260
	MAQC B	1: >3 kb	Poly-A + 5′G cap	3,578	280	667
	MAQC B	1: 2–3 kb	Poly-A + 5′G cap	3,523	284	594
	MAQC B	1: 1–2 kb	Poly-A + 5′G cap	2,844	351	991
	MAQC B	1: unfractionated	Poly-A + 5′G cap	2,814	355	251
3(H)	MAQC A	1: >3 kb	Poly-A + 5′G cap	3,201	312	528
	MAQC A	1: 2–3 kb	Poly-A + 5′G cap	3,135	319	344
	MAQC A	1: 1–2 kb	Poly-A + 5′G cap	2,761	362	767
	MAQC A	1: unfractionated	Poly-A + 5′G cap	2,998	334	477
	MAQC B	1: >3 kb	Poly-A + 5′G cap	3,189	314	572
	MAQC B	1: 2–3 kb	Poly-A + 5′G cap	2,952	339	383
	MAQC B	1: 1–2 kb	Poly-A + 5′G cap	2,779	360	660
	MAQC B	1: unfractionated	Poly-A + 5′G cap	3,069	326	395

[h]PacBio Large Insert Template Prep Kit.

Roche 454 FLX

Labs	Samples	Libraries per sample	Preparation[i]	Median read length	Total reads per picotiter plate	Output (Mb)
1(I)	MAQC A	1	Poly-A–enriched	520	1,061,320	552
	MAQC B	1	Poly-A–enriched	494	1,001,678	495
	MAQC A	1	Poly-A–enriched	497	805,399	400
	MAQC B	1	Poly-A–enriched	496	1,076,634	534
2(P)	MAQC A	1	Poly-A–enriched	455	832,580	379
	MAQC B	1	Poly-A–enriched	470	1,181,610	555
3(C)	MAQC A	2	Poly-A–enriched	505	1,294,497	654
	MAQC B	2	Poly-A–enriched	358	293,471	105

[i]Roche cDNA Rapid Library kit.

summary of the strengths and relative costs of the sequencing technologies included in this study.

The ABRF-NGS study is not intended to be a 'bake-off' between NGS platforms, but rather is an effort to establish a useful reference data set for each platform, which will assist laboratories in improving their methods and in evaluating new chemistries, protocols and instruments. It is encouraging that comparison of gene expression quantification, including results from intra-platform, inter-platform, inter-protocol and even inter-aligner comparisons, demonstrated high correlations overall. This result suggests broader inter-study analyses and data mining can be successfully carried out across multiple platforms despite intrinsic differences between technologies, methods and aligners.

Reference data resources, such as the results from this ABRF-NGS study, are key to understanding the effects of variable sample quality, changes to platform protocols and the adoption of new technologies. Given the rapid pace of advancement in sequencing technologies, techniques and bioinformatics tools, the methods and data described here can facilitate the development of best practices for gene quantification, isoform characterization, dynamic range comparisons, managing inter-site and intra-site variation, analysis pipeline refinement and cross-platform testing of transcriptome hypotheses. These data can also be used to address other aspects of RNA-seq, including polymorphism detection, allele-specific expression, intron retention, RNA editing and gene fusions, and provide an immediately useful resource that can complement current databases, such as the RNA-Seq Atlas[48]. These and other applications, especially clinical molecular diagnostics that rely on nucleic acid biomarkers, will require a level of technical stability across time and both within and between studies, which this study helps to establish. Reference data resources are key to monitoring platform stability, and widespread adoption of standard samples and routine reference benchmarking are challenges that must be addressed to further advance genomics technologies.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: GSE46876.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We greatly appreciate the contribution and distribution of reference sample RNA from L. Shi (FDA) and his valuable interactions to assist in the planning of this study. This work was supported with funding from the National Institutes of Health (NIH), including R01HG006798, R01NS076465, R24RR032341, as well as funds from the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts and the STARR Consortium (I7-A765).

We thank the following contributors for their technical wisdom, including laboratory expertise, data analysis and bioinformatics contributions, and technical design guidance and consultation. Without their help, this study would not have been possible: D. Stopka (Memorial Sloan-Kettering Cancer Institute), G. Grove (Penn State Univ.), D. Hannon (Penn State Univ.), K. Jones (NIH/NCI/SAIC), C. Raley (NIH/NCI/SAIC), H. O'Geen (UC Davis), D. Zheng (Univ. Illinois-Urbana), O. Nguyen (UC Davis), Z.-W. Lu (UC Davis), J. Spisak (Cornell Univ.), D. Lin (NIH/NIAID), J. Pillardy (Cornell Univ.), P.-Y. Wu (Georgia Institute of Technology), J. Phan (Emory Univ.), D. Oschwald (New York Genome Center), H. Arnold (PerkinElmer), S. Tyndale (Univ. Southern California), H. Truong (Univ. Southern California), Y. Zhang (Univ. Florida), N. Panayotova (Univ. Florida), D. Moraga (Univ. Florida), S. Shanker (Univ. Florida), and N. Barker (US Army Environmental Quality Research Program).

We would also like to thank the platform vendors, Illumina, Life Technologies, Pacific Biosciences and Roche Life Sciences, for their support of this study, and their distinguished scientists for providing technical expertise and assistance in study designs, protocols, new methods development and significant contributions of reagents and sequencing kits. In particular, alphabetically by vendor: G. Schroth (Illumina); M. Gallad, J. Smith, T. Bittick, R. Setterquist and G. Scott (Life Technologies); J. Korlach, S. Turner and E. Tseng (Pacific Biosciences); and K. Fredrickson and C. Teiling (Roche Life Sciences).

We are sincerely appreciative of the Association of Biomolecular Resource Facilities (ABRF) for supporting this study and the contributing ABRF Research Groups. Special thanks to our ABRF executive board liaison A. Perera (Stowers Institute for Medical Research).

AUTHOR CONTRIBUTIONS

All authors are members of the Association of Biomolecular Resource Facilities Next-Generation Sequencing (ABRF-NGS) Consortium. S.W.T., C.M.N., D.A.B., G.S.G. and C.E.M. managed the project. S.W.T., C.M.N., D.G., S.L., W.F., A.V., C.W., P.A.S., Y.G., D.K., J.B., B.H., R.K., N.J., N.R., J.G., N.G.-R., C.H., D.R., J.R., T.S., J.G.U., C.E.M. and P.Z. performed sequencing. S.L., S.W.T., C.M.N., D.A.B., G.S.G. and C.E.M. designed the analyses. S.L., P.A.S., J.G.U., P.Z., C.E.M. and D.K. performed the data analyses. S.L., P.Z., M.W., D.K., J.G.U. and C.E.M. made the figures. S.L., S.W.T., C.M.N., D.A.B., G.S.G. and C.E.M. wrote and revised the manuscript. The ABRF-NGS Consortium members contributed to the design and execution of the study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Wang, E.T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
2. Nagalakshmi, U., Waern, K. & Snyder, M. RNA-Seq: a method for comprehensive transcriptome analysis. *Curr. Protoc. Mol. Biol.* **89**, 4.11 (2010).
3. Liu, S., Lin, L., Jiang, P., Wang, D. & Xing, Y. A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Res.* **39**, 578–588 (2011).
4. Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M. & Gilad, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* **18**, 1509–1517 (2008).
5. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
6. Liu, L. *et al.* Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364 (2012).
7. Ratan, A. *et al.* Comparison of sequencing platforms for single nucleotide variant calls in a human sample. *PLoS ONE* **8**, e55089 (2013).
8. Quail, M.A. *et al.* A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341 (2012).
9. Loman, N.J. *et al.* Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439 (2012).
10. Shi, L. *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.* **24**, 1151–1161 (2006).
11. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat. Biotechnol.* doi:10.1038/nbt.2957 (24 August 2014).
12. 't Hoen, P.A. *et al.* Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat. Biotechnol.* **31**, 1015–1022 (2013).
13. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* **21**, 2213–2223 (2011).
14. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
15. Łabaj, P.P. *et al.* Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics* **27**, i383–i391 (2011).
16. McIntyre, L.M. *et al.* RNA-seq: technical variability and sampling. *BMC Genomics* **12**, 293 (2011).
17. Huang, R. *et al.* An RNA-Seq strategy to detect the complete coding and non-coding transcriptome including full-length imprinted macro ncRNAs. *PLoS ONE* **6**, e27288 (2011).
18. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
19. Toung, J.M., Morley, M., Li, M. & Cheung, V.G. RNA-sequence analysis of human B-cells. *Genome Res.* **21**, 991–998 (2011).
20. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325–2329 (2011).
21. Angeletti, R.H. *et al.* Research technologies: fulfilling the promise. *FASEB J.* **13**, 595–601 (1999).
22. Moelans, C.B., Oostenrijk, D., Moons, M.J. & van Diest, P.J. Formaldehyde substitute fixatives: effects on nucleic acid preservation. *J. Clin. Pathol.* **64**, 960–967 (2011).
23. Opitz, L. *et al.* Impact of RNA degradation on gene expression profiling. *BMC Med. Genomics* **3**, 36 (2010).
24. Morlan, J.D., Qu, K. & Sinicropi, D.V. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS ONE* **7**, e42882 (2012).
25. Li, S. *et al.* Detecting and correcting systematic variation in large-scale RNA sequencing data. *Nat. Biotechnol.* doi:10.1038/nbt.3000 (24 August 2014).
26. Pareek, C.S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *J. Appl. Genet.* **52**, 413–435 (2011).
27. Adiconis, X. *et al.* Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat. Methods* **10**, 623–629 (2013).
28. Boland, J.F. *et al.* The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum. Genet.* **132**, 1153–1163 (2013).
29. Glenn, T.C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **11**, 759–769 (2011).
30. Jiang, L. *et al.* Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551 (2011).
31. Zook, J.M., Samarov, D., McDaniel, J., Sen, S.K. & Salit, M. Synthetic spike-in standards improve run-specific systematic error analysis for DNA and RNA sequencing. *PLoS ONE* **7**, e41356 (2012).
32. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
33. Hansen, K.D., Brenner, S.E. & Dudoit, S. Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res.* **38**, e131 (2010).
34. Aird, D. *et al.* Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.* **12**, R18 (2011).
35. Risso, D., Schwartz, K., Sherlock, G. & Dudoit, S. GC-content normalization for RNA-Seq data. *BMC Bioinformatics* **12**, 480 (2011).
36. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
37. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
38. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* **31**, 1009–1014 (2013).
39. Smyth, G.K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S.) 397–420 (Springer New York, 2005).
40. Cui, P. *et al.* A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing. *Genomics* **96**, 259–265 (2010).
41. Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. & Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
42. Baldi, P., Brunak, S., Chauvin, Y., Andersen, C.A. & Nielsen, H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics* **16**, 412–424 (2000).
43. Shi, L. *et al.* The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat. Biotechnol.* **28**, 827–838 (2010).
44. Li, S. & Mason, C.E. The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.* doi:10.1146/annurev-genom-090413-025405 (2 June 2014).
45. Haas, B.J. & Zody, M.C. Advancing RNA-Seq analysis. *Nat. Biotechnol.* **28**, 421–423 (2010).
46. Wenger, Y. & Galliot, B. RNAseq versus genome-predicted transcriptomes: a large population of novel transcripts identified in an Illumina-454 Hydra transcriptome. *BMC Genomics* **14**, 204 (2013).
47. Pipes, L. *et al.* The non-human primate reference transcriptome resource (NHPRTR) for comparative functional genomics. *Nucleic Acids Res.* **41**, D906–D914 (2013).
48. Krupp, M. *et al.* RNA-Seq Atlas–a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics* **28**, 1184–1185 (2012).

ONLINE METHODS

RNA used in the ABRF-NGS study. Standardized, commercial RNAs were sent to multiple sites for RNA-seq library preparation using different methods. Data were generated on five NGS platforms: Illumina HiSeq 2000/2500, Roche 454 GS FLX+, Life Technologies Ion Personal Genome Machine (PGM) and Proton, and the Pacific Biosciences (PacBio) RS. The HiSeq 2500 was used for the libraries from site W; all other Illumina libraries were sequenced on a HiSeq 2000.

Universal Human Reference RNA (740000, Agilent Technologies) and Ambion FirstChoice Human Brain Reference RNA (AM6000, Life Technologies) were used as the primary input RNAs for this study. These samples were labeled as MAQC samples A and B, respectively, in the MicroArray Quality Control (MAQC) experiments initiated in 2005 (ref. 9). The A and B naming convention is maintained here. These RNA samples were selected because they are well-characterized and have been used for many benchmarking studies, including a concurrent complementary RNA-seq study led by the US Food and Drug Administration[11].

ERCC spike-in synthetic transcripts were added at manufacturer recommended amounts (4456739, Life Technologies) to A and B standards. These RNAs, also developed for the 2006 MAQC study, consist of different ratios of artificially generated, poly-adenylated RNA transcripts of various lengths and combined in differing known concentrations[30]. Analyzing the ratios of these synthetic transcripts following library construction enables detection of sample preparation and platform-based biases.

The quality of the RNA samples was assessed before distribution to the participating laboratories, using the Agilent Bioanalyzer 2100, Nanodrop ND-1000 Spectrophotometer (Thermo Scientific) and fluorometry. All shipments were on dry ice. The samples distributed to the participating core laboratories and the libraries produced from these samples are summarized in **Table 2**.

PrimePCR RT-qPCR. PrimePCR RT-qPCR reactions were run in 384-well plates according to the manufacturer's instructions (Bio-Rad). In short, 5 µl reactions contained 1× final SsoAdvanced SYBR Green Supermix (Bio-Rad), 1× final PrimePCR assay components and 25 ng of cDNA, and were run in a CFX384 Touch real-time PCR detection system (Bio-Rad) using standard cycling parameters. Quantification cycle (Cq) value determination was done using CFX Manager software (Bio-Rad) with autocalculated baseline and fixed threshold settings (300 relative fluorescence units). A Cq value of 35 corresponds to a single molecule of cDNA input (see below).

2 µg of each MAQC RNA sample (MAQC A and B, and 1:3 mixture samples MAQC C and D) was reverse transcribed using the iScript advanced cDNA synthesis kit for RT-qPCR (Bio-Rad) in a 20 µl reaction. Prior to reverse transcription, MAQC B RNA was DNase treated using the Heat&Run gDNA removal kit according to the manufacturer's instructions (ArcticZymes). Absence of gDNA contamination in both MAQC A and DNase treated MAQC B was verified by qPCR using 25 ng of RNA as input and DNA-specific assays[49]. MAQC samples A, B, C and D were measured in parallel in the same 384-well plate, each time for 96 different assays ($n = 1$) (according to the sample maximization run layout strategy as described[50]).

All PrimePCR assays have been extensively wet-lab validated (see Bio-Rad tech note 6262 for more details on PrimePCR assay validation and performance characteristics). In accordance with the MIQE guidelines[51], assays were evaluated for specificity, efficiency, linear dynamic range and background signal in negative controls. At least ten qPCR reactions were done for each assay: cDNA from reference RNA, no-template control, gDNA and seven points from a tenfold dilution series of synthetic templates (from 20 million to 20 copies). Amplification efficiencies were calculated from the results of the dilution series. Only assays that displayed good linear performance in the 20 to 20 million copy number range ($R^2 > 0.99$) with efficiencies of 90–110% were considered to be of sufficient quality. The average y-intercept of the standard curves was 35, indicating that a single template molecule results in a Cq value of 35 when amplified with SsoAdvanced SYBR Green Supermix in a qPCR reaction.

Data analysis and bioinformatics protocols. All data, with analysis methods, are freely available at the GEO, GSE46876. Additional study materials, code, scripts and methods are available at: http://physiology.med.cornell.edu/faculty/mason/lab/data3/sac2026/ABRF/index.html. Code and scripts are also in **Supplementary Software**.

Sequence data preprocessing. Sequences were aligned to the hg19 genome assembly (GRCh37). Read counts were calculated using the Rmake pipeline (http://physiology.med.cornell.edu/faculty/mason/lab/r-make/) with GENCODE v12 annotation. Read-level quantification for genes is achieved using Boost Interval Container Library split_interval_map and a reference transcriptome in BED format following the union mode (illustrated in http://www-huber.embl.de/users/anders/HTSeq/doc/count.html). For gene expression analysis, sequences from HiSeq reads were aligned with STAR (https://code.google.com/p/rna-star/; parameters see **Supplementary Table 5**; v2.2.1d) and with ELAND (http://www.illumina.com/software/genome_analyzer_software.ilmn; ELAND_standalone.pl -if $input.R1.fastq.gz -if input.R2.fastq.gz -ref hg19/–bam -it FASTQ -od . -op output -rt -l output.log; casava 1.8.2), PGM and Proton reads were aligned with TMAP (https://github.com/iontorrent/TMAP; command line: tmap map all -f hg19.fa -r <(zcat input.fastq.gz) -i fastq -Y -a 0 -o 1 -g 0 -n 5 -s output.bam stage1 map1 map2 map3; tmap.3.0.1), PacBio reads with GMAP (http://research-pub.gene.com/gmap/; gmap -D gmap_db/ -d hg19 -t 24 -f samse -n 0 input.fastq.gz; version 2013-10-04), ILMN reads with ELAND (http://www.illumina.com/software/genome_analyzer_software.ilmn; ELAND_standalone.pl -if $input.R1.fastq.gz -if input.R2.fastq.gz -ref hg19/–bam -it FASTQ -od . -op output -rt -l output.log; casava 1.8.2), and 454 reads with GS Reference Mapper (http://www.454.com/products/analysis-software/; GUI with "cDNA" settings). Only uniquely mapped reads were used for gene expression quantification. RNA expression levels were calculated as reads per kilobase of transcript per million mapped reads (RPKM) or, for paired-end sequencing, fragments per kilobase of transcript per million mapped reads (FPKM). Splice junction detections were generated by STAR RNA-seq aligner (parameters see **Supplementary Table 5**; v2.3.0e for 454, PGM, Proton, PacBio). Total junctions from 454, PGM, Proton, PacBio and HiSeq (ribo-depleted RNA and Illumina v2 kits from sites L, R, V) were used for comparison. Junction detection efficiency comparisons were normalized for read depth by using all PacBio data and subsets of data from other platforms (454: site I, HiSeq: site L-replicate1-Lane 5, PGM: site S-replicates 1-3, Proton: site S-replicate1). The resulting number of bases per platform used for this calculation ranged from 630 million to 5.451 billion bases.

RNA-seq differential gene expression analysis. The raw read counts were normalized by the trimmed mean of the M-values normalization method, which uses a weighted trimmed mean of the log expression ratios[52–54]. The mean-variance relationship of the counts was estimated, and the appropriate weights for each observation were computed based on their predicted variance, using voom from the limma package[39]. By applying the lmFit(), contrasts.fit() and eBayes() functions, also from the limma package, the $\log_2$ fold differences and standard errors were estimated by fitting a linear model for each gene, and empirical Bayes smoothing was applied for the standard errors. Benjamini-Hochberg adjustment for multiple hypothesis testing was applied at a variety of false discovery rates (FDR = 0.05 or 0.01 or 0.001). Differentially expressed genes were evaluated at $\log_2$ fold change (FC) cutoffs (FC = 1.5 or 2). Data from HiSeq site W, which used both ribo-depletion and poly-A library methods, were used for the comparison of different protocols from Illumina. Other platforms' data from 454 (sites C, P, I), PGM (sites H, S, P) and Proton (sites B, S) were used in the same fashion for cross-platform comparisons.

Expression level CV calculations. The inter-site coefficients of variation for normalized gene expression levels were calculated on the matrix of the same sample with the same platforms from all test sites for each gene. Only genes detected by all replicates for each platform were used for CV calculations.

Surrogate variable analysis. Normalized gene expression values in $\log_2$ scale were used to detect latent variables using the sva package[41]. Using the twostepsva.build() function based on the two-step algorithm of Leek and Storey[55], three latent variables were constructed. Latent variables were

doi:10.1038/nbt.2972

removed in the DEG analysis by adding each latent variable to the design matrix for limma pipeline[39].

RNA-seq quality metrics definitions. (i) Sequencing depth: total number of reads sequenced. (ii) Mapping rate: percentage of reads which mapped uniquely to the reference genome. (iii) Sequence directionality: the number of reads which mapped to the forward and reverse strands compared to those of the AceView gene model. (iv) Nucleotide composition: the total number of A/G/C/T sequenced at each position across the length of the read. (v) Guanine-cytosine (GC) distribution: the number of reads with a particular %GC content. (vi) Read distribution: the fraction of the reads which mapped to either exons, 3′UTRs, 5′UTRs, introns or intergenic regions (or the intersection of any of the categories) as defined by the AceView gene models. (vii) Coverage uniformity: the number of reads covering each nucleotide position of all genes scaled to 100 nt. (viii) Base quality scores (QV): Phred-quality scores as calculated by Illumina's HiSeq Control Software for each nucleotide position across all reads. (ix) Duplication rate: the number of reads with exactly the same sequence content. See Wang *et al.*[56] for a more thorough description of metrics. (x) Mismatch rate calculation: total number of mismatches was calculated by parsing and summing up the number in NM tag in bam files; total bases of indel mismatches comes from cigar field by parsing and summing up the number in front of "I" or "D." The total number of single-base mismatches was the total number of mismatches subtracted by total number of bases of indel mismatches. The mismatch rates were calculated using the number of mismatches divided by the number of mapped bases from cigar field by SAMtools bam check. 454 GSRM alignment's cigar field and PAC GMAP alignment used SAMtools calmd to calculate the MD tag before mismatch rates calculation. PGM and PRO calculations used TMAP alignment without softclipping.

TaqMan gene expression analysis. TaqMan data for samples A, B, C and D were obtained through the Gene Expression Omnibus database (accession number GSE5350)[57]. Data for four replicates of each sample were analyzed. Undetectable C_T values ($C_T > 35$ or $C_T = 0$) were removed before normalization. The data were normalized using the HTqPCR package[58] to the average C_T of POLR2A (lowest s.d. of C_T value) by subtracting the average C_T of POLR2A from each TaqMan target to give the $\log_2$ difference between endogenous control and target genes. TaqMan differential gene analysis was performed as for RNA-seq data, without the trimmed mean and voom transformations.

PrimePCR RT-qPCR gene expression analysis. Undetectable Cq values (Cq > 35 or Cq = 0) were removed from data for samples A, B, C and D. The s.d. of the Cq values for each gene were calculated, and the gene MYSM1 exhibited the lowest s.d. The data were normalized by subtracting the average Cq of MYSM1 from each PrimePCR target to give the $\log_2$ difference between the endogenous control and the target genes. The normalized Cq values were then used to calculate the R^2 correlation to the RNA-seq data using lm() from the R stats package and summary function from the R base package.

Comparison between RNA-seq and TaqMan data. DEGs were validated from the Illumina RNA-seq data from each site, for six comparisons (A-B, A-C, A-D, B-C, B-D, C-D), using the DEGs from the TaqMan data. MCC[42,43] was used to compare performance metrics for external validations. Each DEG from the RNA-seq data was predicted based on its adjusted *P* value and its fold difference. The determination of truth in the performances metric analysis was the detection of DEGs by the TaqMan data. Here, the true-positive rate was the probability of a positive DEG result from RNA-seq given that TaqMan called the same gene differentially expressed, and the false-positive rate is the probability of a positive DEG result from RNA-seq given that TaqMan did not call the gene as differentially expressed. MCC was calculated to measure how accurately RNA-seq can distinguish between DEGs and non-DEGs.

Platform-specific protocols. RNA samples were evaluated for quality before distribution to the participating laboratories using the Agilent Bioanalyzer 2100 (Agilent) (**Supplementary Fig. 38**), Nanodrop ND-1000 Spectrophotometer (Thermo Scientific), and fluorometry. All shipments were on dry ice. The study coordinated shipments of library preparation and sequencing reagents so that all laboratory sites received reagents from similar manufacturing lots as determined by the vendor.

Illumina HiSeq 2000/2500 and MiSeq. *Starting material and enrichment.* TruSeq libraries were synthesized from 2 µg of intact (RIN > 7) and degraded (RIN < 3) MAQC A and MAQC B RNA. Three replicate libraries of intact A and B were prepared at each of three core laboratory sites. Degraded RNA libraries were prepared at two additional core laboratory sites as listed in **Table 2**. Ribosomal RNA-depletion material was generated using the Ribo-Zero Gold system (Epicentre Biotechnologies) according to the manufacturer's instructions. The preparation of three types of artificially degraded MAQC A and B RNA was performed at a single core laboratory site, using 75 µg of each RNA at a concentration of 1 µg/µl. Three techniques were used: (i) heat treatment in deionized water at 99 °C for 10 min; (ii) exposure to 1 ng/µl RNase A for a sufficient time period to result in a RIN of 3, with the RNase then neutralized with 10 µl RNase Inhibitor (RiboLock EO0381, Thermo Scientific); and (iii) sonication within a Covaris S2 MicroVial for 6 × 55 s at 5% DC, intensity 3 and 100 c/b. All resulting RNA sample degradations (i.e., RIN values) were analyzed using the Agilent 2100 Bioanalyzer (**Supplementary Fig. 39**).

Library construction and sequencing. Following ribo-depletion, all recovered RNA was processed using the Illumina TruSeq RNA Sample Preparation Kit v2 protocol at the "elute-fragment-prime" step, followed by the standard TruSeq protocol. Completed libraries were evaluated by DNA quantification and Bioanalyzer analysis, and then submitted to a single core laboratory site for sequencing. Sequencing libraries were constructed with barcodes to allow multiplexing of 12 samples per lane, pooled to target 200 million clusters per channel and 100 million reads per library, and distributed over multiple channels of three flow cells to normalize for lane and run variability. Sequencing was carried out on Illumina HiSeq 2000 and 2500 instruments using protocols HCS 1.5.15.1 and RTA 1.13.48 and paired-end 50-bp reads. One of the replicate libraries for intact MAQC B was also sequenced on a MiSeq instrument using the 4 nM protocol (v2.2.0.2) targeting 15 million clusters with paired-end 250-bp reads. The recently released TruSeq RNA Sample Preparation Kit v3 differs from v2 by including ribosomal RNA depletion and preserves cDNA strand orientation. Lab site W compared poly-A enriched and ribo-depleted libraries constructed using the v3 kit.

Life Technologies Ion Torrent PGM. *Starting material and enrichment.* Four different Ion Torrent PGM libraries were constructed at three core laboratory sites using the MAQC A, MAQC B, ERCC 1 and ERCC 2 RNAs. Five micrograms of each MAQC RNA was enriched for poly-A RNA (MRRK1010, MPG Kit, PureBiotech) using the recommended Life Technologies Ion protocol for Transcriptome Profiling of Low-Input RNA Samples (April 2011). The MPG-Streptavidin was prepared from 100 µg (10 µl) of the complex and resuspended in 5 µl of Release Solution instead of 20 µl. This process was repeated as a second round of enrichment to further deplete rRNA from the samples. The resulting RNA was assessed for yield and purity using an Agilent 2100 Bioanalyzer PicoChip. No enrichments were performed for the ERCC samples.

Library construction. Whole transcriptome library preparation was performed using 5–10 ng of fragmented enriched poly-A RNA according to the manufacturer's protocol (Ion Total RNA-Seq Kit V2 protocol #4476286B Life Technologies). Size selection of a 315-bp product was performed using a standard Pippin prep protocol (Sage Science) followed by purification with AMPure beads (Beckman-Coulter Genomics). The synthetic ERCC 1 and ERCC 2 control RNA library construction was performed directly from 30 ng of the non-poly-A–enriched sample.

Template preparation and sequencing. Emulsion PCR was performed using the One Touch system (Life Technologies). Beads were prepared from 70–100 million copies using the One Touch 200 Template Kit v2 #4471263. Some libraries were prepared for 70–100 million copies and others using the standard 210 million copies as stated in the RNA-seq protocol #4476286B. Sequencing

was conducted using an Ion PGM 200 sequencing kit (#4474004) on the 318 Ion chip. Data were collected using the Torrent Suite v3.0 software.

Life Technologies Proton. *Starting material, enrichment and library construction*. Libraries were prepared from 1 µg of MAQC A, B, C and D RNA containing ERCC controls using either poly-A enrichment (as described for PGM). After enrichment, 8–9 ng of poly-A mRNA was used for ligation reactions. The size selection step was adjusted to 220 bp using a standard Pippin Prep protocol, generating library competent molecules with a template insert size of approximately 150 bp. The Ion Total RNA-seq Kit v2 (4476286 Rev D, Life Technologies) was used to prepare the MAQC libraries for sequencing. The resulting material was quantified using the Agilent Bioanalyzer High Sensitivity Chip.

Ion template preparation and sequencing. Emulsion PCR was performed using the One Touch 2 (OT2) system following the Ion P1 Template OT2 200 protocol (Life Tech 0007488 Rev2.0) by using 315–615 million DNA molecules post-library preparation. Enriched spheres were quantified and ~400–800 million spheres were recovered. P1 Chips were loaded according to the spinning protocol and sequencing was performed using the Proton 200 sequencing kit (MAN0007491 Rev 3.0). Base calls were collected with Torrent Suite using v3.4.1 software.

Roche 454 GS FLX+. *Starting material and enrichment*. The MAQC A and MAQC B RNAs were subjected to two rounds of poly-A enrichment at a single laboratory site before distribution to the other 454 data generation core laboratories. Each RNA sample was enriched using the Oligotex mRNA Mini Kit (Qiagen), starting from 60 µg of RNA, according to manufacturer's instructions, using the spin column <0.25 mg method. A second enrichment step was performed following step 5 of the protocol, according to the manufacturer's instructions. Final elution was performed twice using 50 µl of 700C OEB elution buffer. The resulting enriched RNA was quantified by Nanodrop spectrophotometry and evaluated on the Agilent Bioanalyzer 2100 using an RNA PicoChip.

Library construction. Library synthesis and sequencing was performed with MAQC A and MAQC B samples at three core laboratory sites. Each site constructed one cDNA library from each of the poly-A–enriched RNA samples. Enriched RNA (200 ng) was reduced to 19 µl in a vacuum centrifuge at 60 °C, followed by library construction using the Roche cDNA Rapid Library Preparation Method Manual XL+ (May 2011) with the following modifications: (i) RNA Fragmentation Reagent kit (AM#8740, Life Technologies) was used in place of the RNA Fragmentation Solution; (ii) all magnetic particle concentrator (MPC) pelleting steps were held for 2 min; (iii) Roche rapid library multiplex identifier (RL-MIDs) adaptors were used for the adaptor ligation step; (iv) the final libraries were quantified using a Qubit fluorometer (Life Technologies) and average fragment sizes were determined by analyzing 1 µl of the libraries on the Agilent Bioanalyzer 2100 using a High-Sensitivity DNA LabChip; and (v) the library concentrations were determined using the average fragment size from the Bioanalyzer analysis. Final samples were diluted to 1×10^8 molecules/µl in Tris-HCl pH 8 buffer with 0.001% Tween-20.

Template enrichment and sequencing. Libraries were diluted to 1×10^6 molecules/µl for sequencing. Emulsion-based clonal amplification and sequencing on the Roche 454 Genome Sequencer FLX+ was performed according to the manufacturer's instructions (454 Life Sciences). Each library was sequenced on one full PicoTiterPlate (PTP) per laboratory site. An additional PTP per library was sequenced at one of the laboratories for a total of four PTPs per MAQC sample. Sequencing was done using the Roche XL+ sequencing kit with software version 2.6 or 2.8 with Flow Pattern A. Signal processing and base calling were performed using the bundled 454 Data Analysis Software (v.2.6 and 2.8).

Pacific Biosciences (PacBio) RS. The RNA-seq methods used in this study for full-length cDNA sequencing on the Pacific Biosciences RS are an early access protocol provided to the ABRF-NGS Consortium; refinements to the protocol are under development by the vendor.

Starting material. MAQC A and MAQC B RNA samples were used to generate PacBio libraries. The poly-A+ fraction was purified from total RNA using Invitrogen Dynabeads Oligo(dT)25, according to the manufacturer's protocol (61002, Life Technologies). MAQC A RNA (100 µg) was mixed with 1.33 mg of washed Dynabeads, and 200 µg MAQC B RNA was mixed with 2.66 mg of washed Dynabeads, collected by magnetic precipitation, washed as directed, and eluted into 27 µl of 10 mM Tris-HCl. The amount and purity of the poly-A+ RNA was assessed using an Agilent 2100 Bioanalyzer RNA NanoChip.

cDNA synthesis. Full-length, double-stranded cDNA was synthesized from the poly-A+ mRNA using the first five steps of the Invitrogen SuperScript Full Length cDNA Library Construction Kit II (A13268, Life Technologies); which included: (i) first strand cDNA synthesis; (ii) RNase I treatment; (iii) 5′G Cap-Antibody selection; (iv) second strand cDNA synthesis; and (v) cDNA size fractionation by Sephacryl column chromatography and precipitation. The cDNA was amplified with limited PCR using Phusion Hot Start Flex DNA Polymerase (M0535L, New England BioLabs), including PCR primers adapted from the 3′ oligo-dT primer used for first-strand cDNA synthesis and the 5′ adaptor used for second-strand cDNA synthesis, as described in the Invitrogen Superscript manual (primer sequences: pGGG ACA ACT TTG TCA AAG AAA and pTCG TCG GGG ACA ACT TTG TAC, respectively). The PCR amplification profile was 98 °C for 2 min, 14 cycles × (98 °C for 0.5 min, 64 °C for 0.5 min, and 72 °C for 4 min) and a final extension at 72 °C for 4 min. PCR-amplified cDNA was purified using 0.6× volume of Agencourt AMPure PB Beads (Beckman Coulter, Life Sciences Division), as specified by the supplier (Pacific Biosciences). The purified cDNA was recovered in 50 µl of TE buffer.

Total cDNA was divided into three MW size classes of 1–2 kb, 2–3 kb, and 3–8 kb using SYBR green and 0.8% agarose gel size fractionation, and recovered using Qiagen QIAquick Gel Extraction (Qiagen). Each cDNA size class was re-amplified using PCR conditions identical to those listed above. In order to prepare enough cDNA for multiple sequencing library constructions, a total of eight 100 µl PCR reactions were performed in parallel for each of the three cDNA size classes. The resulting PCR product was purified using 0.6× volume of AMPure PB beads as described above. Purified cDNA was quantified by a fluorometric Qubit assay and evaluated using an Agilent 2100 Bioanalyzer DNA 12000 chip.

Library construction and sequencing. Each of the three size-fractionated cDNA pools from the MAQC A and MAQC B samples were distributed to three core laboratory sites for library preparation and data generation, resulting in a total of 18 libraries. SMRTbell libraries were prepared from 0.5 and 1.0 µg of each cDNA size class according to the PacBio Large Insert Template Library Prep Kit. Double-stranded cDNA was subjected to DNA damage repair, end repair and blunt-end ligation to hairpin adaptors. Incomplete SMRTbell templates were degraded with a combination of Exonuclease III and Exonuclease VII. Intact cDNA SMRTbells were purified by three sequential AMPure PB purifications. Sequencing primers were annealed to the SMRTbell templates and subsequently bound to the sequencing polymerase using the Pacific Biosciences DNA/Polymerase v 2.0 binding kit, following manufacturer's instructions. The samples were sequenced on the PacBio RS using "C2" chemistry with SMRTcell loading via diffusion. The data collection times were adjusted per cDNA size bin. For size bins less than 2 kb, the 2 × 45 min movie protocol was used, whereas bins at or above 2 kb used a 1 × 90 min movie protocol.

Platform-specific library preparation parameters for all the participating sites are summarized in **Table 1**.

Data used in each section. Figure 2. Samples A–D, intact and degraded, sequenced at all sites.

Figure 3. **3a–e,g**. Samples A–B for all sites, except for ILMN, sites L,R,V were used here.

doi:10.1038/nbt.2972

3f. 454: site I all replicates; ILMN: site L, sample A-1_D1DJ4ACXX_CGATGT_L005_R1_001 and sample B-1_D1DJ4ACXX_GCCAAT_L005_R1_001; PAC: all data; PGM: site S, replicates 1-3; PRO: site S, replicate 1.

Figure 4. **4a**. Sample B was used. POLYA: site W replicate 1; RIBO: site W replicate 1; PAC: site H replicate 2; 454: site I replicate 1; PGM: site S replicate 1; PRO: site S replicate 1; MISEQ: site W replicate 1.

4b. Sample A and B for data in PRO, PGM, 454, ILMN site W all data.

Figure 5. ILMN site W all data.

49. Van Peer, G., Mestdagh, P. & Vandesompele, J. Accurate RT-qPCR gene expression analysis on cell culture lysates. *Sci. Rep.* **2**, 222 (2012).
50. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19 (2007).
51. Bustin, S.A. *et al*. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* **55**, 611–622 (2009).
52. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Robinson, M.D. & Smyth, G.K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).
54. Robinson, M.D. & Smyth, G.K. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* **9**, 321–332 (2008).
55. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
56. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
57. Canales, R.D. *et al*. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat. Biotechnol.* **24**, 1115–1122 (2006).
58. Dvinge, H. & Bertone, P. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics* **25**, 3325–3326 (2009).

Erratum: Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study

Sheng Li, Scott W Tighe, Charles M Nicolet, Deborah Grove, Shawn Levy, William Farmerie, Agnes Viale, Chris Wright, Peter A Schweitzer, Yuan Gao, Dewey Kim, Joe Boland, Belynda Hicks, Ryan Kim, Sagar Chhangawala, Nadereh Jafari, Nalini Raghavachari, Jorge Gandara,, Natàlia Garcia-Reyero, Cynthia Hendrickson, David Roberson, Jeffrey Rosenfeld, Todd Smith, Jason G Underwood, May Wang, Paul Zumbo, Don A Baldwin, George S Grills & Christopher E Mason

***Nat. Biotechnol.* 32, 915–925 (2014); published online 24 August 2014; corrected after print 10 October 2014; doi:1038/nbt.2972**

In the version of this article initially published, author Jeffrey Rosenfeld's middle initial "A" was omitted. The error has been corrected in the HTML and PDF versions of the article.