RNA expression analysis of formalin-fixed paraffin-embedded tumors

Shannon K Penland[1,2], Temitope O Keku[1,3], Chad Torrice[1,2,4], Xiaping He[2,4], Janakiraman Krishnamurthy[1,2,4], Katherine A Hoadley[2,4], John T Woosley[5], Nancy E Thomas[6], Charles M Perou[2,4,5], Robert S Sandler[1,3] and Norman E Sharpless[1,2,3,4]

RNA expression analysis is an important tool in cancer research, but a limitation has been the requirement for high-quality RNA, generally derived from frozen samples. Such tumor sets are often small and lack clinical annotation, whereas formalin-fixed paraffin-embedded (FFPE) materials are abundant. Although RT-PCR-based methods from FFPE samples are finding clinical application, genome-wide microarray analysis has proven difficult. Here, we report expression profiling on RNA from 157 FFPE tumors. RNA was extracted from 2- to 8-year-old FFPE or frozen tumors of known and unknown histologies. Total RNA was analyzed, reverse-transcribed and used for the synthesis of labeled aRNA after two rounds of amplification. Labeled aRNA was hybridized to a 3′-based 22K spot oligonucleotide arrays, and compared to a labeled reference by two-color microarray analysis. After normalization, gene expression profiles were compared by unsupervised hierarchical clustering. Using this approach, at least 24% of unselected FFPE samples produced RNA of sufficient quality for microarray analysis. From our initial studies, we determined criteria based on spectrophotometric analyses and a novel TaqMan-based assay to predict which samples were of sufficient quality for microarray analysis before hybridization. These criteria were validated on an independent set of tumors with a 100% success rate (20 of 20). Unsupervised analysis of informative gene expression profiles distinguished tumor type and subtype, and identified tumor tissue of origin in three unclassified carcinomas. Although only a minority of FFPE blocks could be analyzed, we show that informative RNA expression analysis can be derived from selected FFPE samples.

Microarray technology has become an important tool in cancer research. Microarrays analyses have identified novel molecular tumor subtypes as well as prognostic gene expression signatures that predict chemotherapy response or tumor progression. Knowledge derived from microarray analysis has been successfully applied to many types of tumor with established clinical utility. For example, RNA expression analysis identified ZAP-70 as a prognostic marker in CLL[1,2] and facilitated development of the Oncotype DX assay used in node-negative, estrogen receptor (ER)+ breast cancer.[3] Likewise, the discovery of molecularly distinct tumor subtypes such as the 'triple negative' or 'basal cluster' breast carcinoma and germinal center diffuse large B-cell lymphoma are the result of genome-wide RNA expression profiling.[4–7] Several groups in industry and academia are applying RNA expression analyses toward a number of clinical questions in an effort to improve molecular pathology and better predict patient outcome.

An important limitation to this approach, however, is the requirement of clinically annotated high-quality RNA. In fact, the majority of large microarray studies to date have used RNA made from frozen samples collected ad hoc at individual centers. These collections are limited in availability, clinical annotation and patient number. The ability to use RNA from formalin-fixed paraffin-embedded (FFPE) samples would solve many of these problems. Given the wide availability of annotated paraffin-embedded tissue blocks,

[1]Department of Medicine, The University of North Carolina School of Medicine, Chapel Hill, NC, USA; [2]Lineberger Comprehensive Cancer Center, The University of North Carolina School of Medicine, Chapel Hill, NC, USA; [3]Center for Gastrointestinal Biology and Disease, The University of North Carolina School of Medicine, Chapel Hill, NC, USA; [4]Department of Genetics, The University of North Carolina School of Medicine, Chapel Hill, NC, USA; [5]Department of Pathology, The University of North Carolina School of Medicine, Chapel Hill, NC, USA and [6]Department of Dermatology, The University of North Carolina School of Medicine, Chapel Hill, NC, USA
Correspondence: Dr NE Sharpless, MD, The Lineberger Cancer Center, Departments of Medicine and Genetics, The University of North Carolina School of Medicine, CB# 7295, Chapel Hill, NC 27599-7295, USA. E-mail: NES@med.unc.edu

both common and rare diseases can be studied retrospectively.

It has been suggested that RNA of sufficient quality for expression analysis could not be routinely derived from FFPE samples. RNA undergoes significant chemical modification by formalin and further degradation during storage.[8–12] Recently, however, there have been successful reports of the use of total RNA isolated from FFPE for RT-PCR assays.[3,12–16] For example, Cronin et al[12] designed a 92-gene assay using RNA extracted from FFPE breast cancer specimens dating from 1985 to 2001, which yielded analyzable data in all tested specimens. In spite of these achievements, however, little evidence to date exists that genome-wide microarray analysis can be applied to FFPE tissue-derived RNA. For example, Karsten et al[10] concluded that formalin-fixed tissue provided a poor substrate for such analyses.

In this work, we show, in accord with earlier work, that microgram quantities of RNA of sufficient quality for limited TaqMan. RT-PCR analysis can be derived from nearly all FFPE samples up to 8 years of age. We further show that RNA from a subset of these samples can be successfully amplified, labeled and hybridized to 22K feature 3′-biased microarrays, and that data from these arrays can determine tumor type and subtype. We find, however, that only a minority of blocks were of sufficient quality for microarray analysis, and that gene signatures derived from FFPE samples contained fewer transcripts (ie less information) than those derived from frozen material. Importantly, we report novel TaqMan-based and spectrophotometric criteria to determine which samples are suitable for microarray analysis before hybridization with ∼100% accuracy. This work demonstrates that meaningful microarray analysis can be performed on FFPE tumors, and provides a realistic appraisal of the feasibility and limitations of this methodology.

MATERIALS AND METHODS
FFPE Tissue
All frozen and FFPE samples were available at the University of North Carolina (UNC) and obtained from either the Tissue Procurement Facility (TPF) or through the North Carolina Colon Cancer Study 1 (NCCS1) Study.[17,18] The blocks age ranged from 2- to 8-years old, with 85% of samples from 1999 to 2001. The following clinical criteria were provided for each FFPE specimen: cancer type, primary tumor or metastases and age of the specimen. All human studies were approved by the UNC Institutional Review Board.

RNA Extraction and Amplification
Sections were prepared as for RNA in situ hybridization, using an RNAse-free microtome on RNAse-free slides. RNA was prepared from FFPE sections using a column-based purification protocol and eluted in a final volume of 30 mcl (Arcturus Paradise System, Arcturus, Mountain View, CA). For each block, the topmost five sections were not used, as the yield from these superficial sections was inconsistent. One of the top sections was H&E stained to allow a determination of tumor and stroma content. Adjoining deeper sections (two per tumor) were then deparaffinized and macrodissected with a razor blade. Although we did not microdissect the specimens, we were able to harvest areas of tumor enrichment using this approach. Macrodissected slides were scraped into proteinase K digestion buffer, and digested for 16 h at 50°. After extraction and purification, $OD_{260/280}$ ratios and RNA yield were determined. RNA samples that passed our initial pre-hybridization criteria (see Results and Discussion) were then amplified twice with polyA priming and T7-based linear amplification using the Paradise Reagent System in accordance with the manufacturer's instructions. For these analyses, starting RNA quantity was 50 ng (sample and reference) and labeling was performed in the second round of amplification.

Microarray Analysis
Samples were analyzed by a comparative hybridization using a common 'reference' mRNA pool as a standard, as described by Perou et al.[5,19–21] After the first round of amplification, samples were labeled with Cy5-dUTP and the pooled cell line control was labeled with Cy3-dUTP by standard methods using the Agilent low-RNA input Linear Amplification RNA kit (#5188-5339). The Cy3- and Cy5-labeled samples were quantitated and combined and then hybridized overnight at 65°C to an Agilent 22K 3′-biased custom array. Samples were quantitated as follows: Cy-dye incorporation was determined in pmol/ng using the Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA). The ratio of Cy-dye reference/Cy-dye sample $= R$ was determined. If the sample aRNA met hybridization criteria (see Results) 2 μg of reference RNA was added to each array and 2 μg × R of the FFPE extracted RNA was added to each array. Normalization to dye incorporation rather than aRNA quantity proved superior for this application (see Results). Custom Agilent 22K 3′-biased Gene Chip containing probe sets representing approximately 22 000 transcripts were used for hybridization. Array washing was performed in accordance with the manufacturer's protocol. Fluorescent images of hybridized microarrays were obtained by using a GenePix 4000 scanner (Axon Instruments, Foster City, CA, USA). Images were gridded and quantified using GenePix Pro 5.1 software. Scanned, gridded images were then uploaded to the UNC microarray database (http://genome.unc.edu). All primary data from this work are available at the same web site.

For unsupervised analysis, genes were filtered using the following criteria: good quality spot (unflagged and normalized spot intensity > 30), spot intensity more than twice background on at least 90% of the arrays and a twofold or greater increase in expression over median on at least three arrays. Using these criteria, 1334 genes passed filtering and were analyzed by hierarchal clustering using Cluster and Java Treeview (M Eisen; http://www.microarrays.org/software).[22]

Figure 1 Rates of hybridization 'success' (see Materials and methods) of 74 independent samples according to RNA quality analysis strategy. Column A shows the success rates of samples arrayed before the use TaqMan and/or Nanodrop analysis to determine RNA degradation and labeling. Column B shows the improved success rate if hybridization was restricted only to moderately degraded samples as determined by Taqman analysis. Column C shows a further improved success rate of hybridizations of samples that met both TaqMan-based and Nanodrop spectrophotometric criteria. The criteria used in column C on FFPE-derived RNA were: extracted RNA $> 20 \, \text{ng/mcl}$, extracted RNA $\text{OD}_{260/280} > 1.5$, extracted RNA 3′–5′ ratio < 100 $(\Delta C_T(5'3') < 6.5)$ and Cy-dye incorporation in aRNA $> 4.5 \, \text{pmol/ng}$ (see Results).
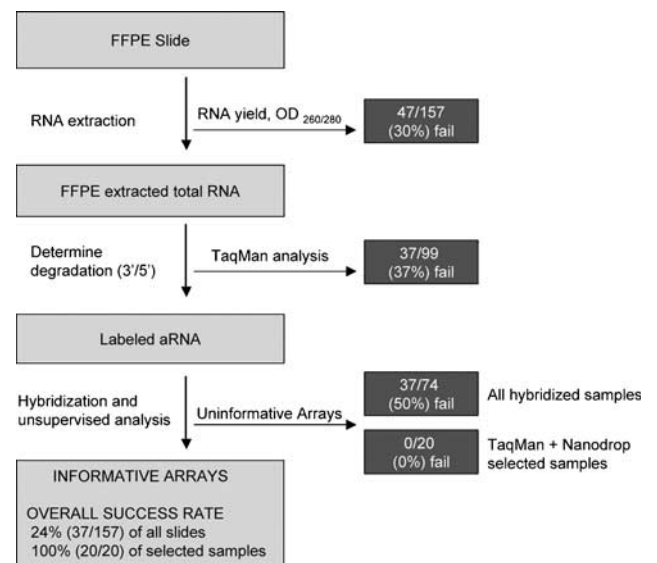
To determine which arrays were 'informative' or 'successful', we employed a stringent definition based on unsupervised hierarchical clustering compared with known high-quality samples (eg from frozen RNA) as well as known poor-quality samples of degraded RNA. An 'informative' hybridization met the following criteria:

1. Little or no green bias (red gain–green gain < 300).
2. $> 70\%$ of spots were of good quality (unflagged).
3. By unsupervised analysis, informative arrays clustered with like hybridizations of tumor types using high-quality (eg frozen) samples and not with hybridizations derived from samples of known degraded RNA.

We discovered *post hoc* that it was possible to move a small number (four or fewer) of hybridizations from the degraded clusters to the informative cluster by altering filtering criteria; but for the purposes of Figures 1–3, these arrays were considered uninformative.

RT–PCR

Total RNA was extracted as described. Transcription into cDNA was performed in a 20-μl volume using oligo-dT or random hexamer and ImProm-II reverse transcriptase (Promega Corp). All PCRs were carried out in a final volume of 20 μl and were performed in duplicate for each cDNA sample in the ABI PRISM 7700 Sequence Detection System (Applied Biosystems) according to the manufacturer's protocol. Sequence-specific TaqMan primers and probe were designed using Primer Express (Applied Biosystems) for β-actin (Supplementary Figure 1). Two primer-probes sets were developed (5′ and 3′) ~ 300 base pairs (bp) apart. The reaction mix consisted of Universal Master Mix No AmpErase UNG (Applied Biosystems), 0.25 μM fluorogenic probe, 0.9 μM of each specific forward and reverse primer and 9 μl of diluted
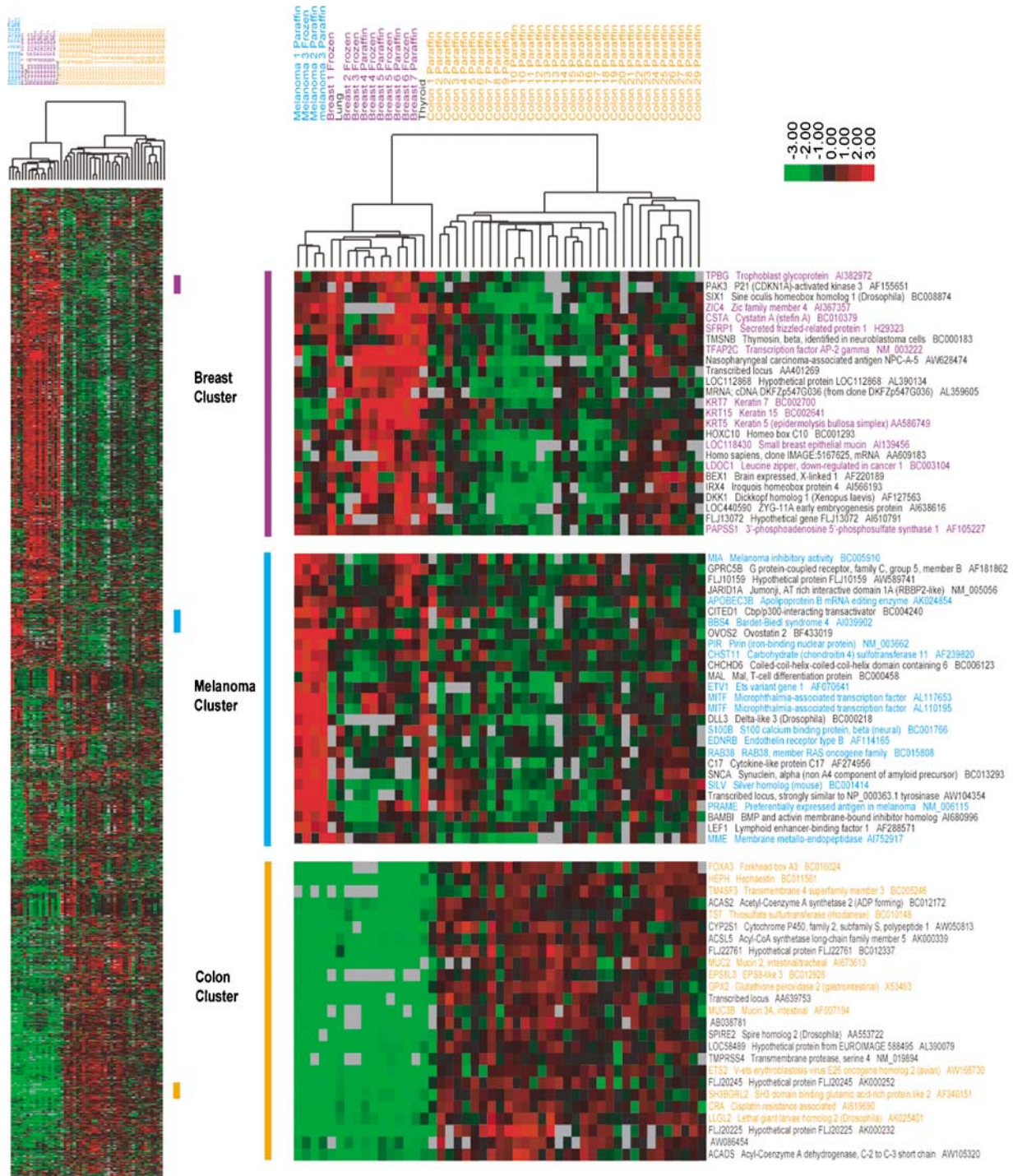

Figure 2 A flow diagram showing pass and fail rates at each step of RNA quality analysis using criteria described in the results. A hybridization was considered 'successful' or 'informative' based on array quality and behavior in unsupervised hierarchical clustering (see Materials and methods).

cDNA. Amplifications were performed under standard conditions. The number of PCR cycles needed to reach the fluorescence threshold (C_T) was determined in duplicate for each cDNA and averaged.[23] We determined empirically that these 5′ and 3′ actin primer pairs amplified with comparable efficiency, and therefore the $\Delta C_T(5'3')$ was defined as the mean $5'C_T$ subtracted from the mean $3'C_T$ value of each sample. Using this methodology, non-degraded, high-quality RNA (eg from a cell line) showed a $\Delta C_T(5'3') = 0$. The 3′/5′ ratio was determined as $1/(2^{\wedge}\Delta C_T(5'3'))$.

RESULTS

We extracted total RNA from 157 FFPE tumor blocks with seven matched frozen specimens. Total RNA was harvested from 5-μm sections of tumor containing paraffin blocks from the UNC TPF and NCCS1. RNA yields from FFPE tissue were unpredictable in that total RNA quantity did not strictly correlate with the size or amount of sample or the age of the block. Sufficient RNA for TaqMan analysis of an extreme 3′ segment of an abundant transcript (β-actin) after cDNA synthesis using oligo-dT primer was possible for all harvested blocks (not shown). This finding suggests, in accord with published results,[3,6,12,13,15] that TaqMan-based strategies using gene-specific primers to detect abundant transcripts is generally achievable using FFPE-derived samples.

Successful RNA expression profiling requiring unbiased amplification and hybridization, however, depends on RNA quality in addition to RNA yield. Although the yield and $\text{OD}_{260/280}$ of the extracted RNA (measuring protein contamination) were generally acceptable, these measures did not predict RNA degradation. Several techniques are available

Figure 3 Unsupervised hierarchical clustering of transcripts that passed pre-defined criteria from informative expression profiles of FFPE and frozen tumors. Data are median-centered, green represents expression below median, red represents above median and gray represents missing data. Three representative clusters are indicated by colored bars and explicitly shown on the right. Genes names in these clusters are colored if they typify the indicated tumor type as demonstrated by greater than median expression in the majority of tumors of the indicated histology in one of two publicly available data sets[20,26] (orange, colon cancer; blue, melanoma; purple, breast).

for determining RNA integrity such as denaturing agarose gel analysis (requiring several micrograms of RNA), ABI RNA bioanalyzer, or RT–PCR (to determine a 3′/5′ ratio). We attempted to assess RNA quality by bioanalyzer, but FFPE samples in general proved too degraded to be interrogated

reliably using this method (Supplementary Figure 1). Therefore, we pursued a quantitative RT-PCR approach.

We designed two sets of β-actin primers ~ 300 bp apart in the 3′ portion of the transcript for RT-PCR with SYBR labeling. We chose to design our primers 300 bp apart as all

oligo probes are within 300 bp of the polyA tail on the Agilent 3′-biased microarray. As poly-dT was used for reverse transcription, the 3′/5′ ratio determined using this strategy should always be >1 as reverse transcriptase is not 100% processive. Despite careful primer design and multiple attempts at optimization, however, determining RNA integrity utilizing SYBR dye produced inconsistent results. For example, estimates of 3′/5 ratios using SYBR for pristine RNA specimens ranged between 0.2 and 4. Additionally, 10 of 12 FFPE specimens in a pilot sample demonstrated 3′/5′ ratios less than 40 when estimated by SYBR RT-PCR, yet would not amplify and label sufficiently for informative hybridizations. Therefore we concluded that both bioanalyzer- and SYBR-based RT-PCR were inadequate to predict successful hybridization using FFPE-derived RNA.

As SYBR detects any dsDNA product, we suspected that determination of the 5′ (less abundant) transcript was systematically biased by spurious amplification products, and therefore designed a TaqMan strategy to determine 3′/5′ ratios (Supplementary Figure 2). Because they anneal to an internal region of the desired PCR product, TaqMan probes provide enhanced specificity over SYBR. Using two sets of β-actin primers located in the 3′ portion of the transcript (at ∼1500 and 1800 bp from the translation start site) we are able to quantitate reliably the transcript 3′–5′ ratio (ie the relative copy number of the 1800 bp message to the 1500 bp message). Both primer pairs amplified a single PCR product by melting point analysis and gel electrophoresis (data not shown). Using the TaqMan strategy on pristine RNA, 3′/5′ ratios of 0.9–1.2 were seen, a much smaller range than that determined with SYBR. Additionally, we found that SYBR systematically underestimated the degree of RNA degradation. For comparison, the average $\Delta C_T(5′3′)$ obtained with SYBR was approximately 2.5 (22 samples, geometric mean 3′/5′ = 5.7), whereas with the TaqMan primer sets, 5.8 (99 samples, geometric 3′/5′ = 56). We noted that samples with a $\Delta C_T(5′3′) < 6.5$ by TaqMan were more likely to provide 'informative' microarray analysis (as defined in the Materials and methods, Figure 1), and therefore, this cutoff was chosen for subsequent hybridizations to identify severely degraded, unusable RNA from less degraded, usable samples.

Utilizing this Taqman assay, the $OD_{260/280}$ values and extracted RNA quantity, we improved the ability to predict which RNA samples would give useful gene signatures when hybridized. For FFPE-derived RNA with $OD_{260/280}$ ratios >1.5, 3′–5′ ratios <100 ($\Delta C_T(5′3′) < 6.5$), and yields of >20 ng/mcl (600 ng total), we noted successful hybridization from 48% of samples, as compared with 17% pre-TaqMan (Figure 1, column A *vs* B). However, it was still not possible to predict in most cases which samples would successfully label and provide informative hybridizations. In addition, many samples labeled with low efficiency and when hybridized, generated arrays with low sample signal. To compensate for the decreased labeling of these samples, the hybridized arrays had to be scanned with increased gain on

the red (sample) channel compared with the green (reference) channel introducing a reproducible artifact. Therefore, we sought to control for the efficiency of labeling as well.

To accomplish this, we employed a multi-wavelength spectrophotometer (Nanodrop ND-1000) capable of analyzing small volumes (1 mcl) of analyte. The ability to analyze small volumes with reliability permitted the determination of progress at every step in aRNA synthesis, eg amplification efficiency and Cy-dye incorporation after labeling were measured. Using this approach, high-quality reference RNA always labeled successfully with average Cy-dye incorporation of 40 pmol/ng of aRNA (see Materials and methods). In contrast, FFPE RNA, even from samples with relatively low 3′/5′ ratios, labeled less efficiently; generally, <15 pmol/ng of aRNA. This appreciation of the decreased efficiency of labeling of FFPE samples allowed for two methodologic improvements. First, we noted that samples which labeled very inefficiently (<4.5 pmol/ng of aRNA) did not yield informative hybridizations; and therefore this labeling criteria was included in the algorithm to predict which samples would produce informative hybridizations. Second, we improved hybridization quality by normalizing the reference and sample aRNA based on Cy-dye incorporation (in pmol per ng of aRNA) instead of total aRNA quantity (the normalization procedure is described in Materials and methods).

On the basis of experience from these initial analyses, empirically determined criteria of RNA quality and quantity were devised to predict which samples would hybridize successfully pre-hybridization:

1. A yield of >600 ng (20 ng per mcl) of extracted total RNA.
2. $OD_{260/280}$ ratio >1.5 of extracted total RNA.
3. A 3′–5′ ratio <100 (or $\Delta C_T(5′3′) < 6.5$) of extracted total RNA.
4. Cy-dye incorporation >4.5 pmol/ng in labeled aRNA.

A flow diagram (Figure 2) shows the failure rates at each step of this algorithm. These criteria, coupled with the practice of normalizing sample *vs* reference aRNA to Cy-dye incorporation rather than aRNA yield, increased performance on an independent set of tumor samples with a success rate approaching 100% (20 of 20; Figure 1, column C). Therefore, identifying which FFPE-RNA samples are of sufficient quality to merit array hybridizations is possible after labeling before hybridization.

Using a rigorous definition of hybridization success based on unsupervised analysis (see Materials and methods), all hybridizations of RNA derived from frozen samples were informative. In contrast, only 50% (37 of 74) of the FFPE-derived samples were informative, although this success rate could be significantly enhanced by pre-hybridization sample selection using the aforementioned criteria. When compared with uninformative arrays, all informative arrays clustered on a common dendrogram branch, whereas uninformative

arrays clustered with samples of known poor RNA quality (data not shown). Although not addressed in this work, we believe useful gene expression information could be further obtained from some, but not all, of the non-informative arrays through statistical approaches and other technical improvements (see Discussion).

Nonetheless, the expression data obtained from informative arrays were of good quality and compared with favorably other analyses of frozen samples that analyzed tumors or cell lines of multiple histologic subtypes.[24–27] For example, by unsupervised analysis of the 45 (FFPE + frozen) informative arrays of distinct subtypes, 1334 genes passed filtering using pre-defined criteria, and hierarchical clustering of these samples demonstrated that the tumors clustered by histologic type (Figure 3). Melanoma, breast and colon cancers clustered on distinct dendrograms, with only a single colon cancer (colon sample #1; Figure 3) clustering on a distinct branch from similar tumors. Histologic review of this tumor was consistent with colon cancer, and it demonstrated overexpression of transcripts overexpressed in other colon cancers (in orange, Figure 3). It is possible that this tumor represents a distinct but rare colon cancer subtype, or that technical features of this hybridization led to misclassification. Two additional tumors, a thyroid cancer and a lung adenocarcinoma, that were of unclear tissue origin before this analysis (see discussion) clustered loosely with the breast tumors. The heterogeneity of the identified breast cluster is not surprising given that the number of samples was relatively small, and included breast tumors of three well-recognized subtypes (Her2 + , ER + and basal cluster).[6] These results show that informative hybridizations of FFPE-derived aRNA produced expression data of sufficient quality to allow the identification of tumor subtypes.

Additional evidence suggested that these microarray data of FFPE samples were comparable with results obtained from frozen material. For example, frozen-FFPE-matched pairs clustered samples with high intra-class correlation coefficients (Pearson $r > 0.7$ for all pairs).[10,16] Additionally, many of the identified transcripts that typified the specific clusters were familiar markers of that tumor type or have been found to characterize these tumor subtypes in other studies. For example, transcripts in Figure 3 were colored if they were overexpressed in the indicated tumor type in two publicly available data sets[25,26]; melanoma in blue, colon in orange and breast in purple. The well-recognized clinical markers of melanoma, silver homolog (PMEL17) and microopthalmia characterized the melanoma samples. GATA3, Keratin 5, Her2 and ER receptors all passed filtering and were overexpressed in certain breast tumors corresponding with their known subtype (Her2 *vs* basal *vs* ER + ; not shown). A large number of transcripts (> 600 of 1334 used in this unsupervised analysis) characterized colon cancer including several markers (eg Mucins 2 and 3b, Hephaestin, Ets2 and FOXA3) associated with colon histology in other series.[25,26] In aggregate, these results demonstrate that meaningful
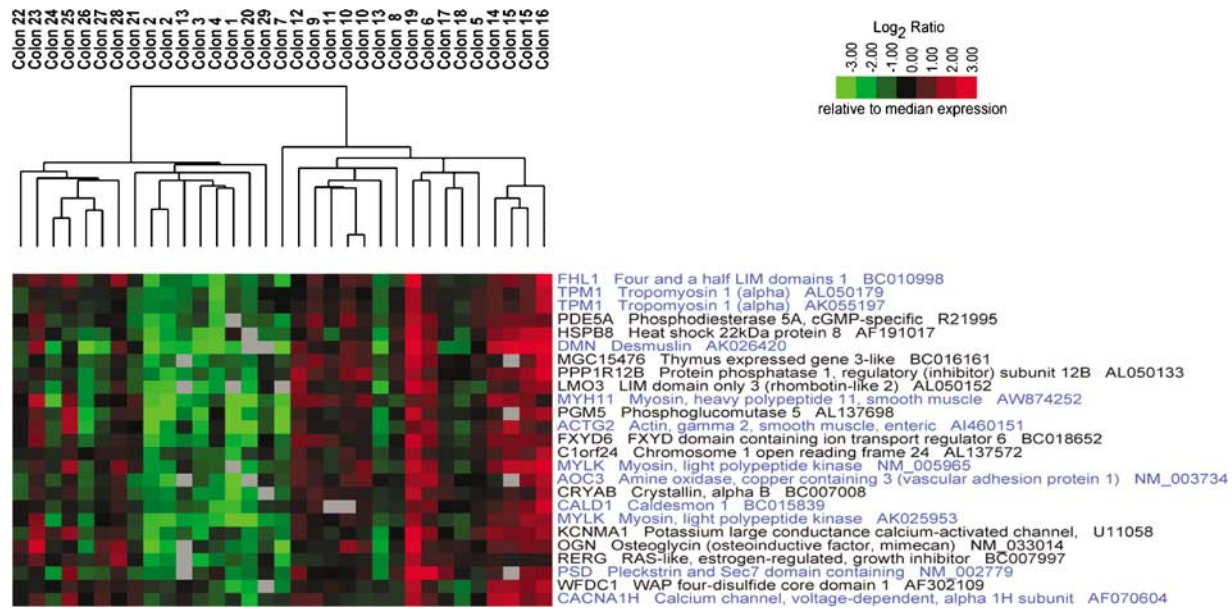
expression analysis can be performed on selected 2–8-year FFPE tumors.

Analysis of FFPE-derived RNA also identified meaningful heterogeneity among tumors of a given histologic subtype. As stated, several markers of specific breast tumor subtype (eg GATA3, ER, Krt5)[5,6,28] passed filtering and demonstrated increased expression in FFPE samples of those breast subtypes (not shown). Moreover, when only the colon samples were considered by unsupervised analysis, the tumors segregated into two clusters of roughly equal size (a representative sub-cluster is shown in Figure 4). This clustering of colon samples into two distinct groups may reflect subtypes of colon cancer (eg MMR + *vs* MMR−) as reported by others,[29] but we believe in part represents an unequal degree of smooth muscle and other stromal contamination of these samples. This conclusion is supported by the finding that the transcripts that best distinguished the two subgroups (Figure 4) are highly expressed in smooth muscle: eg γ-actin, smooth muscle myosin, desmuslin and tropomyosin (smooth muscle transcripts identified using source data of[30]). Histological analysis of these tumors suggested increased stromal contamination corresponded with high expression of smooth muscle-associated transcripts (not shown). These data suggest that analysis of selected FFPE samples can identify tumor-relevant features beyond histologic subtype.

Although these results from unsupervised analysis were encouraging, we noted an additional, perhaps unanticipated limitation to the analysis of FFPE-derived samples. A comparison of independent analyses of the matched frozen and FFPE-derived samples revealed that there was a significant loss of gene signature information using FFPE-derived *vs* frozen material. One measure of this loss of information is suggested by the observation that the number of transcripts that passed standard filtering criteria based on spot quality and range of variation was 40% lower for the FFPE data set *vs* matched frozen samples. This finding could reflect either an inability to detect less abundant transcripts after formalin fixation, or significant differences in stability across transcripts during formalin fixation. Both possibilities have been suggested by previous analyses of FFPE-derived RNA.[10,15] This comparison indicates that a significant quantity of expression information, particularly related to rare or unstable transcripts, is lost in analyses of FFPE-derived RNA.

DISCUSSION
We found that only a quarter of unselected FFPE samples aged 2–8 years provided RNA of sufficient quality for successful expression analysis. Perhaps this is not surprising given the well-described detrimental effects of formalin fixation on RNA, as well as marked heterogeneity in tumor fixation techniques and block storage conditions between institutions. Additionally, even using only successful hybridizations, we noted a loss of information in gene signatures of FFPE-derived samples compared with matched frozen samples. Although precise quantification of this loss of information is

Figure 4 A representative subcluster of a larger unsupervised analysis of FFPE-derived colon tumors. Two distinct clusters of colon carcinoma are identified which are largely, but not entirely, distinguished by the expression of smooth muscle-associated transcripts (identified in blue, see[30]). Histological analysis suggested that high expression of smooth muscle transcripts correlated with stromal contamination of these specimens.

not possible given the chosen experimental approach, a limited analysis of matched FFPE and frozen specimens suggests that 40% of transcripts that pass standard filtering using frozen samples will not pass identical criteria using FFPE samples. Despite these significant drawbacks, however, this work shows that highly informative arrays can be generated from FFPE using a very rigorous definition of success.

We believe the low success rate seen in this study can be improved. First, the definition of hybridization 'success', based on unsupervised analysis, is likely overly stringent. With the application of statistical techniques to control for RNA degradation and block age, it may be possible to glean useful information from hybridizations that we considered uninformative. For example, Chung *et al*[31] have recently reported a statistical approach to account for block age of FFPE samples. Also, supervised analysis[32] with respect to a variable of clinical interest (eg patient outcome) may minimize the effect of certain systematic biases across the data set. Secondly, although not tested directly, it is probable that a subset of the hybridizations performed in the absence of normalization to aRNA labeling—that is based on Cy-dye incorporation determined by Nanodrop analysis—would have provided informative microarrays if the amount of sample aRNA had been increased to account for inefficient labeling. Lastly, new technologies are emerging that may overcome limitations of the current methodology. For example, random hexamer priming and terminal transferase end-labeling[33,34] may enhance cDNA synthesis, a troublesome step in the current approach. It is reasonable to believe these approaches may improve cDNA synthesis over methods relying on traditional oligo-dT primer as polyA tracts seem parti-

cularly prone to formalin-mediated covalent modification.[9] Along these lines, Bibkova *et al*[35,36] have combined random hexamer cDNA synthesis with sensitive, multiplexed assays of gene expression using fiberoptic beads to interrogate simultaneously 200–500 transcripts using RNA from FFPE material. For these reasons, the success rate of 24% reported in this series is likely conservative, and we feel will be improved in future efforts.

Nonetheless, we believe some FFPE blocks yield RNA that is of insufficient quality for informative microarray data by almost any approach. For example, 30% (47 of 157) of blocks yielded minute amounts of total RNA and/or were irrevocably protein contaminated (Figure 2). These problems do not reflect improper extraction technique, but rather indicate significant RNA degradation in these samples, as we attempted to re-harvest the majority of these 47 samples, without subsequent success in a single case. Moreover, 37% (37 of 99) of the samples, which passed the initial criteria of RNA yield and purity ($OD_{260/280}$), were still highly degraded (3′/5′ ratio > 100, Figure 2). In fact, in a few samples, no 5′-PCR product could be detected using a highly sensitive and optimally designed TaqMan primer-probe set, a mere 300 bp from the polyA tail of β-actin, a highly abundant transcript. Therefore, although TaqMan-based analyses using gene-specific primers for cDNA synthesis may be possible on RNA from such samples, we are skeptical that rigorously defined 100% success rates of microarray hybridization will be routinely achievable using unselected FFPE samples.

Two advantages of this study are that we tested a relatively large number of FFPE samples (157) and that blocks in this work came from both community and tertiary-care hospitals

from across the state of North Carolina. Therefore, we expect our results would be generalizable to other FFPE collections from disparate sources. It may be, however, that by choosing blocks that had been ideally handled and stored, we would have determined a higher success rate for microarray analysis of FFPE samples. Our data give pause in this regard, however, as the success rate was comparable for samples from the NCCS study (where blocks largely came from community hospitals) as from the UNC TPF (where blocks were all prepared and stored at an academic medical center). The variables that determine which blocks provide higher quality RNA are not clear from our analysis, but our data would be consistent with other work suggesting that the manner and details of formalin fixation are the crucial variables determining RNA quality from FFPE samples.[8,11,37,38]

Given that degradation in FFPE samples is unpredictable and does not solely correlate with block age, we believe the significant contribution of this work is the ability to discern which RNA samples will provide useful microarray signatures before hybridization. The empirically determined criteria in this work account for both quantitative and qualitative problems with FFPE RNA as they address RNA degradation (by TaqMan) and inability to incorporate Cy-dye label (by Nanodrop). The latter in turn reflects RNA quality (eg chemical modifications of polyA tracts, protein contamination, etc.). These pre-hybridization criteria greatly enhance the feasibility of this methodology, because the RNA harvesting, TaqMan and spectrophotometric analyses are relatively inexpensive compared with the cost of oligonucleotide arrays.

Our results demonstrate an obvious application of this technology: the use of microarray expression profiling on FFPE samples to identify tissue of origin in carcinoma of unknown primary (CUP). Several recent publications have shown that microarray technology can predict the tissue of origin in CUP,[5,39] which represents approximately 3% of all new cancer diagnoses.[40] In practice, however, patients with CUP often only have FFPE samples, and repeat biopsy is impractical in many instances. Therefore, the ability to perform microarray analysis on FFPE samples appears to be an advance in CUP diagnosis. Three examples from this work serve as proof-of-principle of this application. In one example, a tumor had been mis-annotated as colon cancer. This sample did not cluster with the other colon cancers in this study (Figure 3), and subsequent pathological review correctly identified it as a thyroid malignancy. Additionally, a tumor presenting with intraperitoneal carcinomatosis was initially diagnosed as colon cancer, but was clearly established as pancreatic using this approach (not shown). Finally, a widely metastatic tumor that highly expressed the ovarian marker CA-125 was initially diagnosed as CUP, likely ovarian. Microarray analysis followed by comparison to public microarray data sets clearly indicated this tumor was pulmonary in origin, sequencing demonstrated an exon 19 deletion in EGFR and therapy with the kinase inhibitor, erlotinib, produced a durable clinical response. These anecdotal experi- ences suggest that microarray analysis on FFPE samples could be a valuable adjunct to clinical pathology in the diagnosis of CUP. It is, however, unclear from our work if expression profiling of selected FFPE samples will be of value in other potential applications of the technology; for example, to identify transcripts that predict outcome in large FFPE data sets from completed inter-group trials.

In summary, these data suggest that meaningful RNA expression analysis can be performed on FFPE samples, with the caveats that many samples are too degraded for analysis and that there is loss of information using FFPE-derived compared to analysis of frozen samples. Nonetheless, we have identified criteria to predict which blocks will provide informative hybridizations, and have demonstrated a near ready-for-the-clinic application of these methodologies: diagnosis of CUP. We believe further technical refinements will continue to enhance the utility of genome-wide RNA-based assays on FFPE samples.

Supplementary Information accompanies the paper on the Laboratory Investigation website (http://www.laboratoryinvestigation.org)

1. Rassenti LZ, Huynh L, Toy TL, *et al*. ZAP-70 compared with immunoglobulin heavy-chain gene mutation status as a predictor of disease progression in chronic lymphocytic leukemia. N Engl J Med 2004;351:893–901.
2. Wiestner A, Rosenwald A, Barry TS, *et al*. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood 2003;101:4944–4951.
3. Paik S, Shak S, Tang G, *et al*. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med 2004;351:2817–2826.
4. Alizadeh AA, Eisen MB, Davis RE, *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–511.
5. Perou CM, Sorlie T, Eisen MB, *et al*. Molecular portraits of human breast tumours. Nature 2000;406:747–752.
6. Sorlie T, Perou CM, Tibshirani R, *et al*. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 2001;98:10869–10874.
7. Shipp MA, Ross KN, Tamayo P, *et al*. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 2002;8:68–74.
8. Williams C, Ponten F, Moberg C, *et al*. A high frequency of sequence alterations is due to formalin fixation of archival specimens. Am J Pathol 1999;155:1467–1471.
9. Masuda N, Ohnishi T, Kawamoto S, *et al*. 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.
10. Karsten SL, Van Deerlin VM, Sabatti C, *et al*. An evaluation of tyramide signal amplification and archived fixed and frozen tissue in microarray gene expression analysis. Nucleic Acids Res 2002;30:E4.

11. Benchekroun M, DeGraw J, Gao J, *et al*. Impact of fixative on recovery of mRNA from paraffin-embedded tissue. Diagn Mol Pathol 2004;13:116–125.

12. Cronin M, Pho M, Dutta D, *et al*. 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.

13. Abrahamsen HN, Steiniche T, Nexo E, *et al*. Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. J Mol Diagn 2003;5:34–41.

14. Finke J, Fritzen R, Ternes P, *et al*. An improved strategy and a useful housekeeping gene for RNA analysis from formalin-fixed, paraffin-embedded tissues by PCR. Biotechniques 1993;14:448–453.

15. Godfrey TE, Kim SH, Chavira M, *et al*. 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.

16. Specht K, Richter T, Muller U, *et al*. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. Am J Pathol 2001;158:419–429.

17. Satia JA, Keku T, Galanko JA, *et al*. Diet, lifestyle, and genomic instability in the North Carolina Colon Cancer Study. Cancer Epidemiol Biomarkers Prev 2005;14:429–436.

18. Kinney AY, Harrell J, Slattery M, *et al*. Rural–urban differences in colon cancer risk in blacks and whites: the North Carolina Colon Cancer Study. J Rural Health 2006;22:124–130.

19. Eisen MB, Brown PO. DNA arrays for analysis of gene expression. Methods Enzymol 1999;303:179–205.

20. Ross DT, Scherf U, Eisen MB, *et al*. Systematic variation in gene expression patterns in human cancer cell lines. Nat Genet 2000;24:227–235.

21. Novoradovskaya N, Whitfield ML, Basehore LS, *et al*. Universal Reference RNA as a standard for microarray experiments. BMC Genomics 2004;5:20.

22. Eisen MB, Spellman PT, Brown PO, *et al*. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–14868.

23. Krishnamurthy J, Torrice C, Ramsey MR, *et al*. Ink4a/Arf expression is a biomarker of aging. J Clin Invest 2004;114:1299–1307.

24. Ramaswamy S, Tamayo P, Rifkin R, *et al*. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA 2001;98:15149–15154.

25. Scherf U, Ross DT, Waltham M, *et al*. A gene expression database for the molecular pharmacology of cancer. Nat Genet 2000;24:236–244.

26. Su AI, Welsh JB, Sapinoso LM, *et al*. Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 2001;61:7388–7393.

27. Golub TR, Slonim DK, Tamayo P, *et al*. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–537.

28. Perou CM, Jeffrey SS, van de Rijn M, *et al*. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA 1999;96:9212–9217.

29. di Pietro M, Bellver JS, Menigatti M, *et al*. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–1059.

30. Su AI, Cooke MP, Ching KA, *et al*. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci USA 2002;99:4465–4470.

31. Chung CH, Parker JS, Ely K, *et al*. Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor-{kappa}B signaling as characteristics of a high-risk head and neck squamous cell carcinoma. Cancer Res 2006;66:8210–8218.

32. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA 2001;98:5116–5121.

33. Park JW, Beaty TH, Boyce P, *et al*. Comparing whole-genome amplification methods and sources of biological samples for single-nucleotide polymorphism genotyping. Clin Chem 2005;51:1520–1523.

34. Barker DL, Hansen MS, Faruqi AF, *et al*. Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. Genome Res 2004;14:901–907.

35. Bibikova M, Talantov D, Chudin E, *et al*. Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. Am J Pathol 2004;165:1799–1807.

36. Bibikova M, Yeakley JM, Chudin E, *et al*. Gene expression profiles in formalin-fixed, paraffin-embedded tissues obtained with a novel assay for microarray analysis. Clin Chem 2004;50:2384–2386.

37. Foss RD, Guha-Thakurta N, Conran RM, *et al*. Effects of fixative and fixation time on the extraction and polymerase chain reaction amplification of RNA from paraffin-embedded tissue. Comparison of two housekeeping gene mRNA controls. Diagn Mol Pathol 1994;3:148–155.

38. Guerrero RB, Batts KP, Brandhagen DJ, *et al*. Effects of formalin fixation and prolonged block storage on detection of hepatitis C virus RNA in liver tissue. Diagn Mol Pathol 1997;6:277–281.

39. Tothill RW, Kowalczyk A, Rischin D, *et al*. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Res 2005;65:4031–4040.

40. Briasoulis E, Pavlidis N. Cancer of unknown primary origin. Oncologist 1997;2:142–152.