

METHODOLOGY ARTICLE

Open Access

mRNA transcript quantification in archival samples using multiplexed, color-coded probes

Patricia P Reis¹, Levi Waldron², Rashmi S Goswami^{1,8}, Wei Xu⁵, Yali Xuan¹, Bayardo Perez-Ordóñez⁶, Patrick Gullane⁷, Jonathan Irish⁷, Igor Jurisica^{2,3,4} and Suzanne Kamel-Reid^{1,5,6,8*}

Abstract

Background: A recently developed probe-based technology, the NanoString nCounter™ gene expression system, has been shown to allow accurate mRNA transcript quantification using low amounts of total RNA. We assessed the ability of this technology for mRNA expression quantification in archived formalin-fixed, paraffin-embedded (FFPE) oral carcinoma samples.

Results: We measured the mRNA transcript abundance of 20 genes (*COL3A1*, *COL4A1*, *COL5A1*, *COL5A2*, *CTHRC1*, *CXCL1*, *CXCL13*, *MMP1*, *P4HA2*, *PDPN*, *PLOD2*, *POSTN*, *SDHA*, *SERPINE1*, *SERPINE2*, *SERPINH1*, *THBS2*, *TNC*, *GAPDH*, *RPS18*) in 38 samples (19 paired fresh-frozen and FFPE oral carcinoma tissues, archived from 1997-2008) by both NanoString and SYBR Green I fluorescent dye-based quantitative real-time PCR (RQ-PCR). We compared gene expression data obtained by NanoString vs. RQ-PCR in both fresh-frozen and FFPE samples. Fresh-frozen samples showed a good overall Pearson correlation of 0.78, and FFPE samples showed a lower overall correlation coefficient of 0.59, which is likely due to sample quality. We found a higher correlation coefficient between fresh-frozen and FFPE samples analyzed by NanoString ($r = 0.90$) compared to fresh-frozen and FFPE samples analyzed by RQ-PCR ($r = 0.50$). In addition, NanoString data showed a higher mean correlation ($r = 0.94$) between individual fresh-frozen and FFPE sample pairs compared to RQ-PCR ($r = 0.53$).

Conclusions: Based on our results, we conclude that both technologies are useful for gene expression quantification in fresh-frozen or FFPE tissues; however, the probe-based NanoString method achieved superior gene expression quantification results when compared to RQ-PCR in archived FFPE samples. We believe that this newly developed technique is optimal for large-scale validation studies using total RNA isolated from archived, FFPE samples.

Background

A vast collection of formalin-fixed and paraffin-embedded (FFPE) tissue samples are currently archived in anatomical pathology laboratories and tissue banks around the world. These samples are an extremely valuable source for molecular biology studies, since they have been annotated with varied information on disease states and patient follow-up, such as disease progression in cancer and prognosis/survival data. Although FFPE samples provide an ample source for genetic studies, formalin fixation is known to affect the quality of DNA and RNA extracted from FFPE samples and its

downstream applications, such as amplification by the Polymerase Chain Reaction (PCR) or microarrays [1].

Von Ahlfen *et al.*, 2007 [1] described the different factors (e.g. fixation, storage time and conditions) that can influence the integrity of RNA extracted from FFPE tissues, and its downstream applications. They showed that differences in storage time and temperature had a large effect on the degree of RNA degradation. In their study, RNA samples extracted within 1 to 3 days after formalin fixation and paraffin embedding maintained their integrity. Similarly, RNA isolated from FFPE samples that were stored at 4°C showed higher quality compared to samples stored at room temperature or at 37°C. They also reported that RNA fragmentation occurs gradually over time. It is also known that cDNA synthesis from FFPE-derived RNA is limited due to the use of

* Correspondence: suzanne.kamel-reid@uhn.on.ca

¹Division of Applied Molecular Oncology, Princess Margaret Hospital, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada
Full list of author information is available at the end of the article

formaldehyde during fixation. Formaldehyde induces chemical modification of RNA, characterized by the formation of methylene crosslinks between nucleic acids and protein. These chemical modifications can be partially irreversible [2], limiting the application of techniques such as reverse transcription, which uses mRNA as a template for cDNA synthesis. A fixation time over 24 hours was shown to result in a higher number of irreversible crosslinks [3,4]. Overall, fixation time and method of RNA extraction are the main factors that determine the extent of methylene crosslinks [1].

A recently developed probe-based technology, the NanoString nCounter™ gene expression system, has been shown to allow accurate mRNA expression quantification using low amounts of total RNA [5]. This technique is based on direct measurement of transcript abundance, by using multiplexed, color-coded probe pairs, and is able to detect as little as 0.5 fM of mRNA transcripts; described in detail in Geiss *et al.*, 2008 [5]. In brief, unique pairs of a capture and a reporter probe are synthesized for each gene of interest, allowing ~800 genes to be multiplexed, and their mRNA transcript levels measured, in a single experiment, for each sample. In addition, in a recent study, mRNA expression levels obtained using NanoString were more sensitive than microarrays and yielded similar sensitivity when compared to two quantitative real-time PCR techniques: TaqMan-based RQ-PCR and SYBR Green I fluorescent dye-based RQ-PCR [5]. Although NanoString and RQ-PCR were shown to produce comparable data in good quality samples, NanoString is hybridization-based, and does not require reverse transcription of mRNA and subsequent cDNA amplification. This feature of NanoString technology offers advantages over PCR-based methods, including the absence of amplification bias, which may be higher when using fragmented RNA isolated from FFPE specimens. In addition, NanoString assays do not require the use of control samples, since absolute transcript abundance is determined for each single sample and normalized against the expression of housekeeping genes in that same sample [5].

Although NanoString technology has been optimized for gene expression analysis using formalin-fixed samples, to our knowledge we are the first to report the use of this technology for mRNA transcript quantification using clinical, archival, FFPE cancer tissues. In our pilot study, we used the NanoString nCounter™ assay for gene expression analysis of archival oral carcinoma samples. In order to show that mRNA levels obtained by NanoString analysis of FFPE tissues were accurate, we compared quantification data obtained using RNA isolated from paired fresh-frozen and FFPE oral cancer

samples. Our goal was to determine whether this technology could be applied for accurate gene expression quantification using archived, FFPE oral cancer tissues. We also aimed to compare whether quantification data obtained by NanoString achieved a higher correlation than data obtained by SYBR Green I fluorescent dye-based RQ-PCR, using the same paired fresh-frozen and FFPE samples.

Methods

Tissue samples

This study was performed under approval of the Research Ethics Board at University Health Network. Tissues were collected with informed patient consent. Study samples included primary fresh-frozen and formalin-fixed, paraffin-embedded (FFPE) tumor samples from 19 patients with oral squamous cell carcinoma. All patients had surgery as primary treatment. Fresh-frozen tissues were collected at the time of surgical resection, and samples were snap frozen and kept in liquid nitrogen until RNA extraction. RNA from these tumor samples was extracted and kept at -80C for long term storage. Representative FFPE tissue sections were obtained from the same tumor samples. We collected a total of 38 tumor samples (paired fresh-frozen and FFPE) from 19 patients. In addition, we included the analysis of a commercially available human universal RNA (pool of cancer cell lines) (Stratagene) and human normal tongue RNA (Stratagene); these samples were used as quality controls, since they are a source of high quality RNA, and have been previously used in other studies [6,7].

RNA extraction and cDNA synthesis

Total RNA was isolated from fresh-frozen tissues using Trizol reagent (Life Technologies, Inc., Burlington, ON, Canada), followed by purification using the Qiagen RNeasy kit and treatment with the DNase RNase-free set (Qiagen, Valencia, CA, USA). RNA extraction and purification steps were performed according to the manufacturers' instructions.

For FFPE tissues, one tissue section was taken from each specimen, prior to RNA extraction, stained with hematoxylin and eosin (H&E) and examined by a pathologist (B.P-O), to ensure that tissues contained >80% tumor cells. RNA was isolated from five 10 µm sections from FFPE samples, using the RecoverAll™ Total Nucleic Acid Isolation Kit (Ambion, Austin, TX, USA), following the manufacturer's procedures. RNA extracted from both fresh-frozen and FFPE tissues was assessed for quantity using Nanodrop 1000 (Nanodrop), and for quality using the 2100 Bioanalyzer (Agilent Technologies, Canada).

For RQ-PCR experiments, cDNA was synthesized from 1 µg total RNA isolated from fresh-frozen or FFPE tissues, using the M-MLV reverse transcriptase enzyme and according to manufacturer's protocol (Invitrogen).

Gene expression quantification using multiplexed, color-coded probe pairs (NanoString nCounter™)

Genes selected for testing in this technical report are frequently over-expressed in oral cancer (our own data, currently submitted for publication elsewhere). Probe sets for each gene were designed and synthesized by NanoString nCounter™ technologies (Table 1). Probe sets of 100 bp in length were designed to hybridize specifically to each mRNA target. Probes contained one capture probe linked to biotin and one reporter probe attached to a color-coded molecular tag, according to the nCounter™ code-set design.

RNA samples were randomized using a numerical ID, in order to blind samples for sample type (fresh-frozen or FFPE) and sample pairs. Samples were then subjected to NanoString nCounter™ analysis by the University Health Network Microarray Centre (<http://www.microarrays.ca/>) at the Medical Discovery District (MaRS), Toronto, ON, Canada. The detailed protocol for mRNA transcript quantification analysis, including sample preparation, hybridization, detection and scanning followed the manufacturer's recommendations, and are available at http://www.nanostring.com/uploads/Manual_Gene_Expression_Assay.pdf/ under <http://www.nanostring.com/applications/subpage.asp?id=343>. We used 100 ng of total RNA isolated from fresh-frozen tissues, as suggested by the manufacturer. FFPE tissues required a higher amount of total RNA (400 ng) for detection of probe signals. Technical replicates of three paired fresh-frozen and FFPE tissues were included. Data were analyzed using the nCounter™ digital analyzer software, available at <http://www.nanostring.com/support/ncounter/>.

Quantitative real-time RT-PCR

In addition, we performed RQ-PCR analysis in the same fresh-frozen and FFPE samples and compared this to gene expression data determined by the NanoString nCounter assay. RQ-PCR analysis was performed as previously described, using SYBR Green I fluorescent dye [8,9]. Gene IDs and primer sequences are described in Table 2. Primer sequences were designed using Primer-BLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Gene expression levels were normalized against the average Ct (cycle threshold) values for the two internal control genes (*GAPDH* and *RPS18*) and calculated relative to a commercially

available normal tongue reference RNA (Stratagene). Ct values were extracted using the SDS 2.3 software (Applied Biosystems). Data analysis was performed using the $\Delta\Delta C_t$ method [10].

Statistical analysis

Absolute mRNA quantification values obtained by NanoString as well as relative expression values obtained by RQ-PCR were log₂-transformed. Summary statistics as median, mean, range were provided. Pair-wise Pearson product-moment correlation analysis [11] was applied to test the correlation between gene expression data obtained by NanoString and RQ-PCR analysis in fresh-frozen vs. FFPE samples, as well as the correlation between NanoString and RQ-PCR data in fresh-frozen or FFPE samples. Both overall correlation and correlation across sample pairs were calculated. Statistical analyses were performed using version 9.2 of the SAS system and user's guide (SAS Institute, Cary, NC). In addition, Pearson correlation between sample pairs was plotted as heatmaps, in order to visualize the grouping of similar samples. Heatmaps were generated by hierarchical clustering analysis, using *hclust* R function, in R statistical environment [12].

Results

Technical data on sample quality

Bioanalyzer results for fresh-frozen samples showed a mean RNA integrity number (RIN) of 8.3 (range 4.6-9.8), with the majority of fresh-frozen samples (13/19) having a RIN ≥ 8 . FFPE samples were degraded and the mean RIN was 2.3 (range 1.5-2.5); this result was expected since FFPE samples are archival tissues. Representative examples of the Bioanalyzer results for one fresh-frozen and one FFPE sample are shown in Figure 1. FFPE samples used in our study have been archived from a time period between 1997-2008.

Correlation between mRNA transcript quantification in fresh-frozen vs. FFPE samples (NanoString)

Raw data quantification values obtained by NanoString were log₂ transformed, and values derived from the 19 paired fresh-frozen and FFPE samples were compared. The pair-wise Pearson product-moment correlation was 0.90 ($p < 0.0001$). The scatter plot and histogram for log₂ values from fresh-frozen and FFPE samples are shown in Figure 2A. Analysis of the three replicate pairs (log₂ transformed values) demonstrated a correlation of 0.93 ($p < 0.0001$). In addition, we performed unsupervised hierarchical clustering analysis of these data, and heatmaps are shown in Figure 2B.

Table 1 Probe sets for genes of interest used for Nanostring analysis

Gene Symbol	Accession Number	Target Region	Target Sequence
<i>COL3A1</i>	NM_000090.3	180-280	TTGGCACAACAGGAAGCTGTTGAAGGAGGATGTTCCCATCTTGGTCAGTCCTATGCCGATAGAGATGTCTGGAAGCCAGAACCATGCCAAATATGTGTCT
<i>COL4A1</i>	NM_001845.4	780-880	TGGGCTTAAGTTTTCAAGGACCAAAAGGTGACAAGGGTGACCAAGGGTCAGTGGCCTCCAGGAGTACCAGGACAAGCTCAAGTTCAAGAAAAGGAGA
<i>COL5A1</i>	NM_000093.3	6345-6445	GTAAAGGTACCCACCATCACCAAAGCCTCCGTTTTAAACAACCTCCAACACGATCCATTTAGAGGCCAAATGTCATTCTGCAGGTGCCTTCCCGATGG
<i>COL5A2</i>	NM_000393.3	4075-4175	GGTTCATGCTACCCTGAAGTCACTCAGTAGTCAGATTGAAACCATGCGCAGCCCGATGGCTCGAAAAGACCCAGCCCGCAGCTGTGATGACCTAAAG
<i>CTHRC1</i>	NM_138455.2	685-785	CTGTGGAAGGACTTTGTGAAGGAATTGGTCTGGATTAGTGGATGTTGCTATCTGGTTGGCACTTGTTTCAGATTACCCAAAAGGAGATGCTTCTACTGG
<i>CXCL1</i>	NM_001511.1	445-545	AGGCCCTGCCCTTATAGGAACAGAAGAGGAAAGAGAGACACAGCTGCAGAGGCCACCTGGATTGTGCCTAATGTGTTGAGCATCGCTTAGGAGAAGTCT
<i>CXCL13</i>	NM_006419.2	0-100	GAGAAGATGTTTGAAAAAAGTACTCTGCTAATGAGCCTGGACTCAGAGCTCAAGTCTGAACCTCTACCTCCAGACAGAATGAAGTTCATCTCGACATCTC
<i>MMP1</i>	NM_002421.3	1117-1217	AAATGGGCTTGAAGCTGCTTACGAATTTGCCGACAGAGATGAAGTCCGGTTTTCAAGGGAATAAGTACTGGGCTGTTTCAGGGACAGAATGTGCTACAC
<i>P4HA2</i>	NM_001017974.1	1600-1700	TGTGCTTGTGGGCTGCAAGTGGGTCTCCAATAAGTGGTTCATGAACGAGGACAGGAGTTCCTTGAGACCTTGTGGATCAACAGAAGTTGACTGACATCCT
<i>PDPN</i>	NM_006474.4	431-531	CTCCAGGAACCAGCGAAGACCGCTATAAGTCTGGCTTGACAACCTCTGGTGGCAACAAGTGTCAACAGTGAACAGGCATTCGCATCGAGGATCTGCCAAC
<i>PLOD2</i>	NM_182943.2	2590-2690	AAACATTGCACCTAATAACGTGGGAGAAGACTTTCAGGGAGGTGGTTGCAATTTCTAAGGTACAATTGCTCTATTGAGTCACCACGAAAAGGCTGGAGC
<i>POSTN</i>	NM_001135935.1	910-1010	AGAGACGGTCACTTCACACTCTTTGCTCCCAATGAGGCTTTTGAGAAACTTCCACGAGGTGTCCTAGAAAAGGATCATGGGAGACAAAAGTGGCTTCCG
<i>SDHA</i>	NM_004168.1	230-330	TGGAGGGGCAGGCTTGCAGCTGCATTTGGCCTTTCTGAGGCAGGGTTAATACAGCATGTGTTACCAAGCTGTTTCTACCAGGTCACACACTGTTGCA
<i>SERPINE1</i>	NM_000602.2	2470-2570	TGTGTTCAATAGATTTAGGAGCAGAAAATGCAAGGGGCTGCATGACCTACCAGGACAGAACCTTCCCAATTACAGGGTACTCACAGCCGCAATTGGTGAC
<i>SERPINE2</i>	NM_006216.2	240-340	CGCTGCCTCCATCTGCTCCCACTTCAATCCTCTGTCTCTCGAGGAAGTCTAGGCTCCAACACGGGGATCCAGGTTTTCAATCAGATTGTGAAGTCGAGGCC
<i>SERPINH1</i>	NM_001235.2	880-980	ATGGTGGACAACCGTGGCTTTCATGGTACTCGGCTCTATACCTGGGTGTCATGATGATGCACCGGACAGGCCTTACAACACTACGACGACGAGAAGG
<i>THBS2</i>	NM_003247.2	4460-4560	AAACATCCTTGCAAAATGGGTGTGACGCGGTTCCAGATGTGGATTTGGCAAAAACCTATTTAAGTAAAAGGTTAGCAGAGCAAAGTGCAGTGGTCTTTAGCTG
<i>TNC</i>	NM_002160.1	6885-6985	CAGAAATCTTGAAGGCAGGCGCAAACGGGCATAAATTTGGAGGACCCTGGGTGAGAGAGGAATAAGGCGGCCAGAGCGAGGAAAGGATTTTACCAAAG
<i>GAPDH</i>	NM_002046.3	35-135	TCCTCCTGTTGACAGTCAGCCGCATCTTTTTCGCTGCCAGCCGAGCCACATCGCTCAGACACCATGGGGAAGGTGAAGGTGCGAGTCAACGGATTT
<i>RPS18</i>	NM_022551.2	110-210	GCGGCGGAAAATAGCCTTTGCCATCACTGCCATTAAGGGTGTGGGCCGAAGATATGCTCATGTGGTGTGAGGAAAGCAGACATTGACCTACCAAGAGG

GAPDH and *RPS18* were used as internal controls for normalization of Nanostring data.

Table 2 Primer sequences used in the RQ-PCR experiments

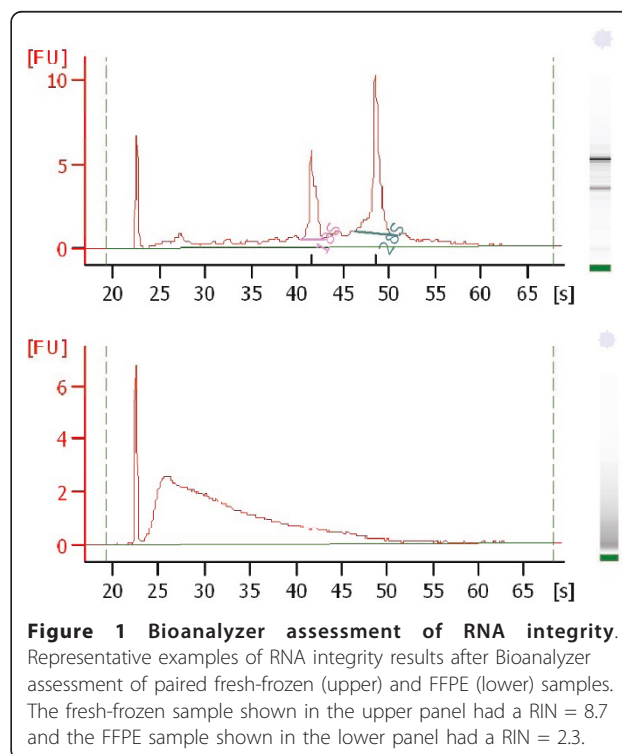
Gene symbol	Primer sequence	Amplicon length
<i>GAPDH</i>	Forward 5'-CCTGTTTCGACAGTCAGCCGCAT-3' Reverse 5'-GACTCCGACCTTCACCTTCCCC-3'	87 bp
<i>RPS18</i>	Forward 5'-GCGGCGGAAAATAGCCTTTGCC-3' Reverse 5'-CCTCTTGGTGAGGTCAATGTCTGC-3'	100 bp
<i>MMP1</i>	Forward 5'-CAAATGGGCTTGAAGTGCTTACG-3' Reverse 5'-GTGTAGCACATTCTGTCCCTGAACA-3'	101 bp
<i>COL4A1</i>	Forward 5'-AAGGACCAAAAAGGTGACAAGGGTGA-3' Reverse 5'-GAACTTGAGCTTGCTCCTGTACTCC-3'	72 bp
<i>COL5A1</i>	Forward 5'-GTCATCCCACCATACCAAAAGCC-3' Reverse 5'-ATCGGGAAGGCACCTGCAGAAATG-3'	92 bp
<i>THBS2</i>	Forward 5'-TTGCAAATGGGTGTGACGCGGT-3' Reverse 5'-AAGCACCCGACTTTGCTCTGCT-3'	86 bp
<i>TNC</i>	Forward 5'-ACGAACACTCAATCCAGTTTGTCTGA-3' Reverse 5'-TGGAAATTTATGCCCGTTTGCGCC-3'	89 bp
<i>COL3A1</i>	Forward 5'-TGGCACAAACAGGAAGCTGTTGAAGG-3' Reverse 5'-ACACATATTTGGCATGGTTCTGGCT-3'	97 bp
<i>COL5A2</i>	Forward 5'-TCATGCTACCTGAAGTCACTCAGT-3' Reverse 5'-AGGTCATCACACGTGCGGGC-3'	93 bp
<i>PDPN</i>	Forward 5'-CAGGAACCAAGCAAGACCGCT-3' Reverse 5'-TGGCAGATCCTCGATGCGAATGC-3'	95 bp
<i>POSTN</i>	Forward 5'-CGGTCACCTCACACTCTTTGCTCCC-3' Reverse 5'-CGGAAGCCACTTTGTCTCCCATGA-3'	95 bp
<i>SERPINE2</i>	Forward 5'-ACCATGAAGTGGCATCTCCCCCT-3' Reverse 5'-TGGAGCCTAGTTCCTCGAGAGACA-3'	100 bp
<i>SERPINH1</i>	Forward 5'-CCGTGGCTTCATGGTGACTCGG-3' Reverse 5'-AGTAGTTGTAGAGGCCTGTCCGGT-3'	74 bp
<i>SDHA</i>	Forward 5'-CTCCAAGCCCATCCAGGGGCAA-3' Reverse 5'-CAGAGTGACCTTCCAGTGCCAA-3'	100 bp
<i>PLOD2</i>	Forward 5'-TGGCTCTTTGCCGAAATGCTAGAG-3' Reverse 5'-GGGGGCTGAGCATTGGGAATGTTT-3'	87 bp

Table 2 Primer sequences used in the RQ-PCR experiments (Continued)

<i>P4HA2</i>	Forward: 5'-AGGAGCTGCCAAAGCCCTGA-3' Reverse: 5'-ACCTGCTCCATCCACAACACCG-3'	170 bp
<i>CTHRC1</i>	Forward: 5'-TTGTTCACTGGCTCACTTCG-3' Reverse: 5'-TTCAATGGGAAGAGGTCTCG-3'	102 bp
<i>CXCL1</i>	Forward: 5'-ATTTCTGAGGAGCCTGCAAC-3' Reverse: 5'-CACATACATTCCTGCCTT-3'	100 bp
<i>CXCL13</i>	Forward: 5'-GAGCCTGTCAAGAGGCAAAG-3' Reverse: 5'-CTGGGGATCTTCAATGCTA-3'	142 bp

SERPINE1 was excluded from RQ-PCR analysis since no primer pairs tested showed good efficiency for amplification in FFPE samples. Primer sequences used yielded short amplicon lengths, as indicated.

We also performed a correlation analysis between mRNA transcript quantification values (log2 transformed values) for each pair of fresh-frozen *versus* FFPE sample (sample by sample comparison). This analysis is important as it allows us to determine whether the amount of mRNA transcripts of a given gene is maintained in individual sample pairs. The mean correlation coefficient obtained was 0.94, with a minimum correlation of 0.77 and a maximum correlation of 0.99.



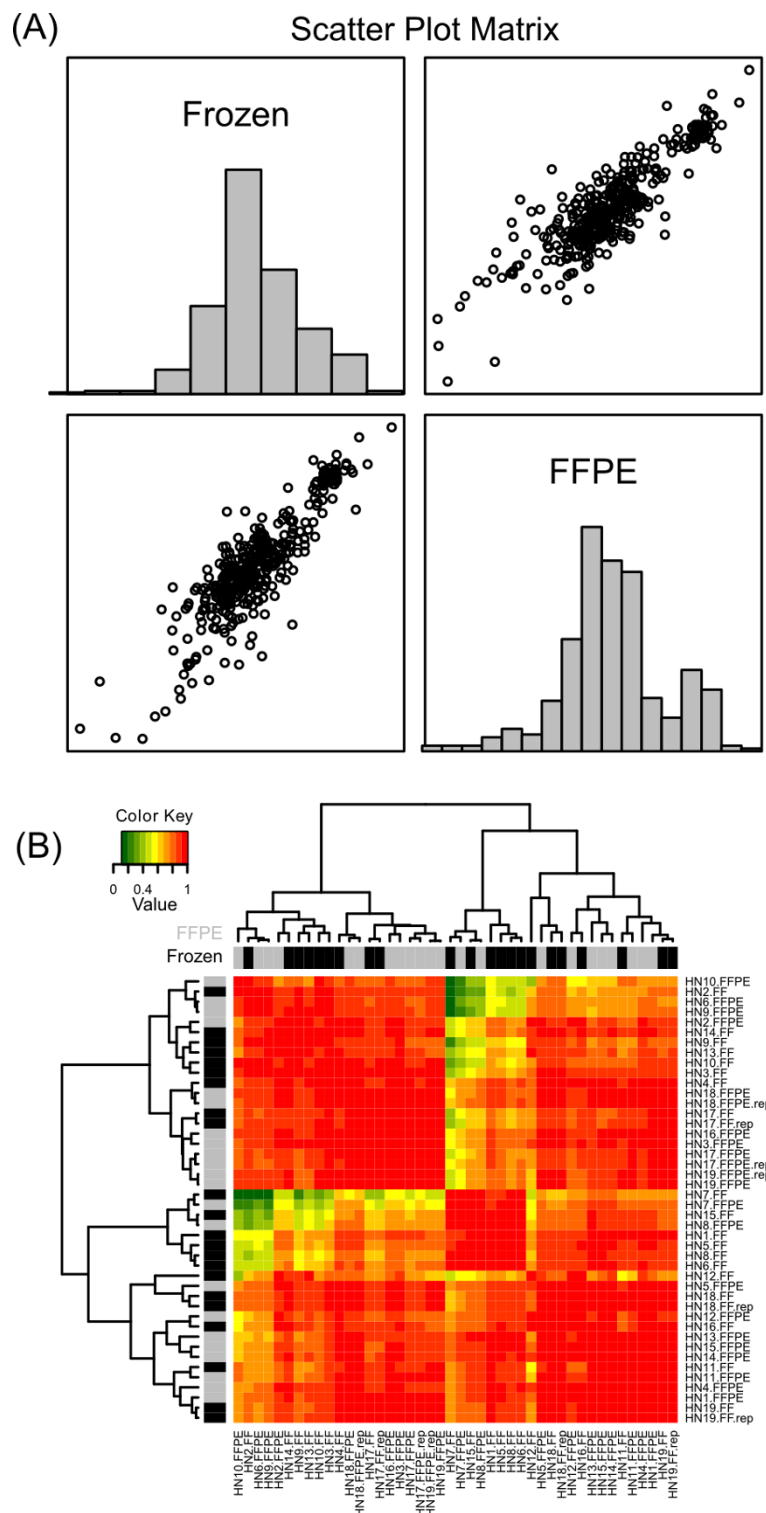


Figure 2 Correlation of results obtained from Nanostring analysis of paired fresh-frozen and FFPE tissues. Scatter plot matrix (left panel, A) for the normalized mRNA transcript quantification values obtained by Nanostring analysis of 19 fresh-frozen vs. FFPE sample pairs (n = 38 samples). In this analysis, the pair-wise Pearson product-moment correlation coefficient was 0.90 ($p < 0.0001$). The right panel (B) shows a heatmap analysis for the Pearson correlation of absolute mRNA transcript abundance as determined by Nanostring, for all pair-wise combinations of samples. These results show a good-high correlation between absolute mRNA transcript quantification data in fresh-frozen vs. FFPE tissues using Nanostring analysis. Fresh-frozen and FFPE tissues are interspersed, and all technical replicates are adjacent in all cases. Gene expression patterns are highly consistent among the large majority of samples.

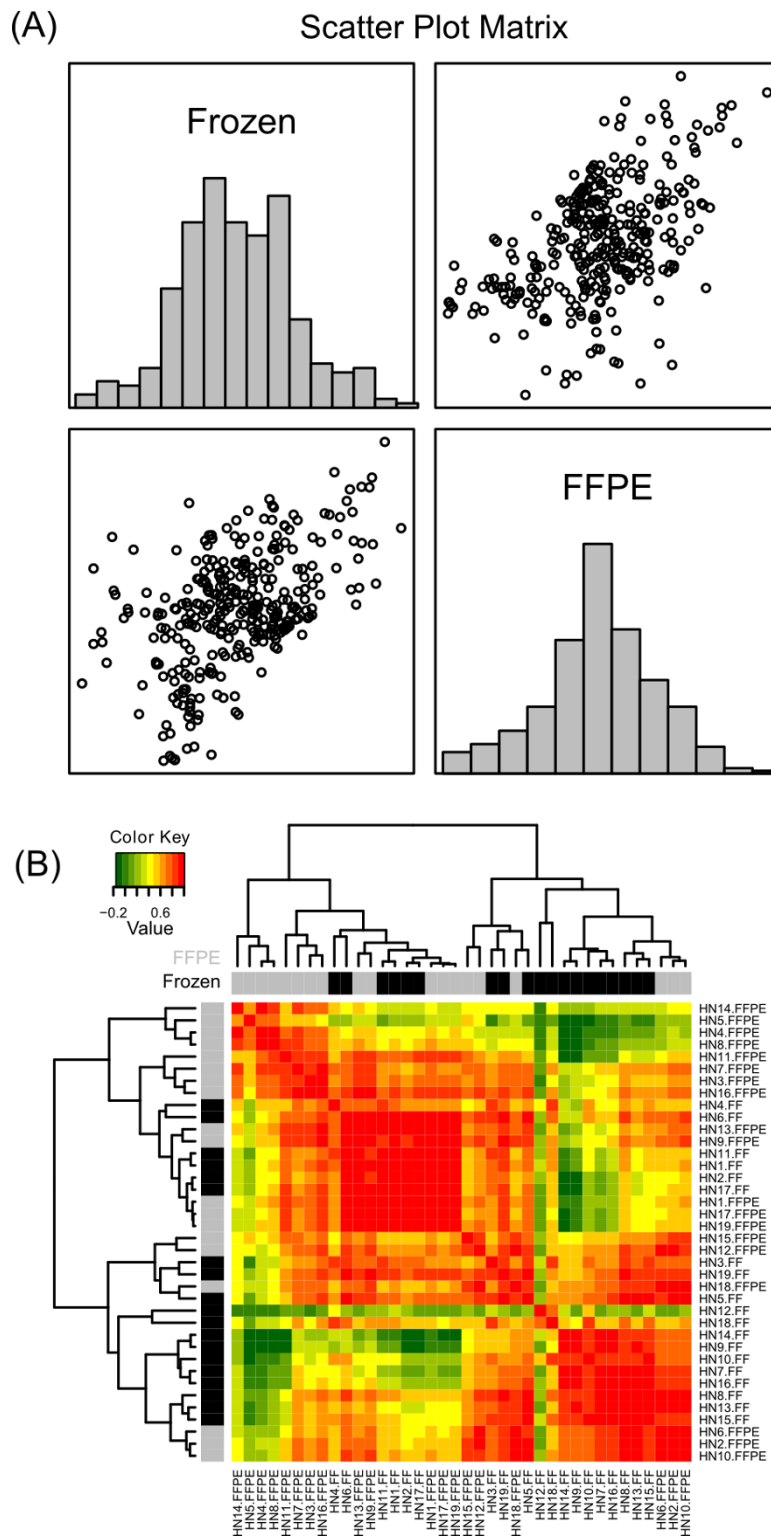


Figure 3 Correlation of results obtained from RQ-PCR analysis of paired fresh-frozen and FFPE tissues. Scatter plot matrix (left panel, A) showing normalized gene expression data obtained by RQ-PCR analysis of the 19 fresh-frozen vs. FFPE sample pairs ($n = 38$ samples). The pair-wise Pearson product-moment correlation coefficient was 0.50 ($p < 0.0001$). The right panel (B) shows a heatmap analysis for the Pearson correlation of gene expression abundance as determined by RQ-PCR, for all pair-wise combinations of samples. A low-moderate correlation is observed between mRNA transcript quantification data in fresh-frozen vs. FFPE tissues, and tissues tend to cluster according to storage method.

Correlation between gene expression levels in fresh-frozen vs. FFPE samples (RQ-PCR)

We also compared gene expression levels determined by RQ-PCR analysis in fresh-frozen *versus* FFPE samples. The overall pair-wise Pearson product-moment correlation coefficient was 0.53 ($p < 0.0001$) (Figure 3A). Heatmap analysis of these data is shown in Figure 3B. A sample-by-sample (fresh-frozen/FFPE sample pair) correlation analysis of RQ-PCR data revealed a mean correlation of 0.54, varying between 0.12 and 0.99, with the majority of sample pairs (12/19) showing a correlation ≥ 0.50 .

Comparison of mRNA quantification data using NanoString versus RQ-PCR

Since all RNA samples isolated from FFPE tissues were degraded, as confirmed by Bioanalyzer analysis, we expected that a probe-based assay would generate more accurate gene expression quantification data compared to amplification-based assays, such as RQ-PCR.

For each sample type (fresh-frozen or FFPE), we compared mRNA transcript quantification as determined by NanoString analysis and gene expression levels as determined by RQ-PCR. For fresh-frozen tissues, this comparison analysis showed that the overall pair-wise Pearson product-moment correlation coefficient was 0.78 ($p < 0.0001$). Figure 4A shows the scatter plot for the $\text{Log}(\text{NanoString})$ *vs.* $\text{Log}(\text{QPCR})$ and their histogram in fresh-frozen tissues. This same analysis in FFPE samples showed a lower overall correlation coefficient of 0.59 ($p < 0.0001$); 11/19 FFPE sample pairs showed a correlation ≥ 0.60 . Figure 4B shows the scatter plot for the $\text{Log}(\text{NanoString})$ *vs.* $\text{Log}(\text{QPCR})$ and their histogram in FFPE tissues. Unsupervised hierarchical clustering analysis of these data was performed and corresponding heatmaps are shown in Figure 4C and 4D.

Discussion

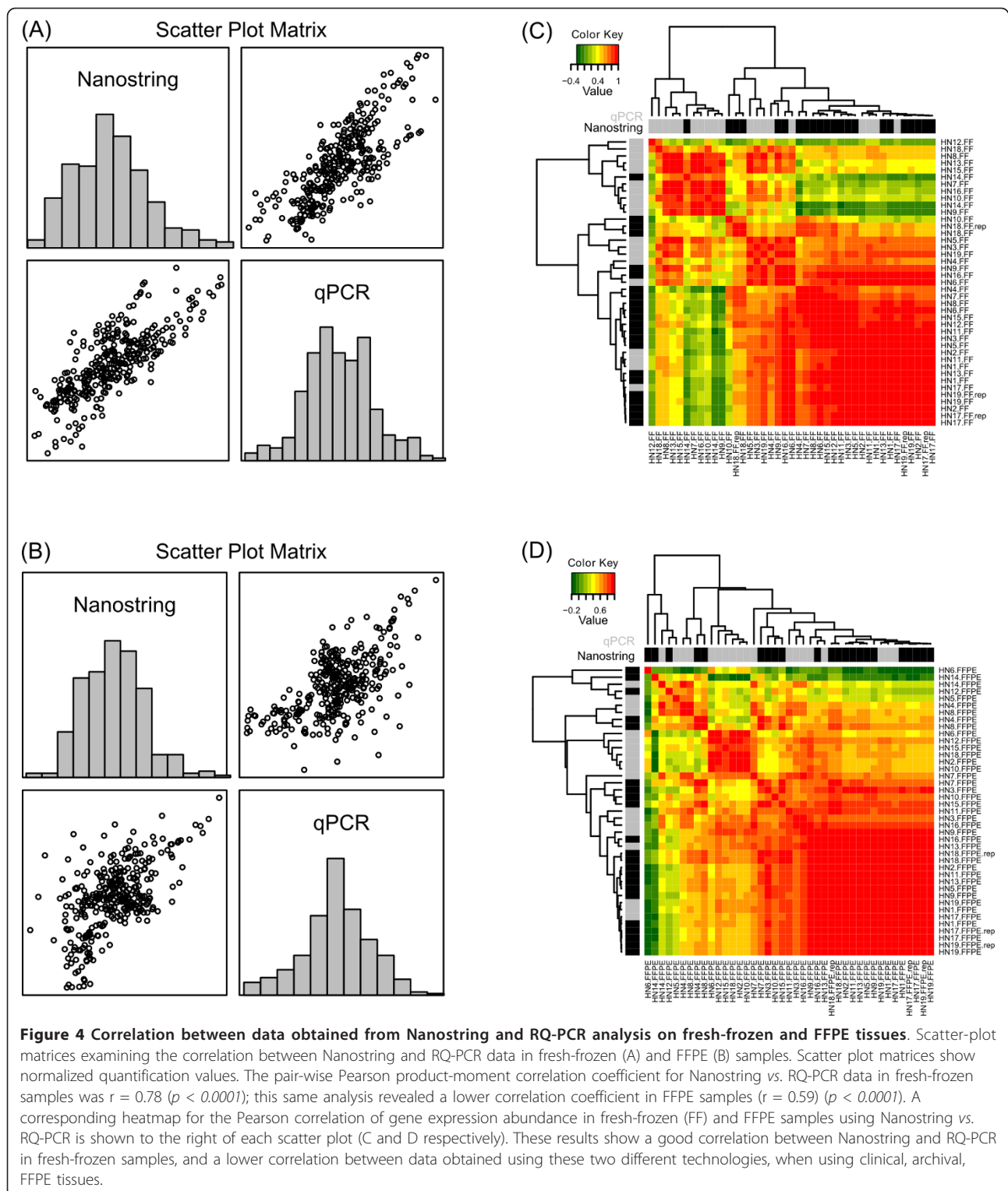
In this pilot study, we showed that NanoString technology is suitable for accurately detecting and measuring mRNA transcript levels in clinical, archival, FFPE oral carcinoma samples. Our results demonstrated that this probe-based assay (NanoString) achieved a good overall Pearson correlation when we compared mRNA transcript quantification results between paired fresh-frozen and FFPE samples. In addition, correlation coefficients were determined in a sample-by-sample comparison, and results showed that mRNA levels in single sample pairs (fresh-frozen and FFPE) were maintained across the sample pairs when using NanoString technology. When we compared gene expression levels obtained by RQ-PCR, we obtained a lower overall correlation coefficient between fresh-frozen and FFPE tissues, and across sample pairs. These results suggest that mRNA transcript levels are more concordant between fresh-frozen and FFPE sample pairs when using NanoString technology.

A recently published study [13] evaluated the performance of quantitative real-time PCR using TaqMan assays (TaqMan Low Density Arrays platform), for gene expression analysis using paired fresh-frozen and FFPE breast cancer samples. The investigators found a good overall correlation coefficient of 0.81 between fresh-frozen and FFPE samples; however, when they compared individual sample pairs, they found a low correlation of 0.33, with variability of 0.005-0.81. These authors suggested that the extensive RNA sample degradation in FFPE samples is likely the cause for the low correlation coefficients observed across sample pairs [13]. Indeed, Bioanalyzer results for our samples showed that fresh-frozen tissues had a good quality RIN and were suitable for gene expression analysis, while FFPE tissues were degraded and had a low RNA integrity number. This RNA degradation in FFPE samples also resulted in higher Ct values initially detectable by RQ-PCR, with loss of amplifiable templates. The low RIN characteristic of FFPE samples did not seem to have an effect on the efficiency of NanoString results, however, when we compared quantification values obtained using RNA isolated from fresh-frozen *vs.* FFPE tissues.

Although quantitative PCR-based assays have been used for gene expression analysis in FFPE samples [13-15], these assays do carry some disadvantages, such as the need for optimization strategies aimed at reducing amplification bias and increasing the number of detectable amplicons when using RNA extracted from FFPE samples. To date, some of the recommended strategies include optimization of the RNA extraction method and designing primers able to detect short amplicons [16]. In our study, primers for RQ-PCR experiments yielded amplicon lengths between 72-170 bp (as detailed in Table 2). Only 2/19 primer pairs yield amplicons >110 bp in size. Such short amplicons are well-suited for PCR amplification using FFPE samples. Our results showed that, although we did obtain gene expression data using RQ-PCR in our FFPE samples, both the overall and the sample-by-sample correlation between fresh-frozen and FFPE samples was notably lower for RQ-PCR data than for data obtained using NanoString. This suggests that this newly developed technology, NanoString nCounterTM, offers advantages over RQ-PCR for gene expression analysis in archival FFPE samples.

Conclusions

We found that the multiplexed, color-coded probe-based method (NanoString nCounterTM) achieved superior gene expression quantification results when compared to RQ-PCR, when using total RNA extracted from clinical, archival, FFPE samples. Such technology could thus be very useful for applications requiring the use of



clinical archival material, such as large scale validation of gene expression data generated by microarrays for generation of tissue specific gene expression signatures.

List of abbreviations

Ct: cycle threshold; FFPE: formalin fixed, paraffin embedded; H&E: hematoxylin and eosin; M-MLV RT enzyme: Moloney Murine Leukemia Virus reverse transcriptase enzyme; PCR: polymerase chain reaction; RIN: RNA integrity number; RQ-PCR: Quantitative real-time PCR; SAS: Statistical analysis system; SDS: Sequence Detection System;

Acknowledgements

RSG is a Research Fellow of The Terry Fox Foundation through an award from the Canadian Cancer Society Research Institute; received funding through a grant from the Canadian Institute of Health Research (Reference No. STP-53912), and through a University of Toronto Fellowship. This study was performed with funds from the Ontario Institute for Cancer Research (OICR).

Author details

¹Division of Applied Molecular Oncology, Princess Margaret Hospital, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada. ²Dept. of Computer Science, University of Toronto, Toronto, ON, Canada. ³Ontario Cancer Institute and the Campbell Family Institute for Cancer Research, Toronto, ON, Canada. ⁴Dept. of Medical Biophysics, University of Toronto, Toronto, ON, Canada. ⁵Dept. of Biostatistics, Princess Margaret Hospital, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada. ⁶Dept. of Pathology, Toronto General Hospital, Ontario Cancer Institute, Toronto, ON, Canada. ⁷Dept. of Otolaryngology/Surgical Oncology, Princess Margaret Hospital, Ontario Cancer Institute, University Health Network, Toronto, ON, Canada. ⁸Dept. of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada.

Authors' contributions

PPR, RSG and SKR designed the study. LW and WX performed statistical data analyses. PPR wrote the manuscript. PPR and RSG performed data interpretation. YX performed quantitative PCR assays. RSG, LW, WX, IJ and SKR edited the manuscript. BPO performed histopathological analyses and evaluated all samples. PG and JI provided samples and reviewed the manuscript. All authors read and approved the final manuscript.

Received: 17 January 2011 Accepted: 9 May 2011 Published: 9 May 2011

References

1. von Ahlfen S, Missel A, Bendrat K, Schlumpberger M: **Determinants of RNA quality from FFPE samples.** *PLoS One* 2007, **2**(12):e1261.
2. Masuda N, Ohnishi T, Kawamoto S, Monden M, Okubo K: **Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples.** *Nucleic Acids Res* 1999, **27**(22):4436-4443.
3. Bresters D, Schipper ME, Reesink HW, Boeser-Nunnink BD, Cuyper HT: **The duration of fixation influences the yield of HCV cDNA-PCR products from formalin-fixed, paraffin-embedded liver tissue.** *J Virol Methods* 1994, **48**(2-3):267-272.
4. Macabeo-Ong M, Ginzinger DG, Dekker N, McMillan A, Regezi JA, Wong DT, Jordan RC: **Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses.** *Mod Pathol* 2002, **15**(9):979-987.
5. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, Fell HP, Ferree S, George RD, Grogan T, et al: **Direct multiplexed measurement of gene expression with color-coded probe pairs.** *Nat Biotechnol* 2008, **26**(3):317-325.
6. Reis PP, Bharadwaj RR, Machado J, Macmillan C, Pintilie M, Sukhai MA, Perez-Ordóñez B, Gullane P, Irish J, Kamel-Reid S: **Claudin 1 overexpression increases invasion and is associated with aggressive histological features in oral squamous cell carcinoma.** *Cancer* 2008, **113**(11):3169-3180.
7. Cervigne NK, Reis PP, Machado J, Sadikovic B, Bradley G, Galloni NN, Pintilie M, Jurisica I, Perez-Ordóñez B, Gilbert R, et al: **Identification of a**

- microRNA signature associated with progression of leukoplakia to oral carcinoma. *Hum Mol Genet* 2009, **18**(24):4818-4829.
8. Dos Reis PP, Bharadwaj RR, Machado J, Macmillan C, Pintilie M, Sukhai MA, Perez-Ordóñez B, Gullane P, Irish J, Kamel-Reid S: **Claudin 1 overexpression increases invasion and is associated with aggressive histological features in oral squamous cell carcinoma.** *Cancer* 2008, **113**(11):3169-3180.
9. Reis PP, Tomenson M, Cervigne NK, Machado J, Jurisica I, Pintilie M, Sukhai MA, Perez-Ordóñez B, Grenman R, Gilbert RW, et al: **Programmed cell death 4 loss increases tumor cell invasion and is regulated by miR-21 in oral squamous cell carcinoma.** *Mol Cancer* 2010, **9**:238.
10. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods* 2001, **25**(4):402-408.
11. Rodgers JL, Nicewander WA: **Thirteen ways to look at the correlation coefficient.** *The American Statistician* 1988, **42**(1):59-66.
12. R Development Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria; 2008.
13. Sanchez-Navarro I, Gamez-Pozo A, Gonzalez-Baron M, Pinto-Marin A, Hardisson D, Lopez R, Madero R, Cejas P, Mendiola M, Espinosa E, et al: **Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen and formalin-fixed, paraffin-embedded breast cancer tissues.** *Biotechniques* 2010, **48**(5):389-397.
14. Cronin M, Pho M, Dutta D, Stephens JC, Shak S, Kiefer MC, Esteban JM, Baker JB: **Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay.** *Am J Pathol* 2004, **164**(1):35-42.
15. Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, et al: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**(27):2817-2826.
16. Antonov J, Goldstein DR, Oberli A, Baltzer A, Pirotta M, Fleischmann A, Altermatt HJ, Jaggi R: **Reliable gene expression measurements from degraded RNA by quantitative real-time PCR depend on short amplicons and a proper normalization.** *Lab Invest* 2005, **85**(8):1040-1050.

doi:10.1186/1472-6750-11-46

Cite this article as: Reis et al.: mRNA transcript quantification in archival samples using multiplexed, color-coded probes. *BMC Biotechnology* 2011 11:46.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

