Successful Application of Microarray Technology to Microdissected Formalin-Fixed, Paraffin-Embedded Tissue

Renata A. Coudry,* Sibele I. Meireles,*
Radka Stoyanova,[†] Harry S. Cooper,[†]
Alan Carpino,[‡] Xianqun Wang,[‡]
Paul F. Engstrom,* and Margie L. Clapper*

From the Divisions of Population Science* and Medical Science,[†]
Fox Chase Cancer Center, Philadelphia, Pennsylvania; and
Arcturus Bioscience,[‡] Sunnyvale, California

The establishment of a reliable method for using RNA from formalin-fixed, paraffin-embedded (FFPE) tissue would provide an opportunity to obtain novel gene expression data from the vast amounts of archived tissue. A custom-designed 22,000 oligonucleotide array was used in the present study to compare the gene expression profile of colonic epithelial cells isolated by laser capture microdissection from FFPE-archived samples with that of the same cell population from matched frozen samples, the preferred source of RNA. Total RNA was extracted from FFPE tissues, amplified, and labeled using the Paradise Reagent System. The quality of the input RNA was assessed by the Bioanalyzer profile, reverse transcriptase-polymerase chain reaction, and agarose gel electrophoresis. The results demonstrate that it is possible to obtain reliable microarray data from FFPE samples using RNA acquired by laser capture microdissection. The concordance between matched FFPE and frozen samples was evaluated and expressed as a Pearson's correlation coefficient, with values ranging from 0.80 to 0.97. The presence of ribosomal RNA peaks in FFPE-derived RNA was reflected by a high correlation with paired frozen samples. A set of practical recommendations for evaluating the RNA integrity and quality in FFPE samples is reported. (*J Mol Diagn* 2007, 9:70–79; DOI: 10.2353/jmoldx.2007.060004)

Recent advances in cDNA microarray technology have revolutionized our ability to comprehensively evaluate the genomic expression profile of a specific tissue type. Because the quality of the input RNA influences the reliability and amount of valuable data that can be obtained from the resulting arrays, snap-frozen tissue continues to be the preferred source of RNA.¹ However, fresh-frozen tissues do not provide sufficient morphological detail for accurate clinical diagnosis.² Pres-

ently, all biopsies and the majority of surgical resection specimens taken for routine diagnostic and staging purposes exist only as formalin-fixed, paraffin-embedded (FFPE) specimens. Thus, the extent to which valuable clinical data, outcome, and histopathological alterations can be correlated with comprehensive gene expression profiles has been compromised severely by the exclusion of routine FFPE specimens from microarray analysis. Limited success in extracting high-quality total RNA from FFPE tissues³ has been attributed to the ability of formalin to crosslink RNA and proteins as well as the addition of monomethylol groups (CH₂OH) to the bases, processes known to interfere with reverse transcription and amplification reactions.⁴ Although it has been demonstrated that RNA isolated from FFPE specimens can be used to perform molecular analyses such as detection of viral nucleic acids in human tissues and quantitative determination of mRNA levels through real-time polymerase chain reaction (PCR), high-throughput microarray analysis of archived samples using commercially available platforms has not been reported.^{5–8} To date, two studies demonstrate that the expression of cancer-related genes can be evaluated in FFPE samples using targeted arrays containing less than 500 genes.^{9,10} Successful establishment of a reliable protocol for the use of archived tissue for expression analyses will support valuable retrospective studies by providing an opportunity to correlate histopathological features, treatment response, prognosis, and survival with biological data.

A new RNA processing system (Paradise Reagent System; Arcturus Bioscience), which allows robust isolation and amplification of RNA from FFPE samples, has been developed. Two primary advantages of this methodology over other published protocols for gene

Supported by grant CA-06927 from the National Cancer Institute and by an appropriation from the Commonwealth of Pennsylvania.

Accepted for publication August 30, 2006.

R.A.C. and S.I.M. are joint first authors.

The contents of this work are solely the responsibility of the authors and do not necessarily represent the official views of the National Cancer Institute.

Address reprint requests to Sibele I. Meireles, Ph.D., Division of Population Science, Fox Chase Cancer Center, 333 Cottman Ave., Philadelphia, PA 19111. E-mail: sibele.meireles@fccc.edu.

expression profiling of RNA from FFPE tissue^{9,10} are its adaptation to the analysis of severely limited amounts of homogeneous cells (as little as 5 ng of total RNA) as acquired by laser capture microdissection (LCM) and the combined use of linear amplification to generate a sufficient amount of RNA for microarray analyses. When compared with RNA amplified by PCR, the RNA amplified linearly using T7 RNA polymerase and oligo(dT) primers exhibits both a greater extent of amplification and a stronger correlation with the expression profile of nonamplified samples.^{11,12} Although reports have begun to emerge on the successful use of the Paradise Reagent System for gene expression profiling of FFPE tissue by quantitative real-time PCR,¹³ an in-depth comparison of microarray data generated from FFPE tissue with that obtained from frozen specimens has not been performed and is essential to validate this promising methodology.

The goal of the present study is to assess the feasibility of obtaining reliable microarray data from FFPE samples collected from resection specimens obtained during routine surgical procedures. This evaluation was accomplished by comparing the gene expression profiles of FFPE samples with those of matched frozen tissue samples obtained from the same patient, the latter representing an optimal source of material for microarray analysis. A further objective is to establish predictive markers, such as the integrity and size of the total RNA obtained from FFPE tissue, as indicators of the anticipated quality of downstream microarray data.

This report demonstrates that it is possible to obtain credible microarray data from FFPE tissue, even when a small amount of RNA, acquired by LCM, is used. A set of practical recommendations for evaluating RNA integrity and quality control, which can guide the implementation of the developed technology in any laboratory setting, has been established. Because formalin fixation-paraffin embedding is the universal protocol for routine pathological diagnosis, the proposed procedure represents an invaluable tool with which to correlate the genomic characteristics of a sample with the wealth of morphological and clinical information available for the same specimen.

Materials and Methods

Tissue Specimens

Matched FFPE and frozen samples of human colon adenocarcinomas were collected from patients at the time of colectomy at the Fox Chase Cancer Center according to an institutional review board-approved protocol and stored in the Department of Pathology at Fox Chase Cancer Center for up to 5 years. Colon was selected as a prototypic tissue for analysis based on its high content of endogenous RNases, proteins known to expedite the degradation of RNA before tissue fixation in formalin. Protocols that yield high-quality RNA under these challenging conditions are anticipated to be applicable to most tissue types. To evaluate the effect of tissue autolysis on RNA integrity, four speci-

mens were collected under closely monitored conditions. Immediately upon arrival in the Pathology Laboratory from surgery, each specimen was divided into four pieces (0.5 to 1.0 cm² in size). One piece was snap-frozen in OCT compound (optimal cutting temperature), and one piece was fixed in 10% formalin. Two pieces were allowed to sit on the "counter top" for 2 hours, after which one piece was snap-frozen in OCT, and the other piece was fixed in 10% formalin. The formalin-fixed tissues were placed in the tissue processor the same day, processed routinely, and embedded in paraffin the next morning. FFPE tissue blocks and frozen specimens were stored at room temperature and at -80°C, respectively.

To evaluate the gene expression profile using oligonucleotide arrays, five pairs of frozen and FFPE specimens were obtained from the archives of the Department of Pathology. Four of the formalin-fixed surgical specimens (subjects 26, 27, 62, and 68) were fixed within 45 to 90 minutes upon arrival from the operating room and placed in the tissue processor after 7 to 28 hours of formalin fixation, as per the routine departmental protocol. The sample from subject 201 was fixed in formalin immediately upon delivery from the operating room and processed within 24 hours of fixation. All frozen specimens were immediately placed in liquid nitrogen upon arrival in the Department of Pathology and subsequently stored at -80°C.

Tissue Sectioning and LCM

Frozen sections (5 µm thickness) were cut at -20°C and immediately transferred to a microslide box kept on dry ice and stored at -80°C. FFPE sections were cut (5 µm thickness), floated in an RNase-free water bath, and transferred to glass slides. These slides were placed under the hood to dry for 1 hour and subsequently stored in a microslide box at -80°C. A new blade was used for each frozen and FFPE sample.

To obtain a homogeneous population of epithelial cells for analysis, LCM was performed using tissue sections from both FFPE and frozen samples. Immediately before LCM, the slide box was taken from the freezer and placed on dry ice. FFPE samples were deparaffinized, stained using Histogene Solution (Arc-turus Bioscience), and dehydrated; 5000 malignant cells were laser capture microdissected from colon tumors using the automated AutoPix System (Arcturus Bioscience). The process described above was also followed for frozen tissues, except for the deparaffinization step. Representative images of the LCM procedure are presented in Figure 1.

RNA Isolation and Amplification

FFPE tissue sections were processed by adding RNA extraction buffer (30 µl) (Paradise Reagent System; Arc-turus Bioscience) directly to a small region of the tissue section, scraping the cells, and transferring the lysate to a microcentrifuge tube. For microdissected specimens,

Figure 1. Representative LCM images of colonic malignant epithelial cells captured from H&E-stained sections. **a:** Images of the tissue before capture. **b:** Cells marked for capture. **c:** Pure population of captured cells in the cap. **d:** Remaining tissue after capture.

the caps containing cells captured from FFPE samples were placed into a microcentrifuge tube containing 30 μ l of RNA extraction buffer. All samples were incubated overnight at 50°C according to the manufacturer's protocol. RNA isolation was performed using the MiraCol Purification Column according to the manufacturer's protocol (Arcturus Bioscience) and eluted in 12 μ l of elution buffer. Total RNA was isolated from microdissected cells and scraped tissue of matched frozen specimens using the PicoPure RNA Isolation kit (Arcturus Bioscience). All RNAs were subjected to DNase treatment during the RNA isolation step following the manufacturer's recommended protocol. The quality and yield of the resulting total RNAs were assessed by loading 1 μ l of the extracted RNA on an RNA 6000 Pico LabChip and evaluating it on an Agilent Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA) using the Eukaryote Total RNA Pico Assay.

For oligonucleotide microarray analyses, total RNA (11 μ l) from five matched pairs of FFPE and frozen microdissected tissue was subjected to two rounds of linear RNA amplification based on T7 polymerase *in vitro* transcription. The first round of amplification was performed using the Paradise Reagent System and the RiboAmp OA kit (Arcturus Bioscience) for FFPE and matched frozen samples, respectively. The second round of amplification was performed subsequently using the Paradise Reagent System for both FFPE and matched frozen samples. A total of 10 (five FFPE and five frozen) labeled antisense (aRNA) probes for microarray hybridization were generated during the second round of amplification in the presence of 5-[3-aminoallyl] uridine 5'-triphosphate (UTP) after indirect incorporation of Cy3 fluorescent dye (Amersham Biosciences Corp., Piscataway, NJ). One hundred nanograms of Human Universal Reference total RNA (BD Biosciences, Palo Alto, CA) was amplified in parallel following the same protocol used for RNA derived from frozen samples and labeled with Cy5 fluorescent dye (Amersham Biosciences Corp.). The concentration of the labeled aRNA (picomoles per microliter) was deter-

mined using an ND-1000 Spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE).

Oligonucleotide Microarrays

Microarray hybridizations were performed using a custom-designed 22,000 oligonucleotide (60-mer) array commercially available from Agilent Technologies. The microarray also contains 314 genes that serve as negative controls. This array was designed to contain probes preferentially located in the 3' end of the transcripts. Cy3-labeled aRNA (75 pmol for FFPE and 55 pmol for frozen samples) was mixed with Cy5-labeled reference aRNA (55 pmol) and hybridized onto the microarray in an Agilent G2545A Hybridization Oven using the *In Situ* Hybridization kit Plus (Agilent Technologies), as described by the manufacturer. The resulting 10 microarrays were scanned using an Agilent DNA Microarray Scanner (model G2565BA).

Data Quantification and Statistical Analysis

Spot and local background intensities of the acquired microarray images were quantified using Feature Extraction Software v.7.5 (Agilent Technologies). The log (base 10) expression ratios of the dye-normalized intensities were calculated by the software. The expression ratio of each gene was determined and expressed as the dye-normalized intensities of the reference sample over the corresponding value for the test (frozen or FFPE) sample. The software also reports the *P* value of the ratios as a measure of the confidence that the gene is differentially expressed in the comparison between reference (Cy5 channel) and test samples (Cy3 channel) (Agilent Feature Extraction Software v.7.5 User Guide).

The subsequent analyses included two steps: 1) determination of the *expressed* genes and 2) assessment of the reproducibility and differences between the expression profiles of FFPE and frozen specimens. The intensity

Table 1. PCR Primers and Conditions Used to Evaluate the Length of aRNA

Gene name/accession no.	Primers	Forward/reverse 5'-3'	Melting temperature (°C)	Distance from poly-A tail (bp)	Amplicon size (bp)
β -Actin/NM_001101	HBAC1650	tccccaacttgagatgatgaag aactgggtctcaagtcaggtacagg	60	136	92
	hbac1355	atccccaaggtcacaatg gtggcttttaggatggcaag	60	407	118
Keratin-20/NM_019010	KRT20_1	gccatcttatcatgaagcac catggattcctactttgtgag	60	168	127
	KRT20_2	gctgctgaggtttgaaaga cccacccttctaatactg	60	398	109
	KRT20_3	aacgccagaacaacgaatac cttcagggtgcttaactga	60	653	132
<i>N</i> -Acetyltransferase-1/NM_000662	NAT-1	gacgacctatcatgtat ctagcataatcaccaa	50	302	214

of each spot (both the Cy3 and Cy5 channels) was first filtered for good quality, based on the default parameters provided by the Feature Extraction Software. Briefly, a spot was defined as *good* if it was not flagged as a saturation and nonuniformity outlier and was positive [significantly above background (Agilent Feature Extraction Software v.7.5 User Guide)] in a single channel. A gene was defined as *expressed* ("E") if it was represented by a *good* spot in both the Cy3 and Cy5 channels, and the *P* value of the ratio was below 0.05. Otherwise, a gene was considered *nonexpressed* ("N"). The percentage of non-control genes ($N = 21,939$) *expressed* was calculated as

a measure of chip performance. In addition, the percentage of *expressed* genes detected in negative controls ($N = 314$) was used to identify potential false-positive signals in each microarray.

The reproducibility between matched frozen and FFPE samples was assessed based on 1) the Pearson's correlation coefficients and Spearman's rank order correlation coefficients using the expression ratios of the *expressed* genes, in matched frozen and FFPE samples and 2) the concordance scores of *expressed* and *nonexpressed* genes in matched frozen and FFPE samples.

Interval of processing

Figure 2. Comparative analysis of total RNA obtained by LCM and from scraped tissue. One microliter of RNA extracted from frozen or FFPE tissue was loaded on a Bioanalyzer Pico RNA Chip and fractionated in an Agilent Bioanalyzer 2100. 18S and 28S ribosomal RNA peaks are observed in all samples, and no difference is observed between laser capture microdissected (large panels) and scraped specimens (small panels). RNA recovered from FFPE samples exhibits a more heterogeneous profile, and ribosomal RNA peaks are broadened. Samples were processed immediately (0 hour) on arrival from surgery or after a delay of 2 hours. There is a discrete level of RNA degradation in tissues where freezing was delayed 2 hours.

Figure 3. Electropherogram of total RNA obtained from cells following LCM. RNA extracted from LCM FFPE or frozen tissue was loaded on a Bioanalyzer Pico RNA Chip and fractionated in an Agilent Bioanalyzer 2100. Sample number and year of collection are indicated on the left side. 18S and 28S ribosomal RNA peaks are observed in RNA derived from all frozen tissues and also in sample 201-FFPE. In contrast, there is a lack of ribosomal peaks in most of the FFPE samples. The profile in the FFPE samples varies from a sharp peak to a large plateau; the first indicates a prevalence of small size RNA fragments (samples 26 and 27), and the second indicates a greater variety of fragment sizes (samples 201, 62, and 68).

Reverse Transcription (RT)-PCR

To evaluate the length of the aRNA, which reflects the extent of degradation of the input total RNA, primers specific for

human β -actin (HBAC), keratin-20 (KRT20), and *N*-acetyltransferase 1 (NAT-1) were designed to target different regions of the mRNA sequence. The primer sequences and corresponding details are found in Table 1.

aRNA (250 ng) was reverse transcribed with random primers using the Superscript First-Strand Synthesis System (Invitrogen Corp., Carlsbad, CA). Two microliters of the resulting reaction mixture was then used as a template for PCR amplification. PCR was performed under standard conditions [25- μ l reaction volume, 1 $\times$ PCR buffer, 1.5 mmol/L MgCl₂, 0.2 mmol/L dNTPs, 0.2 μ mol/L of each primer (sense and antisense), and 1 U of Platinum TaqDNA Polymerase (Invitrogen Corp.)]. PCR cycles were as follows: 94°C for 5 minutes; 35 cycles at 94°C for 30 seconds, 50 or 60°C (according to the melting temperatures for the primers shown in Table 1) for 30 seconds, 72°C for 30 seconds, and a final cycle of 72°C for 10 minutes. Products were resolved using standard 2% agarose gel electrophoresis and visualized by ethidium bromide staining under UV light.

Results

Quality of RNA Isolated from Matched FFPE and Frozen Tissue (LCM versus Tissue Scraping)

The quality of total RNA was assessed by examining the shape of the Bioanalyzer electropherograms. A comparative analysis of total RNA obtained from LCM and scraping from both frozen and FFPE tissues is indicated in Figure 2. 18S and 28S ribosomal RNA peaks are observed in all frozen and FFPE samples, even though the profile of RNA recovered from FFPE samples is more heterogeneous. Both the broadening of ribosomal RNA peaks and the presence of fragmentation products between the 18S ribosomal subunit and the control marker (first peak in the electropherogram) indicate that RNA quality is affected by tissue fixation. In all instances, the quality of RNA extracted from cells isolated by LCM (Figure 2, large panels) was identical to that of scraped specimens (Figure 2, small panels). No significant differences were observed with respect to RNA quality when the tissues were processed immediately (0 hours) or after 2 hours. However, a change in the baseline between the marker and the 18S ribosomal peak was observed in frozen samples processed after 2 hours, indicating partial RNA degradation.

Figure 3 shows the RNA quality of five paired samples obtained from the Fox Chase Cancer Center archives that were subjected to LCM and gene expression analyses. Similar to Figure 2, RNA extracted from epithelial cells of frozen tissue showed 28S and 18S ribosomal RNA peaks, indicative of good-quality total RNA. In contrast, the profiles of RNA isolated from FFPE tissue appeared more heterogeneous than those presented in Figure 2. The electropherograms included a single peak (samples 26 and 27) or a plateau (samples 62 and 68), except for sample 201-FFPE, where ribosomal RNA peaks were present.

Evaluation of aRNA

The yield of aRNA after two rounds of amplification was not significantly different between FFPE and frozen samples (12 to 72 and 20 to 53 μ g, respectively) (Figure 4).

Figure 4. Evaluation of aRNA. **A:** One microgram of aRNA derived from FFPE/frozen samples and Human Reference RNA BD Clontech (Ref), all used for microarray hybridizations, was fractionated on a 2% agarose gel stained with ethidium bromide. The average length observed for frozen tissue-derived aRNAs was slightly longer (~500 bases) than that derived from FFPE tissue (~400 bases). Sample 201-FFPE, which exhibited ribosomal RNA peaks, shows the highest aRNA length among the other FFPE samples. The yield obtained for each aRNA is indicated below each lane. **B:** RT-PCR was used to evaluate the fragment length of aRNA from matched frozen and FFPE samples (24, 26, 27, 62, and 68). A control RNA provided in the Paradise system (**C**) was also evaluated. aRNA (250 ng) was reverse transcribed and amplified using primers specific for β -actin (HBAC), N-acetyltransferase 1 (NAT-1), and cytokeratin 20 (KRT20) and located within 700 bp of the 3' region of the mRNA sequence. The position of each amplicon relative to the mRNA sequence is pictured at the bottom.

However, the length of the aRNA, as determined by agarose gel electrophoresis, was slightly smaller for FFPE samples compared with frozen-derived aRNAs (maximum length 500 and 400 nucleotides, respectively). The length of the aRNA generated from sample 201-FFPE, which was the only FFPE RNA preserving ribosomal RNA peaks, was comparable to that of aRNA derived from the matched frozen specimen.

RT-PCR was also used to evaluate the length of the aRNA. The amplicons generated from six primer sets targeting different portions of the HBAC, KRT20, and NAT-1 mRNA sequences are shown in Figure 4B. The presence or absence of an amplicon is a qualitative measure of the length of the cDNA template. cDNA from all frozen and most FFPE samples was amplified successfully using these primer sets. No product was generated when FFPE samples 26 and 27 were subjected to amplification with primers KRT20_3 and NAT-1.

Microarray Data Analysis

The percentage of false positives and *expressed* genes (as defined in Materials and Methods) along with the

Table 2. Summary of Microarray Performance Using Matched FFPE and Frozen Samples

Sample	False positive* (%)	Expressed genes [†]		
		Total (%)	Mean signal Cy5 channel	Mean signal Cy3 channel
201 FFPE	0	56.08	6099	6105
201 Frozen	0	51.98	5980	6251
62 FFPE	0	45.23	5605	5514
62 Frozen	1.59	40.12	6509	6678
68 FFPE	0	36.57	7135	7140
68 Frozen	0	48.35	7240	7247
27 FFPE	0	27.06	6484	6735
27 Frozen	0	24.66	8529	8834
26 FFPE	1.27	28.42	7586	7740
26 Frozen	0	41.42	6565	6574

*Percentage of negative control genes ($n = 314$) detected as expressed genes.

[†]Percentage of expressed genes (from a total of 21,939) represented in the microarray along with dye-normalized mean signal intensities in the Cy5 and Cy3 channels.

dye-normalized mean signal intensities in both channels are summarized in Table 2. The false-positive rate (ie, the percentage of negative control genes detected as *expressed*) is insignificantly low across all hybridizations. The percentage of genes *expressed* was very similar in matched frozen and FFPE samples—differing by approximately 5% in samples 27, 62, and 201 and by 13 and 12% in samples 26 and 68, respectively.

To determine whether the same gene could be detected in matched FFPE and frozen samples, the concordance of *expressed* (“EE”) and *nonexpressed* (“NN”) genes was calculated for the two types of samples (Table 3). Among the five paired specimens, sample 201 presented the highest score (42.5%) of *expressed* genes and the lowest percentage (34.5%) of *nonexpressed* genes of all samples evaluated.

A high correlation coefficient ($r = 0.80$ – 0.96) was observed between *expressed* genes in frozen and FFPE samples (Table 4). The default P value cutoff used to define *expressed* genes is 0.05 (as described in Materials and Methods). In addition, the impact of different numbers of *expressed* genes on the similarity of gene expression profiles was explored by varying the P value cutoff that was used to define the *expressed* genes. By making the P value cutoff more rigorous (from 0.05 to 0.01 and 0.001), fewer *expressed* genes and slightly higher correlation coefficients were observed. The impact on the correlation coefficients is less for those samples with better input RNA quality (1 to 2% for samples 201 and 62) than those samples with poor input RNA quality (4 to 5% for samples 26 and 27). This observation confirmed that the similarity of gene expression profiles between FFPE and frozen samples is less sensitive to the selection of genes for those samples with better RNA quality. The scatter plots in Figure 5 illustrate the overall distribution of expression ratios using *expressed* genes ($P < 0.05$) in matched FFPE and frozen samples. Samples with lower correlation coefficients (26, 27, and 68) had genes that were differentially *expressed* more than 10-fold in matched specimens. The highest Pearson’s correlation coefficient ($r = 0.96$) was observed for the “EE” in sample 201.

Discussion

Results from the present study indicate that reliable microarray data can be obtained from clinical surgical specimens after routine fixation in formalin (as practiced in surgical pathology laboratories), long-term storage, and subsequent LCM to isolate homogeneous cell populations. Gene expression profiling using oligonucleotide microarrays represents a powerful approach with which to assess the molecular characteristics of both normal and diseased human tissues. To date, high-throughput gene expression profiling has facilitated the elucidation of gene function, accelerated target identification for drug discovery, and enhanced disease diagnosis. Successful application of this technology to diagnostic FFPE specimens will revolutionize our ability to correlate clinical outcomes, medical history, and histopathological findings with comprehensive gene expression profiles.

Based on the establishment of distinct expression profiles for specific cell types, it is no longer optimal to evaluate transcript levels in mixed cell populations.^{1,14} LCM ensures accurate data interpretation by restricting cell collection to only those cells of interest and circumventing contamination from other tissue components.

One of the primary limitations of microarray analysis is the large amount of labeled input RNA (5 to 10 μ g) required for hybridization.¹⁵ Thus, the minute quantities of RNA extracted from LCM material are insufficient for global gene expression profiling, dictating the need for RNA amplification. Moreover, reliable strategies to maintain a high correlation between the expression profile of amplified and nonamplified samples, such as linear amplification based on T7 RNA polymerase and oligo(dT) primers,¹¹ should be used. The Paradise Reagent System (Arcturus Bioscience) had been designed to address both the requirement for a reliable amplification protocol and the compromised quality of FFPE specimens. Extraction of RNA from 5000 colonic epithelial cells, isolated by LCM, yielded as little as 5 ng of total RNA, an amount sufficient to perform microarray analysis after two rounds of linear RNA amplification using the Paradise Reagent System. Although others have evaluated the gene ex-

Table 3. Concordance of the Expression Profile of Matched FFPE and Frozen Samples

Categories	Samples				
	201	62	68	27	26
EE	42.5	30.9	27.1	15.8	22.1
NN	34.5	45.6	42.2	64.1	52.2
NE	9.4	9.2	21.2	8.8	19.3
EN	13.5	14.3	9.4	11.2	6.3

EE, percentage of genes designated as expressed ("E") in both FFPE and frozen; NN, percentage of genes designated as nonexpressed ("N") in both FFPE and frozen; NE and EN, percentage of genes designated as nonexpressed in FFPE and expressed in the frozen sample or vice versa.

pression profile of RNA from FFPE samples using different approaches,^{6,8-10,16} the characteristics that the input RNA must exhibit to support these analyses have not been reported. The present study represents a direct comparison of the quality of total RNA extracted from FFPE and frozen colon tissues after LCM and scraping. The resulting data demonstrate that the integrity of RNA extracted from both frozen and FFPE tissue is not affected by LCM (Figure 2), suggesting that one can reliably use scraped tissue to determine the RNA quality of the sample before LCM, a laborious and expensive procedure. Surprisingly, a 2-hour delay in processing of the surgical specimen in the Surgical Pathology Laboratory did not significantly affect the quality of the RNA extracted from either frozen or fixed specimens (Figure 2, bottom panels). As illustrated in Figures 2 and 3 (right panels), the quality of RNA obtained from FFPE tissue is

Table 4. Correlation Coefficients for Expressed Genes in Matched FFPE and Frozen Samples

Samples	$P < 0.05^*$		$P < 0.01$		$P < 0.001$	
	R [†]	"EE" (%)	R	"EE" (%)	R	"EE" (%)
201	0.96	42.5	0.97	37.0	0.97	31.9
62	0.89	30.9	0.91	25.5	0.92	20.6
68	0.83	27.1	0.85	22.0	0.87	17.3
26	0.86	22.1	0.88	17.2	0.90	13.5
27	0.80	15.8	0.83	12.2	0.85	9.3

*Three cutoff P values of the expression ratio were analyzed ($P < 0.05$, $P < 0.01$, and $P < 0.001$).

[†]Pearson's correlation coefficients of FFPE and frozen samples using log (base 10) expression ratio of the dye-normalized mean signal intensity (Cy5/Cy3).

*Percentage of genes (from a total of 21,939) expressed in both FFPE and frozen samples.

quite variable. Although ribosomal RNA peaks are preserved in some cases, fluorescence is detected by the Bioanalyzer over a longer period of time. This profile is observed especially if the tissue is processed within 2 hours after collection. The highest quality of RNA was derived from FFPE sample 201. This sample, which was subjected to immediate formalin fixation, presented the highest correlation values with the paired frozen specimen. This observation validates the importance of using the Bioanalyzer as a quality control tool. In contrast, ribosomal RNA peaks were absent in the remaining FFPE samples. Extensive variability in the length of the RNA produced a plateau (FFPE samples 62 and 68), whereas the prevalence of short RNA fragments, indicative of

Figure 5. Scatter plots show the correlation based on ratios of genes expressed in both frozen and FFPE tissues. Expressed genes are defined as spots that are not flagged as saturation and nonuniformity outliers and are positive and significantly above background in both the Cy3 and Cy5 channels; the P value of the ratio is less than 0.05. The total number of expressed genes and the correlation value (R) for each pair of matched samples are indicated. Nonexpressed genes are excluded from the plots (shown by the empty area in the center of the scatter plots). Lines correspond to the 45° diagonal line and ± 1 ratio unit lines (10-fold difference between frozen and FFPE samples).

degradation, yielded a sharp peak within the first seconds of the electropherogram run (FFPE samples 26 and 27). Based on these criteria, the FFPE samples can be ranked with respect to RNA quality (highest to lowest): 201, 62, 68, 26, and 27.

The demonstrated success in obtaining high-quality RNA from sample 201-FFPE as well as the samples in Figure 2 can be explained by the fact that this surgical specimen was fixed in formalin immediately (or within a 2-hour interval) after collection in the operating room. An additional five samples collected in a similar manner also showed preservation of the ribosomal subunit peaks (data not shown). Although sample 201 was stored for the shortest time, no correlation was observed between the length of archival storage (1 to 5 years) and RNA quality. For example, sample 62, which was stored for 5 years, showed better RNA quality than samples stored for 4 years (samples 26, 27, and 68). Thus, autolysis resulting from a delay in formalin fixation of the surgical specimen, rather than storage time, seems to be the primary factor affecting the quality of extracted RNA. Comprehensive studies designed to establish the optimal conditions to preserve the RNA quality of archived tissues are necessary but will require several years for completion and will not facilitate the use of samples already archived. In this respect, the present study provides guidelines for the use of archived samples already available.

Data from this study indicate that the length of the resulting aRNA is directly related to the method for preserving the specimen (frozen or FFPE). The maximum aRNA length observed was 500 and 400 nucleotides for frozen and FFPE-derived aRNAs, respectively (Figure 4A). Because of the high quality of total RNA obtained from sample 201-FFPE, the length of its derived aRNA is similar to that observed in frozen samples. Consistent with other reports,^{6,7} the average length of amplified RNA from FFPE specimens is slightly smaller than that obtained from frozen samples. A complementary study to evaluate the length of the aRNA before microarray hybridization was performed using RT-PCR. This reaction assessed the presence or absence of specific portions of the transcripts of HBAC, KRT20, and NAT-1 in both frozen and FFPE-derived RNA. Three of five FFPE samples and all frozen samples were suitable for amplification of fragments up to 214 nucleotides in length and up to 653 nucleotides upstream to the poly-A tail (Figure 4B). The fragment amplified using the KRT20_3 primers requires a longer template for successful annealing (Table 1) and, therefore, is less likely to be detected in low-quality RNAs (such as 26- and 27-FFPE). On the other hand, failure of the NAT-1 fragment to amplify in samples 26- and 27-FFPE is probably due to the fact that this gene is weakly expressed in many tissues, including colon (gene expression data available at Cancer Genome Anatomical Project (<http://cgap.nci.nih.gov/Genes/GenelInfo?ORG=Hs&CID=200738>); and its low levels of mRNA might be difficult to detect in RNAs of compromised quality. It should be noted that the HBAC gene, which was amplified successfully in all samples, is a highly expressed housekeeping gene and should not be used exclusively as a marker of RNA quality. Likewise, the same primer

sets were not informative when used in quantitative real-time PCR (data not shown).

The results from the present study, using an oligonucleotide array representing approximately 22,000 genes, indicate that the microarray performance obtained from laser microdissected FFPE samples is comparable to that derived from matched frozen samples. Furthermore, the similarities are enhanced when ribosomal RNA peaks can be detected in FFPE samples (Figure 3). This result is in agreement with Bibikova et al,¹⁰ who reported a high correlation coefficient ($R^2 = 0.69$) when comparing the level of expression of approximately 500 genes in FFPE and fresh-frozen samples. However, microdissection was not used, and higher amounts of input RNA (50 ng) were used. In contrast, other investigators have reported that FFPE tissues are not a reliable substrate for cDNA synthesis and labeling based on the gene expression profile of 95 relatively abundant human genes in matched FFPE and frozen specimens.⁹

Although a slight loss of sensitivity in signal intensities was observed in the present study when FFPE samples were evaluated (Table 2), a high correlation coefficient and significant concordance between expressed genes in both FFPE and frozen samples were obtained (Tables 3 and 4). The performance of each sample in the microarray hybridization correlated directly with the quality of the input RNA, as assessed by the Bioanalyzer, RT-PCR, and agarose gel electrophoresis. The concordance among the expressed genes in FFPE and frozen tissue, as indicated by the fraction of expressed genes called "EE," reflects the quality of the total RNA derived from FFPE tissue ranked from highest to lowest: samples 201, 62, 68, 26, and 27. Likewise, a gradual decrease in the fraction of "EE" (0.43, 0.31, 0.27, 0.22, and 0.16, respectively) was observed (Table 3). In addition, the high quality of RNA derived from sample 201-FFPE was reflected by the highest correlation coefficient between frozen and FFPE expression ratios ($r = 0.96$) and the highest fraction of "EE" compared with the fraction of nonexpressed genes ("NN"), using the P value of the expression ratio <0.05 . In contrast, sample 27-FFPE exhibited the lowest correlation coefficient ($r = 0.80$) and the highest fraction of "NN" (Tables 3 and 4). Very similar results were observed when different cutoff values were considered for identifying expressed genes (Table 4) and also when Spearman's Rho correlation was used (data not shown), confirming the precision of the analysis.

This study demonstrates that the transcript profile of tissue samples that have been archived in laboratories worldwide can now be evaluated for the first time using the Paradise Reagent System. Subsequent correlation of these genomic data with medical history, histopathological findings, and clinical outcome is anticipated to greatly facilitate the development of highly efficacious interventions for a number of diseases.

Acknowledgments

We thank Dr. Rachel Jones for her helpful scientific discussions in the use of the Paradise System, Drs. José and

Irma Russo for kindly allowing us to use their Agilent microarray equipment, and Dr. Gabriela Balogh for helpful microarray assistance. Special thanks to Drs. Andrew Godwin, Karthik Devarajan, Rajiv Raja, and Parisa Hana-chi for their valuable comments and Maureen Climaldi for her excellent assistance in preparing this manuscript for publication. The following facilities at Fox Chase Cancer Center were used for this project: the LCM Unit; the DNA Microarray Facility; and the Cancer Prevention Biomarker and Genotyping Facility.

References

- Elkahloun AG, Gaudet J, Robinson GS, Sgroi DC: In situ gene expression analysis of cancer using laser capture microdissection, microarrays and real time quantitative PCR. *Cancer Biol Ther* 2002, 1:354–358
- Lehmann U, Kreipe H: Real-time PCR analysis of DNA and RNA extracted from formalin-fixed and paraffin-embedded biopsies. *Methods* 2001, 25:409–418
- Ramaswamy S: Translating cancer genomics into clinical oncology. *N Engl J Med* 2004, 350:1814–1816
- 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, 27:4436–4443
- Soguero C, Ribalta T, Campo E, Sanchez-Tapies JM, Saiz JC, Bruguera M: Detection of hepatitis C virus RNA in more than 20-year-old paraffin-embedded liver tissue. *Lab Invest* 1999, 79:365–366
- Godfrey TE, Kim SH, Chavira M, Ruff DW, Warren RS, Gray JW, Jensen RH: Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5' nuclease quantitative reverse transcription-polymerase chain reaction. *J Mol Diagn* 2000, 2:84–91
- Cronin M, Pho M, Dutta D, Stephans 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, 164:35–42
- Specht K, Richter T, Muller U, Walch A, Werner M, Hofler H: Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am J Pathol* 2001, 158:419–429
- Karsten SL, Van Deerlin VM, Sabatti C, Gill LH, Geschwind DH: An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. *Nucleic Acids Res* 2002, 30:E4
- Bibikova M, Talantov D, Chudin E, Yeakley JM, Chen J, Doucet D, Wickham E, Atkins D, Barker D, Chee M, Wang Y, Fan JB: Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am J Pathol* 2004, 165:1799–1807
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH: Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* 1990, 87:1663–1667
- Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM: High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* 2000, 18:457–459
- Ma XJ, Wang Z, Ryan PD, Isakoff SJ, Barmettler A, Fuller A, Muir B, Mohapatra G, Salunga R, Tuggle JT, Tran Y, Tran D, Tassin A, Amon P, Wang W, Enright E, Stecker K, Estepa-Sabal E, Smith B, Younger J, Balis U, Michaelson J, Bhan A, Habin K, Baer TM, Brugge J, Haber DA, Erlander MG, Sgroi DC: A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* 2004, 5:607–616
- Lechpammer M, Sgroi DC: Laser capture microdissection: a rising tool in genetic profiling of cancer. *Expert Rev Mol Diagn* 2004, 4:429–430
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. *Nat Genet* 1999, 21:20–24
- Resnick MB, Sabo E, Meitner P, Kim SS, Cho Y, Kim H, Tavares R, Moss SF: Global analysis of the human gastric epithelial transcriptome altered by *Helicobacter pylori* eradication in vivo. *Gut* 2006, 55:1717–1724