Towards standardization of RNA quality assessment using user-independent classifiers of microcapillary electrophoresis traces

Sandrine Imbeaud*, Esther Graudens, Virginie Boulanger, Xavier Barlet, Patrick Zaborski, Eric Eveno, Odilo Mueller[1], Andreas Schroeder[1] and Charles Auffray

Array s/IMAGE, Genexpress, Functional Genomics and Systems Biology for Health, LGN–UMR 7091, CNRS and Pierre and Marie Curie University of Paris 6, 7, rue Guy Môquet, 94801 Villejuif, France and [1]Agilent Technologies, Waldbronn, Germany

ABSTRACT

While it is universally accepted that intact RNA constitutes the best representation of the steady-state of transcription, there is no gold standard to define RNA quality prior to gene expression analysis. In this report, we evaluated the reliability of conventional methods for RNA quality assessment including UV spectroscopy and 28S:18S area ratios, and demonstrated their inconsistency. We then used two new freely available classifiers, the Degradometer and RIN systems, to produce user-independent RNA quality metrics, based on analysis of microcapillary electrophoresis traces. Both provided highly informative and valuable data and the results were found highly correlated, while the RIN system gave more reliable data. The relevance of the RNA quality metrics for assessment of gene expression differences was tested by Q-PCR, revealing a significant decline of the relative expression of genes in RNA samples of disparate quality, while samples of similar, even poor integrity were found highly comparable. We discuss the consequences of these observations to minimize artifactual detection of false positive and negative differential expression due to RNA integrity differences, and propose a scheme for the development of a standard operational procedure, with optional registration of RNA integrity metrics in public repositories of gene expression data.

INTRODUCTION

Purity and integrity of RNA are critical elements for the overall success of RNA-based analyses, including gene expression profiling methods to assess the expression levels of thousands of genes in a single assay. Starting with low quality RNA may strongly compromise the results of downstream applications which are often labor-intensive, time-consuming and highly expensive. However, in spite of the need for standardization of RNA sample quality control, presently there is no real consensus on the best classification criteria. Conventional methods are often not sensitive enough, not specific for single-stranded RNA, and susceptible to interferences from contaminants present in the sample. For instance, when using a spectrophotometer, a ratio of absorbances at 260 and 280 nm (A_{260}:A_{280}) greater than 1.8 is usually considered an acceptable indicator of RNA purity (1,2). However, the A_{260} measurement can be compromised by the presence of genomic DNA leading to over-estimation of the actual RNA concentration. On the other hand, the A_{280} measurement will estimate the presence of protein but provide no hint on possible residual organic contaminants, considered at 230 nm (3–5). Pure RNA will have A_{260}:A_{230} equal to A_{260}:A_{280} and >1.8 (1). A second check involves electrophoresis analysis, routinely performed using agarose gel electrophoresis, with RNA either stained with ethidium bromide (EtBr) (6–9), or the more sensitive SYBR Green dye (10). The proportion of the ribosomal bands (28S:18S) has conventionally been viewed as the primary indicator of RNA integrity, with a ratio of 2.0 considered to be typical of 'high quality' intact RNA (1). However, these methods are highly sample-consuming, using 0.5–2 μg total RNA and often not sensitive enough to detect slight RNA degradation. Today, microfluidic capillary electrophoresis with the Agilent 2100 bioanalyzer (Agilent Technologies, USA) has become widely used, particularly in the gene expression profiling platforms (11,12). It requires only a very small amount of RNA sample (as low as 200 pg), the use of a size standard during electrophoresis allows the estimation of sizes of RNA bands and the measurement appears relatively unaffected by contaminants. Integrity of

*To whom correspondence should be addressed. Tel: +33 0 1 49 58 35 13; Fax: +33 0 1 49 58 35 09; Email: imbeaud@vjf.cnrs.fr
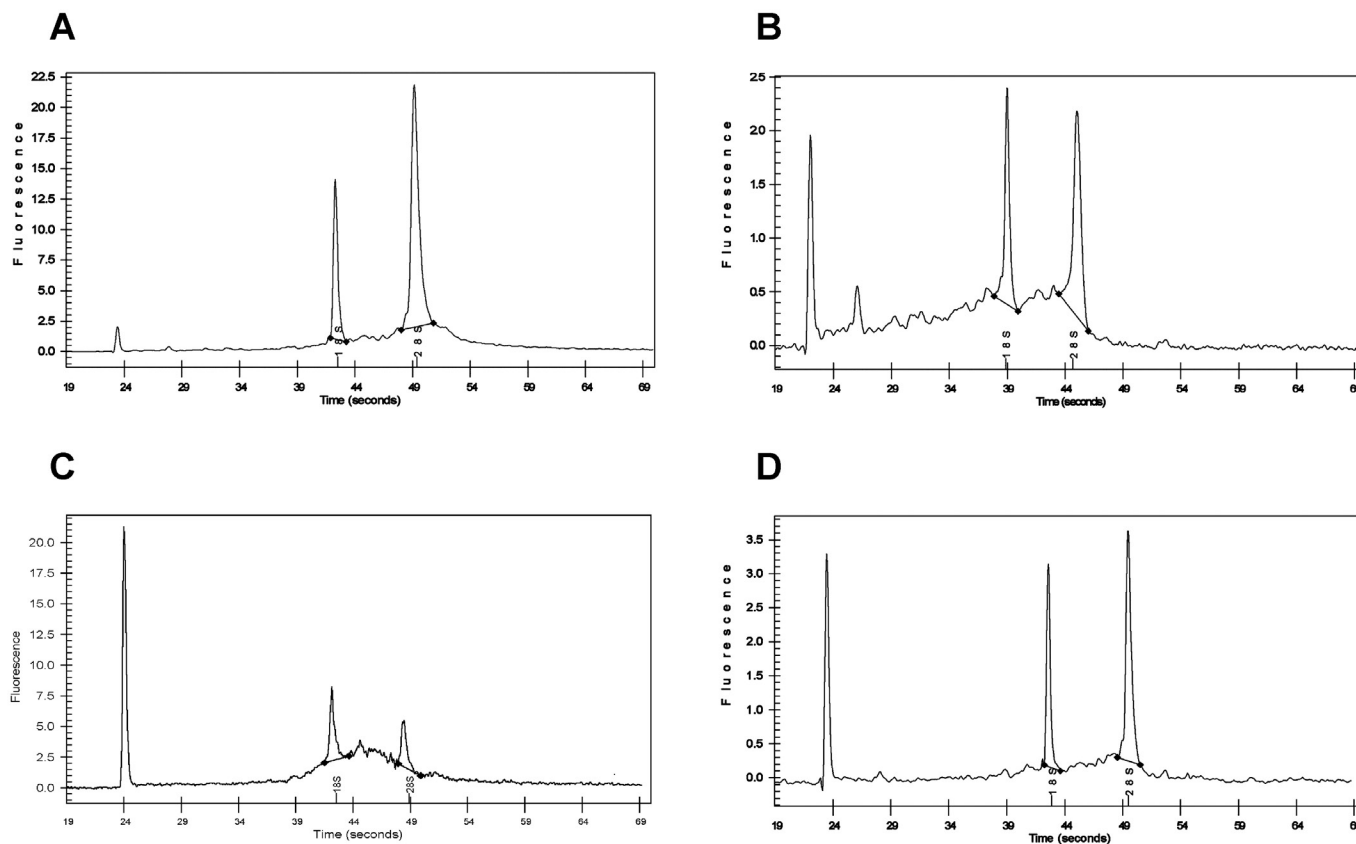
Figure 1. Chromatograms of microcapillary electrophoresis from four RNA samples showing different degrees of degradation. (**A**) Typical electropherogram of high-quality RNA including a clearly visible 28S:18S rRNA peak ratio of 2.0. (**B**) Partially degraded sample as indicated by a shift in the electropherogram to shorter fragment sizes, 28S:18S rRNA ratio of 1.0. (**C**) Partially sheared genomic DNA contamination; and (**D**) RNA judged intact but 28S:18S rRNA ratio <2.0 (ratio equals 1.5).

the RNA may be assessed by visualization of the 28S and 18S ribosomal RNA bands (Figure 1A and B); an elevated threshold baseline and a decreased 28S:18S ratio, both are indicative of degradation. A broad band shows DNA contamination (Figure 1C). As it is apparent from a review of the literature, the standard of a 2.0 rRNA ratio is difficult to meet, especially for RNA derived from clinical samples, and it now appears that the relationship between the rRNA profile and mRNA integrity is somewhat unclear (13–23). On the one hand, this may reflect unspecific damage to the RNA, including sample mishandling, postmortem degradation, massive apoptosis or necrosis, but it can reflect specific regulatory processes or external factors within the living cells. Altogether, it appears that total RNA with lower rRNA ratios is not necessarily of poor quality especially if no degradation products can be observed in the electrophoretic trace (Figure 1D).

For all these reasons, the development of a reliable, fully integrated and automated system appropriate for numeric evaluation of RNA integrity is highly desirable. Standardized RNA quality assessment would allow a more reliable comparison of experiments and facilitate exchange of biological information within the scientific community. With that prospect in mind, and with the aim of anticipating future standards by pre-normative research, we identified and tested two software packages recently developed to gauge the integrity of RNA samples with a user-independent strategy: one open

source, the degradometer software for calculation of the degradation factor and 'true' 28S:18S ratio based on peak heights (24) and the freely available RIN algorithm of the Agilent 2100 expert software, based on computation of a 'RNA Integrity Number' (RIN) (25). Both tools were developed separately to extract information about RNA integrity from microcapillary electrophoretic traces and produce a user-independent metrics. Using these tools, we assessed the purity and integrity of 414 RNA samples, derived from 14 different human adult tissues and cell lines, many of which representing tumors. Those results were compared with conventional RNA quality measurement approaches as well as with highly expert human interpretation. We evaluated the simplicity for users and examined the potential, accuracy and efficiency of each method to contribute to standardization of RNA integrity assessment upstream of biological assays. These procedures were further validated by real-time RT–PCR quantitation of the expression levels of three housekeeping genes, using the same RNA samples, at different levels of degradation.

MATERIALS AND METHODS

Sample preparation

Total RNA was prepared from human cell lines (especially from the ATCC bio-resource center, *N* = 50) and tissue samples

(clinical samples, $N = 285$) from 13 different human adult tissue types, i.e. blood, brain, breast, colon, epithelium, kidney, lymphoma, lung, liver, muscle, prostate, rectum and thyroid. RNA purification was performed by cesium chloride ultracentrifugation according to Chomczynski and Sacchi (26), by phenol-based extraction methods (TRIzol reagent, Invitrogen, USA), or silica gel-based purification methods (RNeasy Mini Kit, Qiagen, Germany; Strataprep kit, Stratagene, USA or SV RNA isolation kit, Promega, USA) according to the manufacturer's instructions with some modifications. Material was maintained at $-80°C$ with minimal handling. RNA extraction was carried out in an RNase-free environment (see Supplementary Table 1 online).

The commercially available RNA samples were the 'Universal Human Reference' ($N = 75$) distributed by Stratagene (USA), and human brain ($N = 2$) and muscle ($N = 2$) RNAs supplied by Clontech (USA).

Once extracted, RNA concentration and purity was first verified by UV measurement, using the Ultrospec3100 *pro* (Amersham Biosciences, USA) and 5 mm cuvettes. The absorbance (A) spectra were measured from 200 to 340 nm. A_{230}, A_{260} and A_{280} were determined. A_{260}:A_{280} and A_{260}:A_{230} ratios were calculated. For microcapillary electrophoresis measurements, the Agilent 2100 bioanalyzer (Agilent Technologies, USA) was used in conjunction with the RNA 6000 Nano and the RNA 6000 Pico LabChip kits. In total, 39 assays were run in accordance with the manufacturer's instructions (see Supplementary Notes online). To evaluate the reliability of the classifier systems described in this study, replicate runs were done on a set of 56 RNA samples loaded on different chips, resulting in 2 ($N = 41$), 3 ($N = 12$), 7 ($N = 2$) and 50 ($N = 1$) data points per sample.

Human RNA integrity categorization

RNA integrity checking was performed by expert operators who classified each total RNA sample within a predefined discrete category from 1 to 5, examining the integrity of the RNA from electropherograms (see Supplementary Table 2 online). A low number indicates high integrity. Reference criteria parameters include ribosomal peaks definition, baseline flatness, existence of additional or noise peaks between ribosomal peaks, low molecular weight species contamination and genomic DNA presence suspicion. A smearing of either 28S and 18S peaks, or a decrease in their intensity ratio indicate degradation of the RNA sample and results in the classification into the higher categories. To evaluate the robustness of this human interpretation, five highly experienced operators, trained in these cataloging steps, separately classified a subset of 33 samples from breast cancers. It included samples with varying levels of integrity: intact RNA (33%), low quality samples (20%) and a wide range of degradation (47%).

Degradometer analysis

Bioanalyzer electrophoretic data were exported in the degradometer software folder (.cld format). For comparison of samples, the original data were re-scaled by the classifier system, first along the time-axis to compensate for differences in migration time, then along the fluorescence intensity-axis to compensate for variation in total RNA amount. As a result, fluorescence curves that have the same shape will have the same peak heights after re-scaling. Then, Degradation Factors (DegFact) and corrected 28S:18S ratios were calculated (see Supplementary Table 3 online) using the mathematical model developed by Auer *et al.* (24), examining additional 'degradation peak signals' appearing in the lower molecular weight range and comparing them to ribosomal peak heights. Calculation of the DegFact is based on a numbering of continuous metrics, ranging from 1 to ∞; increasing DegFact values correspond to more degradation, and a new group of integrity is defined after 8 graduation steps. Once the classification of the RNA samples is completed, 4 groups of integrity are displayed, 3 showing an alert warning indicative of some measurable degradation (Yellow: 8–16, Orange: 16–24 and Red: $\geqslant$24), while all non-reliable data come together and form the fourth group (Black). We introduced a fifth class labeled White (<8), when no alert was produced by the software.

Software and manual are freely available at http://www.dnaarrays.org/downloads.php. Degradometer version 1.4.1 (released in May 2004) of the software was used.

'RNA Integrity Number' (RIN) algorithm analysis

Bioanalyzer electrophoretic sizing files (.cld format) collected with biosizing software version A.02.12.SI292 (released in March 2003) were imported in the Agilent 2100 expert software (RIN beta release). The RIN algorithm allows calculation of RNA integrity using a trained artificial neural network based on the determination of the most informative features that can be extracted from the electrophoretic traces out of 100 features identified through signal analysis. The selected features which collectively catch the most information about the integrity levels include the total RNA ratio (ratio of area of ribosomal bands to total area of the electropherogram), the height of the 18S peak, the fast area ratio (ratio of the area in the fast region to the total area of the electropherogram) and the height of the lower marker.

A total of 1300 electropherograms of RNA samples from various tissues of three mammalian species (human, mouse and rat), showing varying levels of degradation and an adaptive learning approach were used in order to assign a weight factor to the relevant features that describe the RNA integrity. A RIN number is computed for each RNA profile (see Supplementary Table 4 online) resulting in the classification of RNA samples in 10 numerically predefined categories of integrity. The output RIN is a decimal or integer number in the range of 1–10: a RIN of 1 is returned for a completely degraded RNA samples whereas a RIN of 10 is achieved for intact RNA sample.

In some cases, the measured electropherogram signals are of an unusual shape, showing for example peaks at unexpected migration times, spikes or abnormal fluctuation of the baseline. In such cases, a reliable RIN computation is not possible. Several separate neural networks were trained to recognize such anomalies and display a warning to the user or even suppress the display of a RIN number. Combining the results of the neural network for the RIN computation and the neural networks to detect anomalies, the RIN algorithm achieves a mean square error of 0.1 and a mean absolute error of 0.25 on an independent test set.

The beta release of the software and manual are freely available at http://www.agilent.com/chem/RIN. Agilent 2100

Table 1. Description of the quality metrics from 16 aliquots of a unique batch of RNA

Aliquot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
RIN	9	8.5	8.3	8.1	8.1	7.9	7.8	7.3	6.4	6.3	6.2	5.5	4.7	4.6	4.0	2.5
DegFact	4.9	4.8	5.5	5.2	5.0	6.9	6.1	6.8	13.1	11.6	12.3	16.6	16.1	22.3	26.0	30.5
HC	Ref	HC-level2-3							HC-level4					HC-level5		

Samples are grouped in three quality categories (HC), according to both metrics, i.e. RIN and DegFact values. Ref corresponds to the highest quality sample, used as calibrator (reference) sample to calculate the relative expression of the gene expression levels using Q-PCR.

expert version B.01.03.SI144 (released in November 2003) of the software was used.

Real-time RT–PCR

Expression levels of three housekeeping genes (HKG)—*GAPD*, *GUSB* and *TFRC*—were measured by quantitative PCR using the TaqMan Gene Expression Assays according to the manufacturer's instructions (Applied Biosystems, USA). Sixteen aliquots of a unique batch of RNA sample (Universal Human Reference RNA, Stratagene, USA) of various levels of integrity (cf. Table 1) were used to test the influence of RNA quality on the relative expression of those three genes. In parallel, a 5′ to 3′ comparison was done using two separate *GUSB* and *TFRC* TaqMan probes.

An homogeneous quantity (0.8–1 µg) of the RNA samples was subjected to a reverse transcription step using the high-capacity cDNA archive kit (Applied Biosystems, USA) as described by the manufacturer. Single-stranded cDNA products were then analyzed by real-time PCR using the TaqMan Gene Expression Assays according to the manufacturer's instructions (Applied Biosystems, USA). Single-stranded cDNA products were analyzed using the ABI PRISM 7700 Sequence Detector (Applied Biosystems, USA). The efficiency and reproducibility of the reverse transcription were tested using 18S rRNA TaqMan probes. Five assays were used, GAPDH-5′ (Hs99999905_m1), GUSB-5′ (Hs00388632_gH), GUSB-3′ (Hs99999908_m1), TFRC-5′ (Hs00951086_m1) and TFRC-3′ (Hs00951085_m1). In each case, duplicate threshold cycle (Ct) values were obtained and averaged; then expression levels were evaluated by a relative quantification method (27). The fold change in one tested HKG (target gene) was normalized to the 18S rRNA (reference gene) and compared to the highest quality sample (calibrator sample), using the following formula: Fold change = $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct = (C_{t\text{-target}} - C_{t\text{-reference}})_{\text{sample-n}} - (C_{t\text{-target}} - C_{t\text{-reference}})_{\text{calibrator-sample}}$. Sample-n corresponds to any sample for the target gene normalized to the reference gene and calibrator-sample represents the expression level (1×) of the target gene normalized to the reference gene considering the highest quality sample. Mean $2^{-\Delta\Delta Ct}$ and SD were calculated, considering the samples either individually or grouped by quality metrics categories, based on RIN metrics or DegFact values, together with the lower and upper bound mean of 95% Intervals of Confidence (IC). Using this analysis, if the expression levels of the HKG are not affected by the RNA degradation, the values of the mean fold change at each condition should be very close to 1 (since $2^0 = 1$) (27).

Statistical analysis

Descriptive statistics were executed using the XLSTAT software, version 7.1 (Addinsoft, USA), $P = 0.05$. Mean, SD and coefficient of variation (variation or CV) between and within groups of samples were calculated, together with a measure of the dispersion (range), inter-quartile range (1st and 3rd quartiles, Q1–Q3) and evaluation of the lower and upper bound mean of 95% Interval of Confidence (IC). Comparative statistical analyses between groups were completed, $P = 0.05$, using non-parametric statistical tests: two-independent Mann–Whitney U-test and k-independent Kruskal–Wallis test.

RESULTS

We analyzed 414 total RNA sample profiles from various human tissues (69%) and cell lines (31%) of either tumoral (85%) or normal (15%) origin, with varying levels of RNA integrity.

Conventional RNA quality characterization

Following extraction, RNA purity was directly determined by UV measurement from the $A_{260}:A_{280}$ and $A_{260}:A_{230}$ ratios. We found an average $A_{260}:A_{280}$ ratio of 1.89 [Interval of Confidence (IC) 1.88–1.90] with 92% of the values falling within the theoretically expected range (1.8–2.1) (Figure 2A), and an average $A_{260}:A_{230}$ ratio of 1.99 (IC 1.92–2.06), with 73% of the values greater than 1.8 (See Supplementary Table 1 online for details). Significant differences in $A_{260}:A_{280}$ ratios were observed between specific groups of samples (i.e. tumoral versus normal or tissues versus cell lines). For instance, RNA extracted from normal samples displayed an improved ratio of 1.97, with 97% falling within the desired range (Figure 2A). In contrast, the distribution of $A_{260}:A_{280}$ ratios was not found to correlate with either purification methods or tissues of origin.

RNA integrity was further assessed by resolving the 28S and 18S ribosomal RNA bands using the Agilent 2100 bioanalyzer and the RNA 6000 protocol. The analysis was done on 399 RNA profiles; data from 15 samples was not obtained due to device problems during the runs. The system automatically provided 28S:18S ratios for 348 (87%) of the 399 profiles. Figure 2B shows the distribution of the 28S:18S computed values, with a median ratio around 1.7 and a variation of 54% from the mean (IC 1.9–2.1 and Q1–Q3 1.4–2.5). In addition, a significant degree of variability of the 28S:18S ratio (19–24%) was found for identical samples from replicate runs (2–50 times). Among those RNA samples, 28S:18S ratios of 2.0 or greater were rare, less than 44% of the values measured being within the theoretically desired range, except for the samples prepared from cultured cells (Figure 2B). The integration failed in the remaining 51 cases, displaying an atypical migration, with no clear 28S and 18S rRNA bands, and no 28S:18S ratio was computed (data not shown).
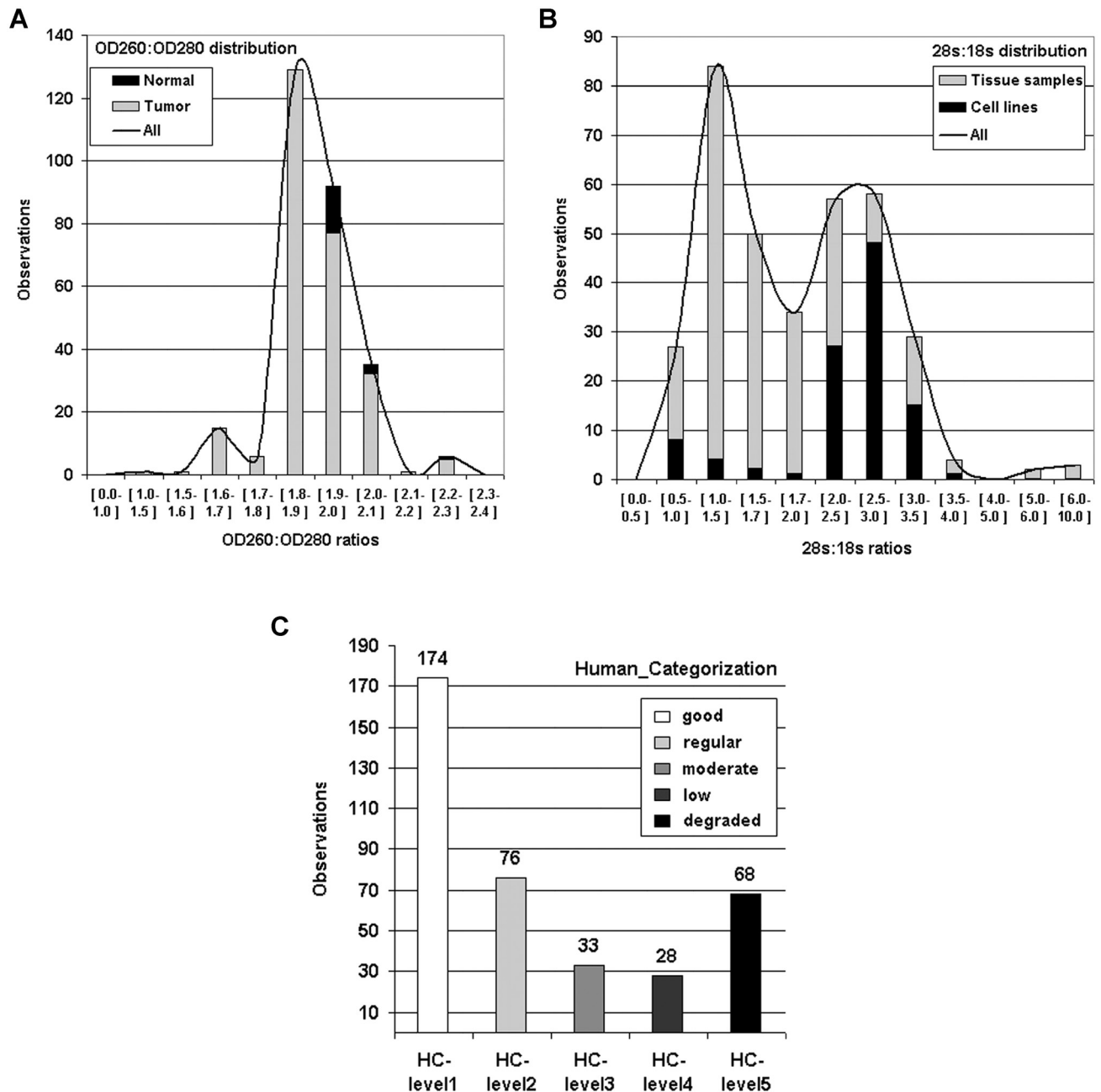
Figure 2. Conventional RNA integrity definition. (A and B) A_{260}:A_{280} and 28S:18S ratio distributions were checked for 399 RNA sample profiles. Predefined classes and number of observations are, respectively, indicated in the *x*- and *y*-axes. Lines refer to mean distribution. (**A**) The A_{260}:A_{280} ratio distribution from normal (in black) and cancerous (in gray) sample types is shown. (**B**) rRNA area ratios were successfully computed from 348 RNA sample profiles by the Agilent biosizing software based on a routine baseline detection and definition of ribosomals peak areas. Distribution of cell lines (in black) and tissue type (in gray) samples is shown. (**C**) Human categorization was done from Agilent electropherograms by trained operators who indexed 379 RNA sample profiles into 5 discrete classes (Human Categorization-level, HC-level). The distribution is represented by the number of records in each class. The remaining 20 samples could not be allocated.

Human categorization

Expert operators categorized the set of RNA samples by inspecting the electrophoretic traces of successful assays. Over the 399 RNA profiles checked, 379 (95%) were scored within predefined categories (Figure 2C), namely good [Human Categorization (HC)-level 1], regular (HC-level 2), moderate (HC-level 3), low (HC-level 4) and degraded (HC-level 5). The remaining 20 (5%) were flagged as displaying a temperature-sensitive profile: RNA samples initially

found intact became highly degraded when heated, although no RNase contamination was observed (data not shown). Estimation of the robustness of this cataloging was done through comparison of qualifying criteria using a set of 33 breast cancer samples (see Materials and Methods). Integrity of the samples was evaluated independently by five expert operators, and categorization was found highly reliable with a coefficient variation (CV) ∼16%. This is low considering that individual interpretation is involved, but can be explained

by the fact that very experienced operators accomplished the scoring based on a clearly defined set of instructions, thus limiting frequently observed subjective visual interpretation and inconsistency of human categorization.

Predictably, a 28S:18S ratio of 2.0 denoted high quality for a majority of RNA samples, 91% being classified in HC-levels 1 to 3. However, 83% of total RNAs with 28S:18S $\geq$ 1.0 but a low baseline between the 18S and 5S rRNA or front marker were also classified in HC-levels 1–3 (see Figure 1D) and could be considered suitable for most downstream applications.

User-independent classification

RNA degradation was first assessed using the degradometer software (see Materials and Methods). Over the 399 RNA profiles checked, all were scored in one of the five predefined classes (Figure 3A). Altogether, 334 (84%) Degradation Factors (DegFact) values were computed, the remaining 65 RNA samples (16%) displaying profiles that could not be interpreted reliably; no DegFact values could be scored, and samples were flagged in the Black category (Figure 3A). Most of them (80%) correspond to samples previously classified by our operators as degraded (HC-level 5). The remaining cases had an average degradation factor of 7.5 (IC 6.7–8.3) with large variations over the entire set of samples (over 103% from the mean, range 1–52). A lower variability was persistently found when identical samples from replicate runs were considered, resulting in observed DegFact values with a 26–32% CV. In addition, statistically significant differences were found between DegFact values of samples sorted by types. The highest DegFact values were found characteristic of tissue samples, 41% of them displaying a DegFact > 8, as compared with 6% for the cell lines (data not shown).

Remarkably, we found a significant linear relationship between the DegFact values distribution and the explicit human categorization. Most HC classes corresponded to an unambiguous DegFact distribution (Figure 3B), while HC-levels 2 and 3 form a single class: HC-level 1, mean DegFact of 3.3, SD of 2.8 (IC 2.8–3.7); HC-level 2 and 3, mean DegFact of 8.8, SD of 6.8 (IC 7.5–10.2); HC-level 4, mean DegFact of 15.9, SD of 7.8 (IC 12.7–19.1); HC-level 5, mean DegFact of 26.0, SD of 7.5 (IC 21.9–30.1). It is worth mentioning that the normalized heights of 18S and 28S peaks, and the interval between them after rescaling gradually decrease and then reverse with increasing degradation (Figure 3B).

Integrity of RNA samples was measured in parallel based on the RNA Integrity Number metrics using an artificial neural network trained to distinguish between different RNA integrity levels by examining the shape of the microcapillary electrophoretic traces (see Materials and Methods). Over the 399 RNA profiles checked, 363 (91%) were scored successfully (Figure 4A), with an average RIN of 7.7 (IC 7.4–8.0). The remaining 36 (9%) samples were associated with various unexpected signals, disturbing computation of the RIN using default anomaly detection parameters. In each case, a flag alert was added corresponding to critical anomalies including unexpected data in sample type, (or) ribosomal ratio, (or) baseline and signal in the 5S region (data not shown).

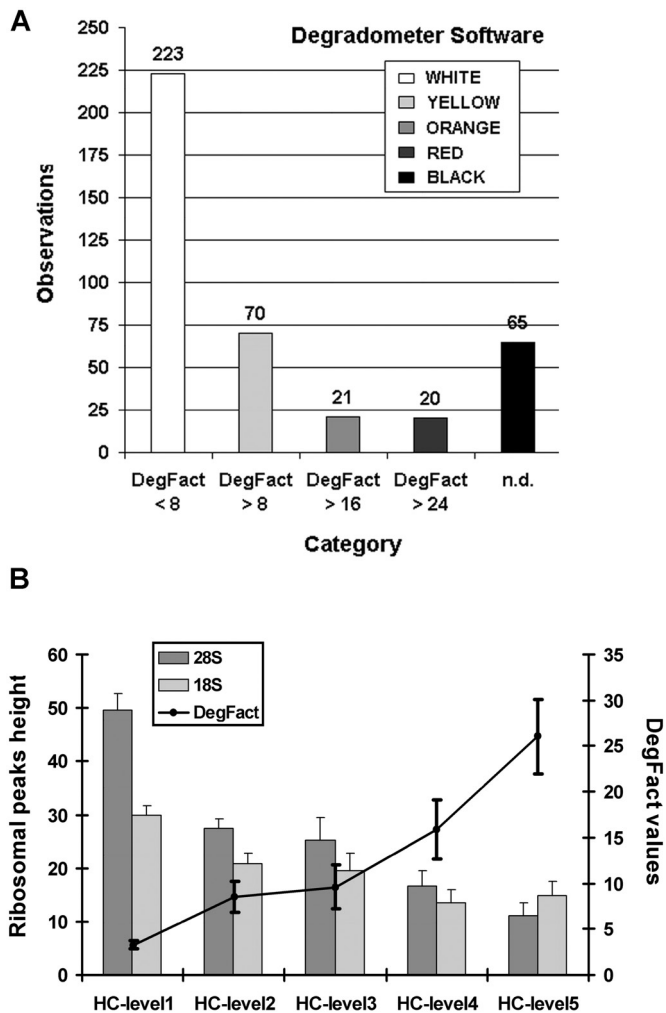RIN categorization was found regular, variability between replicate runs, compared to the other methods, being

Figure 3. RNA degradation characterization. Integrity of 399 RNA sample profiles was scored using the degradometer software. (**A**) A total of 334 RNA profiles were successfully categorized into 5 predefined alert classes using a mathematical model that quantifies RNA degradation and computes a degradation factor (DegFact). Four classes (White, Yellow, Orange and Red) are associated with different levels of degradation. A fifth class, Black alert corresponds to samples that the system was not able to qualify with accuracy (n.d.). The distribution is represented by the number of records in each class. (**B**) Comparative analysis was done using human evaluation (*x*-axis) based on electrophoresis analysis as a reference for RNA integrity classification; observations of rRNA peak heights and DegFact values were taken at each of the 5 HC levels. Histograms refer to the mean 28S and 18S rRNA peak heights and 95% confidence intervals (fluorescence intensities; left scale). Mean DegFact values and 95% confidence intervals (arbitrary unit, right scale) are plotted with the means joined.

consistently very small (CV 8–12%). As expected, the highest RIN were characteristic of cell line samples, 72% of them displaying a RIN $\geq$ 9, as compared with 47% for the tissue samples (data not shown).

A first group, corresponding to 295 (82%) of the 363 RNA profiles, was analyzed using the default settings of the RIN system, but with a lower threshold of RNA quantity loaded (20 ng) for reliable detection of anomalies than that recommended by the manufacturer (50 ng). A significant linear relationship was found between the RIN number and both the explicit human classification provided by our operators,
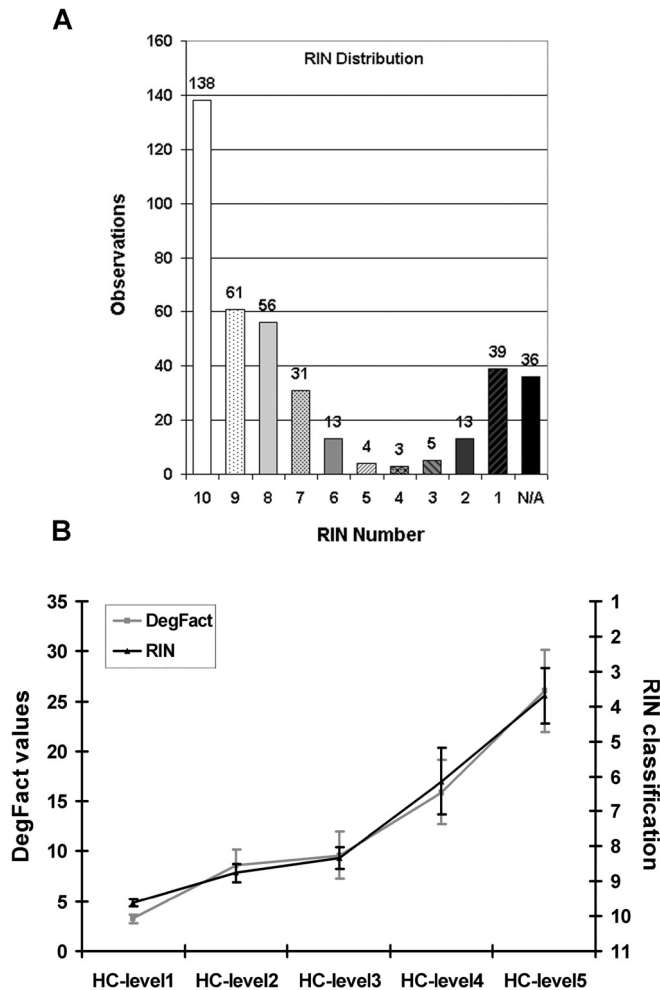
A

B

Figure 4. RIN categorization. Integrity of 399 RNA sample profiles was scored using the 2100 RIN expert software. (**A**) Classification was done based on a numbering metrics, from 1 to 10, with 1 being the most degraded profile and 10 being the most intact. RIN numbers were successfully computed for 363 RNA profiles; number integer category is indicated in the *x*-axis, N/A refers to remaining failed computation. The distribution is represented by the number of records in each class. (**B**) Comparative analysis was done using human evaluation (*x*-axis) based on electrophoresis analysis as a reference for RNA integrity classification; observations of RIN number and DegFact values were taken at each of the 5 HC levels. Mean RIN numbers (right scale) or DegFact values (left scale) and 95% confidence intervals are plotted with the means joined.

and the DegFact values calculated by the degradometer software (Figure 4B). Each distinct HC class corresponds to an explicit RIN number, with HC-levels 2 and 3 forming once again a single class: HC-level 1, mean RIN of 9.6, SD of 0.7 (IC 9.5–9.7); HC-level 2 and 3, mean RIN of 8.6, SD of 0.9 (IC 8.4–8.9); HC-level 4, mean RIN of 6.1, SD of 1.5 (IC 5.2–7.1); HC-level 5, mean RIN of 3.7, SD of 2.0 (IC 2.9–4.5).

For the remaining 68 samples (assay done with <20 ng of RNA), two separate groups were considered: 41 samples with a computed RIN below 5.0, and 27 above 7.0. All samples in the first group were derived from RNA 6000 Nano assays, with mean RNA quantities loaded below 10 ng (Q1–Q3, 5–12 ng), i.e. below the lower limit of quantitation indicated by the manufacturer. All but 8 of these samples were estimated by

our operators to be of poor quality (HC-level 4; $N = 3$) or degraded (HC-level 5; $N = 30$), and all but 4 were flagged Black by the degradometer software and no DegFact values were scored. These RNA profiles could not be interpreted reliably, possibly due to either the low RNA concentration or the unusual migration behavior and shifted baseline values of degraded samples. Thus, the two automated systems were in disagreement for these samples; while human interpretation was in most cases in agreement with the RIN system, with less than 20% of inconsistency. In the second group of 27 samples, 20 of the profiles were derived from RNA 6000 Pico assays with RNA quantities loaded being on average below 4 ng (Q1–Q3, 0.5–0.8 ng), which is within the manufacturer specifications. All but 3 of them were estimated by our operators to range from high (HC-level 1; $N = 12$) to correct (HC-level 2 and 3; $N = 12$) quality levels. In addition, all RNA profiles except 1 were scored by the degradometer software, most of them displaying an alert flag ($N = 20$); some slight degradation was detected, associated to a low mean DegFact value of 9.7 (IC 8.1–11.3; Q1–Q3, 6.2–12.6). Thus, both automated systems and human interpretations agreed in most of these cases, with <11% of inconsistency.

Gene expression profiling versus RNA quality

The influence of RNA quality categorization obtained with both user-independent classifiers on gene expression profiling was explored using real-time RT–PCR. The expression levels of three housekeeping genes (HKG)—*GAPDH*, *GUSB* and *TFRC*—were measured in 16 aliquots of a unique RNA displaying various integrity metrics (Table 1). The mean correlation coefficient (r) between the threshold cycle (Ct) among the 16 samples and both quality metrics was found high: $r = -0.87$ considering the RIN metrics and $r = 0.85$ considering the DegFact values. The values of the mean fold changes, calculated according to the $2^{-\Delta\Delta Ct}$ quantification method (see Materials and Methods), were found lower than 1.0, corresponding to the expression level (1×) in the sample exhibiting the highest RNA quality (Table 2 and Figure 5). Considering that HKG expression was measured relative to the reference sample, an obvious decline of the relative expression levels was observed, up to 24, 70 and 82%, in samples categorized according to the RIN metrics (Figure 5A) and DegFact values (Figure 5B). These results indicate that 2- to 7-fold differences may be expected in the relative expression levels of genes in samples that differ only by their quality (Table 2). These fold differences are much larger than those measured for RNA samples of comparable integrity, consistently lower than 1.6 (Table 2 and Figure 5). In addition, an unambiguous gap in the distribution may be defined (Figure 5A and B), distinguishing the RNA samples of the higher quality categories (RIN ⩾ 8 and DegFact values < 7) from those of the lower categories (RIN < 8 and DegFact values ⩾ 12).

It would be expected that measuring expression of an intact mRNA would yield approximately equal results regardless of the region being probed, and if mRNA fragmentation had occurred, then some sequences may be more abundant than others. We thus tested the effect of PCR probe location on the RNAs. The 5′ and 3′ *GUSB* probes, separated by 1209 nt, were associated with highly similar threshold cycle (Ct) measures ($r = 0.98$, β parameter = 0.88) (Figure 5C). Similar results were

Table 2. Q-PCR results collected on three housekeeping genes (HGK)—*GAPD*, *GUSB* and *TFRC*—considering the sample aliquots grouped by quality metrics categories

	Quality metrics			Q-PCR			Relative expression (%)	Fold differences				
	Metrics	IC (P = 0.05)	N	$2^{-\Delta\Delta Ct}$	IC (P = 0.05)	IC-rep (P = 0.05)		Between categories			Within categories	Technical variation
RIN	9	—	1	1.00	—	0.04	—	—			—	1.1
	8.1	0.2	6	0.76	0.06	0.03	−24	[1.2–1.4]	—		<1.2	1.1
	6.1	0.6	6	0.30	0.04	0.00	−70	[3.0–3.8]	[2.1–3.2]	—	<1.3	1.0
	3.7	1.0	3	0.18	0.04	0.01	−82	[4.5–7.4]	[3.2–6.1]	[1.2–2.5]	<1.6	1.1
DegFact	4.9	—	1	1.00	—	0.04	—	—			—	1.1
	5.8	0.6	7	0.69	0.08	0.03	−31	[1.3–1.6]	—		<1.3	1.1
	13.9	1.8	5	0.31	0.05	0.01	−69	[2.9–3.8]	[1.8–3.0]	—	<1.3	1.0
	26.3	3.8	3	0.18	0.04	0.01	−82	[4.5–7.4]	[2.8–5.7]	[1.2–2.6]	<1.6	1.2

The *N*-value corresponds to the number of samples by category. The mean quality metrics, i.e. RIN and DegFact and the mean fold change ($2^{-\Delta\Delta Ct}$) relative to the reference sample are indicated, together with the 95% confidence intervals. Observed technical variation (IC-rep, P = 0.05) is also specified, considering duplicate (two per gene per target sample) and replicate (six per gene per calibrator sample) measures. The reference sample exhibits a RIN of 9, a DegFact value of 4.9 and by default mean fold change set to 1. The observed decrease in the expression (relative expression, %) relative to the reference sample is calculated. The fold differences refer to the fold-ratios that are expected in the expression levels for a gene, across categories (between categories), given that the samples only differ by their quality, and within each category (within categories), considering RNA of comparable integrity. The fold-ratios (technical variation) that may be expected by chance in the gene expression levels, P = 0.05, from some technical reasons, are also considered.

obtained for *TFRC*, with probes separated by 2066 nt ($r = 0.84$, β parameter = 0.92, data not shown). It seems therefore that the region being probed is not a source of variation in our results.

DISCUSSION

It is universally accepted that RNA purity and integrity are of foremost importance to ensure reliability and reproducibility of downstream applications. In the biomedical literature (PubMed, November 2004), from the 485 090 articles that relate to RNA, and the 287 515 or 40 395 including respectively the 'quality' or 'integrity' term, less than 100 were found to contain 'RNA quality' or 'RNA integrity' terms. Interestingly, half of them were published between 2001 and 2004; but none is proposing a standard operational procedure for RNA quality assessment to the scientific community. Except for two studies (24,25), those reports are based on 10 to 15 years old methods (1), indicating that they represent the established and currently mostly used methods. Our results strongly challenge the reliability and usefulness of those conventional methods, demonstrating their inconsistency to evaluate RNA quality.

First, the A_{260}:A_{280} and A_{260}:A_{230} ratios are reflecting RNA purity, but are not informative regarding the integrity of the RNA. Available RNA extraction and purification methods yield highly pure RNA with very little DNA or other contaminations, resulting most often in both ratios $\gg 1.8$, although 18% of the samples were found degraded and 7% more of poor quality. The high A_{260}:A_{280} ratios are indicative of limited protein contaminations, whereas high A_{260}:A_{230} ratios are indicative of an absence of residual contamination by organic compounds such as phenol, sugar or alcohol, which could be highly detrimental to downstream applications. Nonetheless, samples displaying low A_{260}:A_{230} ratios ($\ll 1.8$) did not exhibit any inhibition during downstream applications, such as cDNA synthesis and labeling or *in vitro* transcription (data not shown). Second, due to a lack of reliability, the 28S:18S rRNA ratios may not be used as a gold standard for assessing RNA integrity. When ribosomal ratios were calculated from identical samples but through independent runs, a large degree of variability (CV 19–24%) was observed. Moreover, using the biosizing software, we found 28S:18S rRNA ratios evaluation compromised by the fact that their calculation is based on area measurements and therefore heavily dependent on definition of start and end points of peaks. In 13% of the cases, the system was unable to localize the ribosomal peaks, and therefore no 28S:18S ratios were computed. For the remaining samples, no clear correlation between 28S:18S ratios and RNA integrity was found although RNAs with 28S:18S >2.0 were usually of high quality. Most of the RNAs we studied (83%), displaying a 28S:18S > 1.0, could be considered of good quality. Interestingly, Auer *et al*. (24) in a study on 19 tissues from seven organisms, reported that an objective measurement of the RNA integrity may possibly be done through comparison of re-scaled 28S and 18S peak heights, but not of the corresponding areas. Actually, we observed a linear relationship between RNA integrity and differences in normalized 28S and 18S peak heights. Increased degradation resulted in a significant decrease in the scaled corrected heights of the ribosomal peaks, with inversion of the ratio at the highly degraded stages (cf. Figure 3B). In comparison to the area computation, 28S:18S rRNA re-scaled peak height measurement produced more consistent values, with a CV reduced to 12–14%, and displayed clear concentration-independent values (see Supplementary Tables 1 and 3 online). Human evaluation of the integrity of RNA through visual inspection of the electrophoresis profiles provided very consistent data. Variability between classifications produced by five independent expert operators (CV 16%) was lower than with automated management of more conventional control 28S:18S area values (CV 19–24%). It is, however, very time-consuming and strongly dependent on individual competence. Even with highly trained specialists, 5% of the set of RNA samples could not be allocated to any of the five predefined categories; their corresponding profiles were considered by our experts as atypical, displaying a temperature-sensitive shape (data not shown).

These strategies appear unsuitable for standardization and quality control of RNA integrity assessment, which require simple but consistent expert-independent classification, facilitating information exchanges between laboratories.
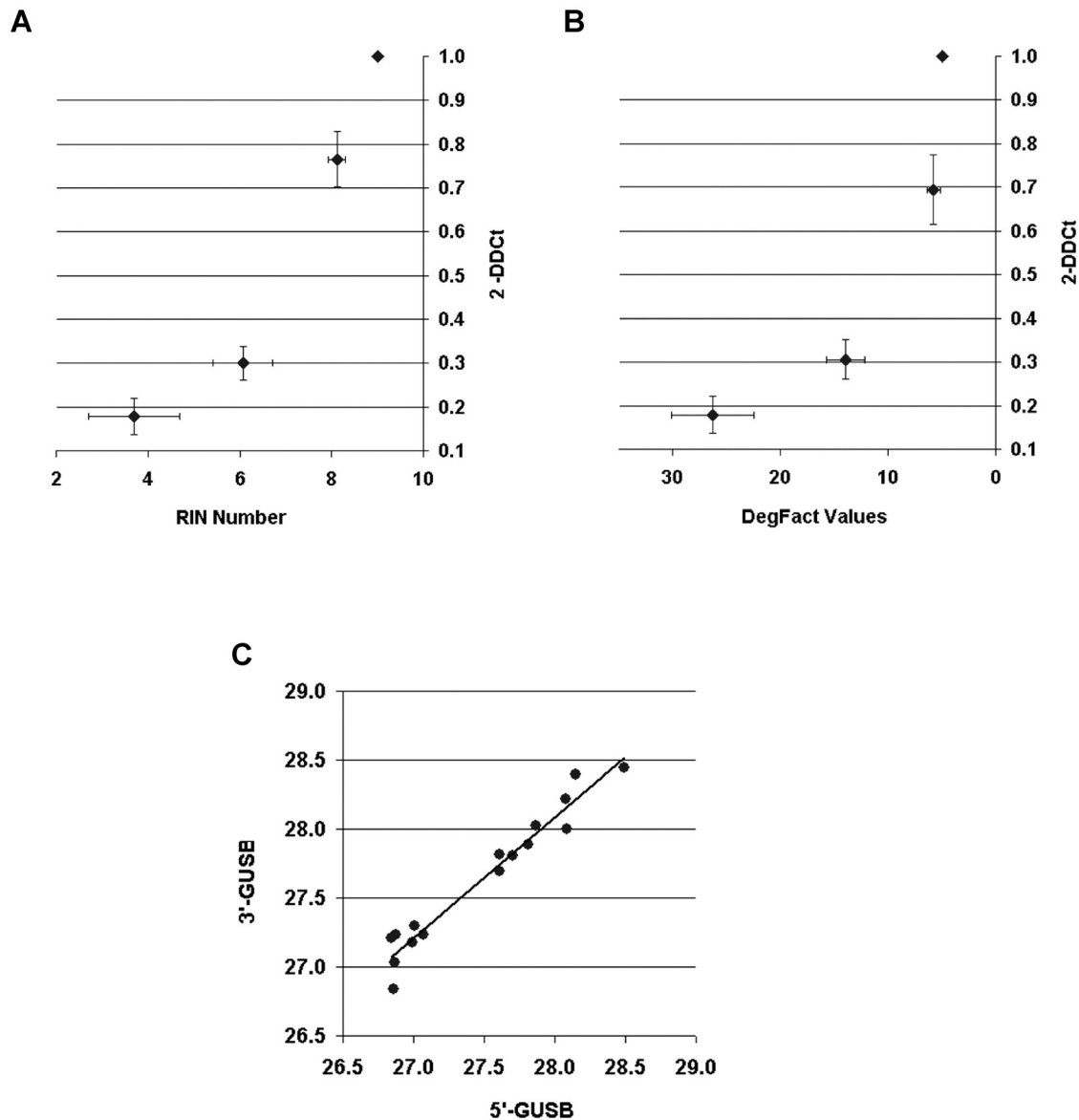
Figure 5. Relative expression between quality metrics categories. Q-PCR collected on 16 aliquots of unique batch of RNA of various RNA integrities (cf. Table 1) and 3 housekeeping genes (*GUSB*, *TFRC* and *GAPD*). (**A** and **B**) Representation of the mean fold change ($2^{-\Delta\Delta Ct}$) of the RNA aliquots grouped by quality metrics categories, including (A) RIN number and (B) DegFact values, with expected 95% confidence intervals. (**C**) Distribution of the threshold cycle (Ct) measured on the 16 RNA aliquots using GUSB-5′ (*x*-axis) and GUSB-3′ (*y*-axis) Taqman probes. The linear regression curve is indicated.

We therefore investigated the performance of two recently developed user-independent software algorithms (24,25). The degradometer software provided a reliable evaluation of RNA integrity based on the identification of additional 'degradation peak signals' and their integration in a mathematical calculation together with the ribosomal peak heights. It allowed characterization of the integrity of 84% of the samples tested, one-third with an alert flag, which was first found to be fairly informative, as it strongly reduces the complexity of the metrics by introducing three distinct classes labeled Yellow, Orange and Red, and can be used as a first straightforward simple filtering step. However, degradation factors (DegFact) metrics yield precise measures with less than 32% CV and are much more valuable than flag alerts for the purpose of standardization. The same is true for the RNA Integrity Number

'RIN' software which allowed the characterization of the integrity of 91% of the RNA samples tested, with a RIN value for 363 RNA sample profiles with less than 12% CV. In general, there was a good agreement between the human classification, the degradation factor and the RIN (see Figure 4B). This provided a cross-validation of the user-independent qualification systems tested. Both resulted in the refinement of human interpretations, validating four statistically relevant classes of samples, namely good (HC-level 1), regular/moderate (HC-level 2 and 3), poor (HC-level 4) and degraded (HC-level 5). Moreover, the 5% RNA samples previously flagged by the operators as displaying an atypical temperature-sensitive shape were unambiguously assigned to one or the other category of samples [RIN = 7.3 (IC 6.8–7.8); DegFact = 11.9 (IC 9.5–14.2); data not shown].
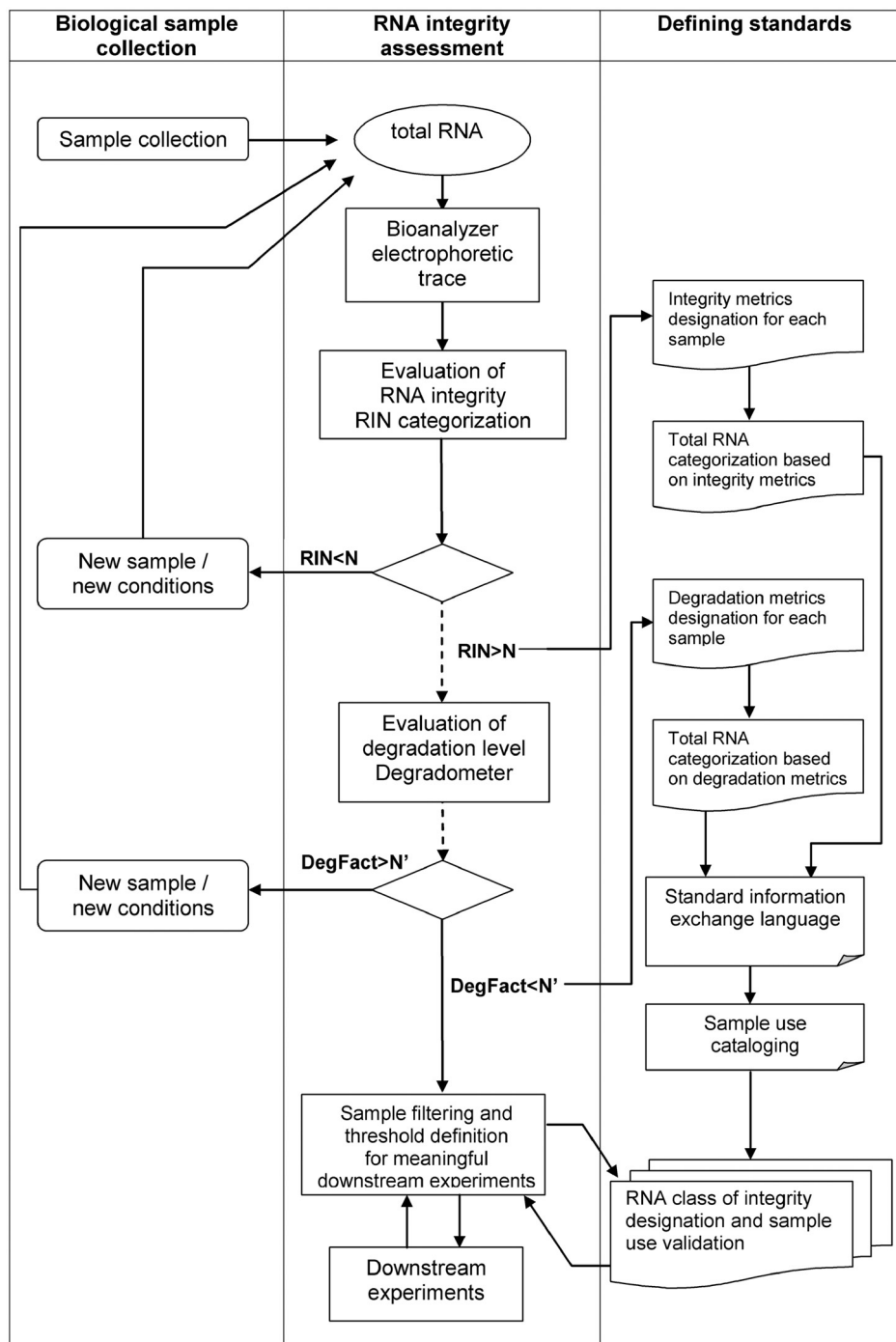
Figure 6. Workflow of operational procedure for RNA quality assessment. Integrity of the RNA, once extracted and purified from cell lines, clinical or biological tissues samples, is controlled from the widely used bioanalyzer electrophoretic traces. As standard part of the Agilent analysis software (25), a RIN metrics is first calculated, scoring each RNA sample into 10 numerically predefined categories of integrity (RIN, from 1 to 10; N is a threshold value). As an independent control, a degradation factor metrics (DegFact, from 1 to ∞; N' is a threshold value) may optionally be allocated to each RNA sample using the bioanalyzer-independent degradometer software (24). In a standard operating procedure, RIN and/or DegFact metrics will first be used as a standard exchange language to document RNA integrity and degradation, second to classify the RNA in homogeneous groups, and finally to select samples of comparable RNA integrity to improve the scheme of meaningful downstream experiments. The standard operating procedure will benefit from feedback information that will help users to define threshold integrity metrics values based on the results of RNA-based analyses.

Altogether, we found the degradometer and RIN algorithms to be highly reliable user-independent methods for automated assessment of RNA degradation and integrity. The RIN system is a slightly more informative tool, able to compute assessment metrics for 91% of the RNA profiles, compared to 84% with the degradometer software; the remaining being flagged respectively as N/A or Black alert. For samples available below a low limit of 20 ng ($N = 80$) the RIN system provided

metric values for 85% of them, compared to only 46% with the degradometer software. Similarly, the RIN system was able to provide metric values for 81% of poor quality samples (including low quality and degraded samples; $N = 96$), whereas the degradometer software could classify only 44% of them. Another advantage with the RIN classifier is that, if there are critical anomalies detected (including genomic DNA contamination, wavy baseline, etc.), threshold settings may be changed and a reliable RIN value computed. This was the case for 25 of the 363 RNA sample profiles successfully classified by the system.

While intact RNA obviously constitutes the best representation of the natural state of the transcriptome, there are situations in which gene expression analysis may be desirable even on partially degraded RNA. Some studies report collection of reasonable microarray data from RNA samples of impaired quality (28), leading to meaningful results if used carefully. Moreover, Auer *et al.* (24) recently concluded that degradation does not preclude microarray analysis if comparison is done using samples of comparable RNA integrity. We confirmed the direct influence of the RNA quality on the distribution of gene expression levels, by detecting using Q-PCR a significant (up to 7-fold) difference in the relative expression of genes in samples of slightly decreased RNA integrity, which is much larger than the variation within comparable RNA quality categories (cf. Figure 5 and Table 2). This may correlate with ratio discrepancies in gene expression experiments, and therefore with false positive and false negative rates of differential gene expression when comparing two samples. Therefore, computing reliable metrics of RNA integrity, even if the RNA is found to be partially degraded, may be highly valuable. The straight and unambiguous relationships established between human interpretations and both RIN and DegFact distributions indicates that, using these metrics, it should be possible to distinguish specific samples that are too disparate to be included in comparative gene expression analyses without compromising the results. Although the information provided by these user-independent classifiers is not a guarantee for successful downstream experiments, it gives a more comprehensive picture of the samples and can be used as a safeguard against performing useless and costly experiments.

Thus, the RIN system may be used as simple metrics that can be easily integrated in any sample tracking information system for definition of standard operating procedures under quality assurance following a scheme such as the one described in Figure 6. In this context, we suggest that the growing number of laboratories performing RNA Quality Control by microcapillary electrophoresis should be offered the option to report objective RNA quality metrics as part of the 'Minimum Information About a Microarray Experiment' MIAME standards (29). Through registration of RNA profiles in a public electronic repository, such standardized information should enable and facilitate comparisons of RNA-based bioassays performed across laboratories with RNA samples of similar quality, in much the same way as sequencing traces are compared.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

REFERENCES

1. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual, 2nd edn.* Cold Spring Harbor Laboratory Press, NY.
2. Manchester,K.L. (1996) Use of UV methods for measurement of protein and nucleic acid concentrations. *Biotechniques*, **20**, 968–970.
3. Manchester,K.L. (1995) Value of A_{260}/A_{280} ratios for measurement of purity of nucleic acids. *Biotechniques*, **19**, 208–210.
4. Glasel,J.A. (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques*, **18**, 62–63.
5. Wolfe,A.R. and Meehan,T. (1994) The effect of sodium ion concentration on intrastrand base-pairing in single-stranded DNA. *Nucleic Acids Res*., **22**, 3147–3150.
6. Le Pecq,J.B. and Paoletti,C. (1966) A new fluorometric method for RNA and DNA determination. *Anal. Biochem*., **17**, 100–107.
7. Vendrely,R., Alexandrov,K., De Sousa Lechner,M.C. and Coirault,Y. (1968) Fractionation of ribonucleic acids by 'Sephadex' agarose gel electrophoresis. *Nature*, **218**, 293–294.
8. Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. *Biochemistry*, **16**, 4743–4751.
9. Bonini,J.A. and Hofmann,C. (1991) A rapid, accurate, nonradioactive method for quantitating RNA on agarose gels. *Biotechniques*, **11**, 708–710.
10. Schneeberger,C., Speiser,P., Kury,F. and Zeillinger,R. (1995) Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. *PCR Methods Appl*., **4**, 234–238.
11. Mueller,O., Hahnenberger,K., Dittmann,M., Yee,H., Dubrow,R., Nagle,R. and Ilsley,D. (2000) A microfluidic system for high-speed reproducible DNA sizing and quantitation. *Electrophoresis*, **21**, 128–134.
12. Bustin,S.A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J. Mol. Endocrinol*., **29**, 23–39.
13. Mori,N., Mizuno,D. and Goto,S. (1978) Increase in the ratio of 18S RNA to 28S RNA in the cytoplasm of mouse tissues during aging. *Mech. Ageing Dev*., **8**, 285–297.
14. Houge,G., Robaye,B., Eikhom,T.S., Golstein,J., Mellgren,G., Gjertsen,B.T., Lanotte,M. and Doskeland,S.O. (1995) Fine mapping of 28S rRNA sites specifically cleaved in cells undergoing apoptosis. *Mol. Cell. Biol*., **15**, 2051–2062.
15. Monstein,H.J., Nylander,A.G. and Chen,D. (1995) RNA extraction from gastrointestinal tract and pancreas by a modified Chomczynski and Sacchi method. *Biotechniques*, **19**, 340–344.
16. Reno,C., Marchuk,L., Sciore,P., Frank,C.B. and Hart,D.A. (1997) Rapid isolation of total RNA from small samples of hypocellular, dense connective tissues. *Biotechniques*, **22**, 1082–1086.
17. Payao,S.L., Smith,M.A., Winter,L.M. and Bertolucci,P.H. (1998) Ribosomal RNA in Alzheimer's disease and aging. *Mech. Ageing Dev*., **105**, 265–272.
18. Banerjee,S., An,S., Zhou,A., Silverman,R.H. and Makino,S. (2000) RNase L-independent specific 28S rRNA cleavage in murine coronavirus-infected cells. *J. Virol*., **74**, 8793–8802.
19. Margan,S.H., Handelsman,D.J., Mann,S., Russell,P., Rogers,J., Khadra,M.H. and Dong,Q. (2000) Quality of nucleic acids extracted from fresh prostatic tissue obtained from TURP procedures. *J. Urol*., **163**, 613–615.

20. Schoor,O., Weinschenk,T., Hennenlotter,J., Corvin,S., Stenzl,A., Rammensee,H.G. and Stevanovic,S. (2003) Moderate degradation does not preclude microarray analysis of small amounts of RNA. *Biotechniques*, **35**, 1192–1196, 1198–1201.

21. Skrypina,N.A., Timofeeva,A.V., Khaspekov,G.L., Savochkina,L.P. and Beabealashvilli,R. (2003) Total RNA suitable for molecular biology analysis. *J. Biotechnol.*, **105**, 1–9.

22. Dumur,C.I., Nasim,S., Best,A.M., Archer,K.J., Ladd,A.C., Mas,V.R., Wilkinson,D.S., Garrett,C.T. and Ferreira-Gonzalez,A. (2004) Evaluation of quality-control criteria for microarray gene expression analysis. *Clin. Chem.*, **50**, 1994–2002.

23. Wex,T., Treiber,G., Lendeckel,U. and Malfertheiner,P. (2003) A two-step method for the extraction of high-quality RNA from endoscopic biopsies. *Clin. Chem. Lab. Med.*, **41**, 1033–1037.

24. Auer,H., Lyianarachchi,S., Newsom,D., Klisovic,M.I., Marcucci,G., Kornacker,K. and Marcucci,U. (2003) Chipping away at the chip bias: RNA degradation in microarray analysis. *Nature Genet.*, **35**, 292–293.

25. Mueller,O., Lightfoot,S. and Schroeder,A. (2004) RNA Integrity Number (RIN)—standardization of RNA quality control. Agilent Application Note, Publication Number-5989-1165EN, 1–8.

26. Chomczynski,P. and Sacchi,N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.

27. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.

28. Dash,A., Maine,I.P., Varambally,S., Shen,R., Chinnaiyan,A.M. and Rubin,M.A. (2002) Changes in differential gene expression because of warm ischemia time of radical prostatectomy specimens. *Am. J. Pathol.*, **161**, 1743–1748.

29. Brazma,A., Hingamp,P., Quackenbush,J., Sherlock,G., Spellman,P., Stoeckert,C., Aach,J., Ansorge,W., Ball,C.A., Causton,H.C. *et al.* (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genet.*, **29**, 365–371.