

RESEARCH ARTICLE

Open Access

# Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens

Al Muktafi Sadi<sup>1†</sup>, Dong-Yu Wang<sup>1†</sup>, Bruce J Youngson<sup>2</sup>, Naomi Miller<sup>2</sup>, Scott Boerner<sup>2</sup>, Susan J Done<sup>1,2,3,4</sup> and Wey L Leong<sup>1,5\*</sup>

## Abstract

**Background:** The ability of gene profiling to predict treatment response and prognosis in breast cancers has been demonstrated in many studies using DNA microarray analyses on RNA from fresh frozen tumor specimens. In certain clinical and research situations, performing such analyses on archival formalin fixed paraffin-embedded (FFPE) surgical specimens would be advantageous as large libraries of such specimens with long-term follow-up data are widely available. However, FFPE tissue processing can cause fragmentation and chemical modifications of the RNA. A number of recent technical advances have been reported to overcome these issues. Our current study evaluates whether or not the technology is ready for clinical applications.

**Methods:** A modified RNA extraction method and a recent DNA microarray technique, cDNA-mediated annealing, selection, extension and ligation (DASL, Illumina Inc) were evaluated. The gene profiles generated from FFPE specimens were compared to those obtained from paired fresh fine needle aspiration biopsies (FNAB) of 25 breast cancers of different clinical subtypes (based on ER and Her2/neu status). Selected RNA levels were validated using RT-qPCR, and two public databases were used to demonstrate the prognostic significance of the gene profiles generated from FFPE specimens.

**Results:** Compared to FNAB, RNA isolated from FFPE samples was relatively more degraded, nonetheless, over 80% of the RNA samples were deemed suitable for subsequent DASL assay. Despite a higher noise level, a set of genes from FFPE specimens correlated very well with the gene profiles obtained from FNAB, and could differentiate breast cancer subtypes. Expression levels of these genes were validated using RT-qPCR. Finally, for the first time we correlated gene expression profiles from FFPE samples to survival using two independent microarray databases. Specifically, over-expression of *ANLN* and *KIF2C*, and under-expression of *MAPT* strongly correlated with poor outcomes in breast cancer patients.

**Conclusion:** We demonstrated that FFPE specimens retained important prognostic information that could be identified using a recent gene profiling technology. Our study supports the use of FFPE specimens for the development and refinement of prognostic gene signatures for breast cancer. Clinical applications of such prognostic gene profiles await future large-scale validation studies.

\* Correspondence: [wey.leong@uhn.on.ca](mailto:wey.leong@uhn.on.ca)

† Contributed equally

<sup>1</sup>Ontario Cancer Institute and Campbell Family Institute for Breast Cancer Research, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada  
Full list of author information is available at the end of the article

## Background

Gene profiling is beginning to have an impact on personalized breast cancer care [1,2]. Gene expression profiling of breast cancers using DNA microarray technology is able to classify breast tumors into distinct biological subgroups and has been shown to predict treatment response and prognosis in several studies [3-9]. This high-throughput molecular technique requires fresh bio-specimens to allow extraction of RNA of sufficient quantity and quality for analysis. There are however limitations to the collection of fresh samples for prospective studies including time-sensitive tissue processing, lengthy patient accrual and follow-up; and bio-banks are not always readily available as a source of fresh frozen samples. Also, a requirement for fresh tissue to be used inevitably leads to a bias towards only larger tumors being studied. To overcome some of these critical shortcomings of prospective studies, the use of archival formalin-fixed paraffin-embedded (FFPE) samples offers a potential solution as most hospitals worldwide have collections of FFPE tumor specimens dating back many years. FFPE is the most widely used standard of practice for tissue fixation for the purpose of diagnostic histology and long-term storage.

The FFPE tissue preserving process was developed long before molecular biologists were concerned with the preservation of RNA. FFPE samples have not been considered a reliable source of RNA due to the tissue processing-associated degradation and chemical modifications of RNA. Formalin fixation creates cross-linking between nucleic acids and proteins and adds monomethylol to amino groups on all four RNA bases [10]. Thus, a number of recent studies have started to look into the prospect of overcoming the RNA quality issues in FFPE specimens. Several studies used modification of standard RNA extraction methods to generate RNA of sufficient quality and quantity for DNA microarray analysis [11-14]. Some innovations in DNA microarray techniques were also reported [15,16]. A major breakthrough has been a new microarray technique developed by Illumina Inc. (San Diego, Ca), which involves cDNA-mediated annealing, selection, extension, and ligation (DASL), as well as random priming for detection of degraded RNA from FFPE samples [16-22]. This DNA microarray technique improved detection of fragmented RNA compared with conventional techniques. If the technique for gene profiling using FFPE specimens becomes sufficiently reliable, we anticipate that prognostic and predictive gene signatures can be identified using the vast available libraries of archival FFPE specimens with long-term treatment outcomes. In addition, any biomarker developed from FFPE samples could be more readily translated into clinical practice.

In this study, we aimed 1) to explore the feasibility of obtaining reliable microarray data from archival FFPE samples; 2) to compare gene expression profiles of FFPE samples with those of matched samples obtained from the same patients by FNAB; 3) to test the reproducibility of such experiments using quantitative real-time reverse transcription-polymerase chain reaction (RT-qPCR); and 4) to correlate the gene profiling of FFPE specimens with clinical outcome data using published microarray data sets.

## Methods

### Case selection and sample size

From the clinical tissue archive stored on site in the hospitals at University Health Network, a total of 50 FFPE blocks were evaluated by our breast pathologists (BY, NM, SD). Thirty nine blocks were finally selected for this study based on the following criteria: 1) sufficient invasive carcinoma remained in the FFPE block for RNA extraction and 2) > 70% malignant cells. In addition, 25 cases with FFPE blocks that also had available pre-operative fresh FNAB specimens that had been collected from an ongoing breast cancer gene profiling study in our institution on an unselected cohort of surgical patients. Clinically, the estrogen receptor (ER, also known as *ESR1*) and Her2/neu (Her2, also known as *ERBB2*) levels were evaluated by immunohistochemistry or by fluorescence in situ hybridization, according to the standard clinical protocols. Based on ER and Her2 status, the 25 cases were divided into four distinct subgroups: 10 cases in ER+/Her2-, 4 in ER+/Her2+, 6 in ER-/Her2+, and 5 in ER-/Her2- status for the next step of the analyses. The clinical descriptors for these 25 cases are provided in Table 1. The study protocol was approved by the Research Ethics Board at the University Health Network.

### Tissue sampling and RNA extraction

FNAB samples and RNA extractions were prepared as previously described [23]. For FFPE samples, the selected blocks were sectioned at 10  $\mu$ m thickness in an RNase-free environment. Total RNA was isolated from FFPE samples using a modified protocol described by Abramovitz et al [24] with RecoverAll Total Nucleic Acid kit (Ambion Austin, TX). In brief, the sections were deparaffinized with xylene and air dried tissue pellets were homogenized by overnight incubation at 50°C with Proteinase K in a lysis buffer. The Proteinase K was then inactivated by incubating the sample at 80°C for 15 minutes the next day. RNA was purified and extracted after a DNase I treatment using a spin column. Quantity of the RNA was measured by ND-1000 Spectrophotometer (Nanodrop Technologies, Wilmington, DE USA). To

**Table 1 The 25 patient and tumor characteristics.**

Patient No.	Group No.	Patient Age	Tumor <sup>1</sup> Size	Tumor Grade	Positive <sup>2</sup> Nodes	ER	PR	Her2	FNAB RIN	FFPE RIN
1	E(+)H(-)-1	56	1.2	2	0(1)	+	+	-	7.5	1.7
2	E(+)H(-)-2	47	2.2	2	1(12)	+	+	-	9.2	2.3
3	E(+)H(-)-3	63	2.4	2	0(1)	+	+	-	7.0	2.5
4	E(+)H(-)-4	76	2.8	3	0(3)	+	+	-	8.9	2.5
5	E(+)H(-)-5	58	10.6	2	6(24)	+	+	-	6.6	2.5
6	E(+)H(-)-6	50	3.7	2	0(2)	+	+	-	8.6	2.4
7	E(+)H(-)-7	55	1.9	2	2(15)	+	+	-	9.4	1.5
8	E(+)H(-)-8	68	2.0	2	4(17)	+	+	-	7.2	2.7
9	E(+)H(-)-9	42	2.6	2	0(3)	+	+	-	7.8	2.3
10	E(+)H(-)-10	64	1.5	2	0(4)	+	+	-	7.4	2.1
11	E(+)H(+)-1	61	2.2	3	2(18)	+	-	+	8.3	2.4
12	E(+)H(+)-2	43	2.5	3	0(5)	+	-	+	7.9	2.4
13	E(+)H(+)-3	41	1.6	3	0(3)	+	+	+	9.3	2.2
14	E(+)H(+)-4	37	1.5	3	0(18)	+	+	+	9.0	2.1
15	E(-)H(+)-1	44	3.2	3	0(1)	-	-	+	6.7	2.7
16	E(-)H(+)-2	56	1.4	3	0(3)	-	-	+	6.9	1.9
17	E(-)H(+)-3	76	2.3	3	2(20)	-	-	+	5.6	2.3
18	E(-)H(+)-4	64	2.3	3	5(21)	-	-	+	8.8	2.2
19	E(-)H(+)-5	73	1.6	3	0(2)	-	-	+	6.5	2.4
20	E(-)H(+)-6	54	2.9	3	0(12)	-	-	+	8.0	2.5
21	E(-)H(-)-1	51	1.6	3	1(23)	-	-	-	8.3	1.7
22	E(-)H(-)-2	54	3.6	2	0(11)	-	-	-	6.9	2.4
23	E(-)H(-)-3	62	1.6	3	1(21)	-	-	-	6.6	2.4
24	E(-)H(-)-4	47	1.5	3	0(4)	-	-	-	5.6	2.3
25	E(-)H(-)-5	57	2.9	2	0(4)	-	-	-	8.9	2.2

E, ER; H, Her2; FNAB, fine needle aspiration biopsy; FFPE, formalin-fixed, paraffin-embedded tissue; RIN, RNA integrity number. <sup>1</sup>Tumor size in cm, <sup>2</sup>Total number of nodes removed shown in parenthesis.

assess the quality and level of degradation of RNA, RIN (RNA integrity number) was assigned by the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

#### Illumina whole genome direct hybridization and DASL assays

All Illumina related services were provided by The Centre for Applied Genomics at the Hospital for Sick Children (Toronto, ON). Illumina Human-Ref8 BeadChip V3 24 K whole genome gene chips (Illumina Inc., San Diego, CA, USA) were used for both direct hybridization and DASL assays. A direct hybridization assay was used for FNAB specimens by using Illumina's standard protocol <http://www.illumina.com>. DASL assays were used for FFPE specimens according to Illumina's standard protocol. In brief, the input amount of RNA for the DASL assays was 400 ng of total RNA obtained from FFPE blocks. The RNA was first converted to cDNA through a reverse transcription reaction with biotinylated primers. The biotinylated cDNA was then annealed to assay oligonucleotides, and bound to streptavidin conjugated paramagnetic particles. After the

oligonucleotide hybridization, mishybridized and non-hybridized oligonucleotides were washed away. The hybridized oligonucleotides were then extended and ligated. These products formed a synthetic template that was transferred to a PCR reaction containing a fluorescently labelled primer. The labelled PCR product strand was then isolated, and the fluorescent products hybridized to a BeadChip. The BeadChip was then washed and imaged on the BeadArray Reader.

#### Quantitative real-time reverse transcription-PCR (RT-qPCR)

The RNA expression level of the 14 most differentially expressed genes (Table 2) was assayed using RT-qPCR with TaqMan gene expression assays in a 7900 sequence detector (Applied Biosystems, Foster City, CA). One µg of total RNA from a FFPE specimen was reverse transcribed in a 20 µL final reaction volume, using SuperScript VILO cDNA synthesis kit. The amount of cDNA corresponding to 10 ng of RNA was used in 10 µL reactions with the TaqMan Universal PCR Master Mix and corresponding sequence-specific primers assay mix (Applied Biosystems). Human HeLa cell RNA was used as the calibration sample, and the housekeeping gene

**Table 2 14 selected differential expression genes.**

Gene Symbol	The 25 breast cancers					NKI-295 dataset			GIS-251 dataset		
	FNAB array p <sup>1</sup>	FFPE array p <sup>1</sup>	FFPE qPCR p <sup>1</sup>	FFPE vs. FNAB r <sup>2</sup>	FFPE vs. qPCR r <sup>3</sup>	Array p <sup>1</sup>	OS p <sup>4</sup>	DMFS p <sup>4</sup>	Array p <sup>1</sup>	OS p <sup>4</sup>	RFS p <sup>4</sup>
AKR7A3	0.001	0.000	0.000	<b>0.701</b>	<b>0.653</b>	0.000	0.000	0.248	0.000	0.284	0.787
ANKRA2	0.001	0.009	0.021	<b>0.522</b>	0.390	0.000	0.021	0.071	0.000	0.064	0.139
ANLN	0.000	0.002	0.007	<b>0.731</b>	<b>0.543</b>	0.000	0.002	0.010	0.000	0.009	0.004
CA12	0.000	0.000	0.000	<b>0.607</b>	<b>0.710</b>	0.000	0.010	0.299	0.000	0.358	0.468
DNALI1	0.000	0.000	0.000	<b>0.681</b>	<b>0.646</b>	0.000	0.009	0.097	0.000	0.565	0.582
ERBB2	0.000	0.008	0.000	0.370	<b>0.815</b>	0.000	0.005	0.024	0.000	0.006	0.001
ESR1	0.000	0.000	0.000	<b>0.787</b>	<b>0.862</b>	0.000	0.004	0.030	0.000	0.698	0.742
FUT3	0.001	0.001	0.003	<b>0.542</b>	0.444	0.000	0.061	0.009	0.000	0.023	0.287
GREB1	0.001	0.001	0.000	<b>0.732</b>	<b>0.711</b>	0.000	0.054	0.667	0.000	0.023	0.487
KIAA1407	0.000	0.000	0.010	<b>0.460</b>	0.337	0.000	< 0.001	0.000	0.000	0.006	0.197
KIF2C	0.000	0.001	0.020	<b>0.710</b>	<b>0.647</b>	0.000	< 0.001	0.000	0.000	0.022	0.004
LONRF2	0.001	0.007	0.004	<b>0.398</b>	<b>0.588</b>	0.000	0.001	0.168	0.000	0.294	0.393
MAPT	0.000	0.000	0.000	<b>0.740</b>	<b>0.677</b>	0.000	0.007	0.023	0.000	0.001	0.014
PGR	0.000	0.000	0.000	<b>0.870</b>	<b>0.674</b>	0.000	< 0.001	< 0.001	0.000	0.022	0.206

FNAB, fine needle aspiration biopsy; FFPE, formalin-fixed, paraffin-embedded tissue; qPCR, Quantitative RT-PCR; OS, Overall survival; DMFS, Distant metastasis free survival; RFS, Relapse free survival; <sup>1</sup>P-value of ANOVA test among the 4 ER/Her2 groups; <sup>2</sup>Correlation coefficient, Pearson correlation test between FFPE array and FNA array (value in boldface type is statistically significant,  $P < 0.05$ ); <sup>3</sup>Correlation coefficient, Pearson correlation test between FFPE array and FFPE qPCR (value in boldface type is statistically significant,  $P < 0.05$ ); <sup>4</sup>P-value of Log-rank Test.

GAPDH served for the standardization of the individual PCR reactions.

#### Data analysis

Scanned microarray image data were used to process expression data by Illumina Gene Expression Module of GenomeStudio Software. The microarray gene expression data was normalized using background subtraction and Quantile methods. Sequence Detection Software (Applied Biosystems) was used to obtain the RT-qPCR amplification plots to quantify gene expression values using the cycle threshold method. All data were represented as log<sub>2</sub> ratios for the expression analysis of gene transcriptions. ANOVA and t-test were used to evaluate the variant significance of gene expression in different groups. Pearson correlation was used to measure the expression similarities between FNAB and FFPE specimens as well as between the microarray and RT-qPCR levels. Hierarchical clustering analysis was used to present gene expression patterns. Kaplan-Meier analysis was used to compare patients' survivals in differential gene expression groups, and the differences were determined by the log-rank test. The microarray data have been deposited in NCBI's Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo> and are accessible through GEO series accession number GSE23386. Two publicly available microarray datasets from 295 breast cancers of the Netherlands Cancer Institute (NKI-295) [7], and from 251 breast cancers of the Genome Institute of Singapore (GIS-251) [25] were used for validation analyses.

## Results

### Quality and quantity of FFPE RNA

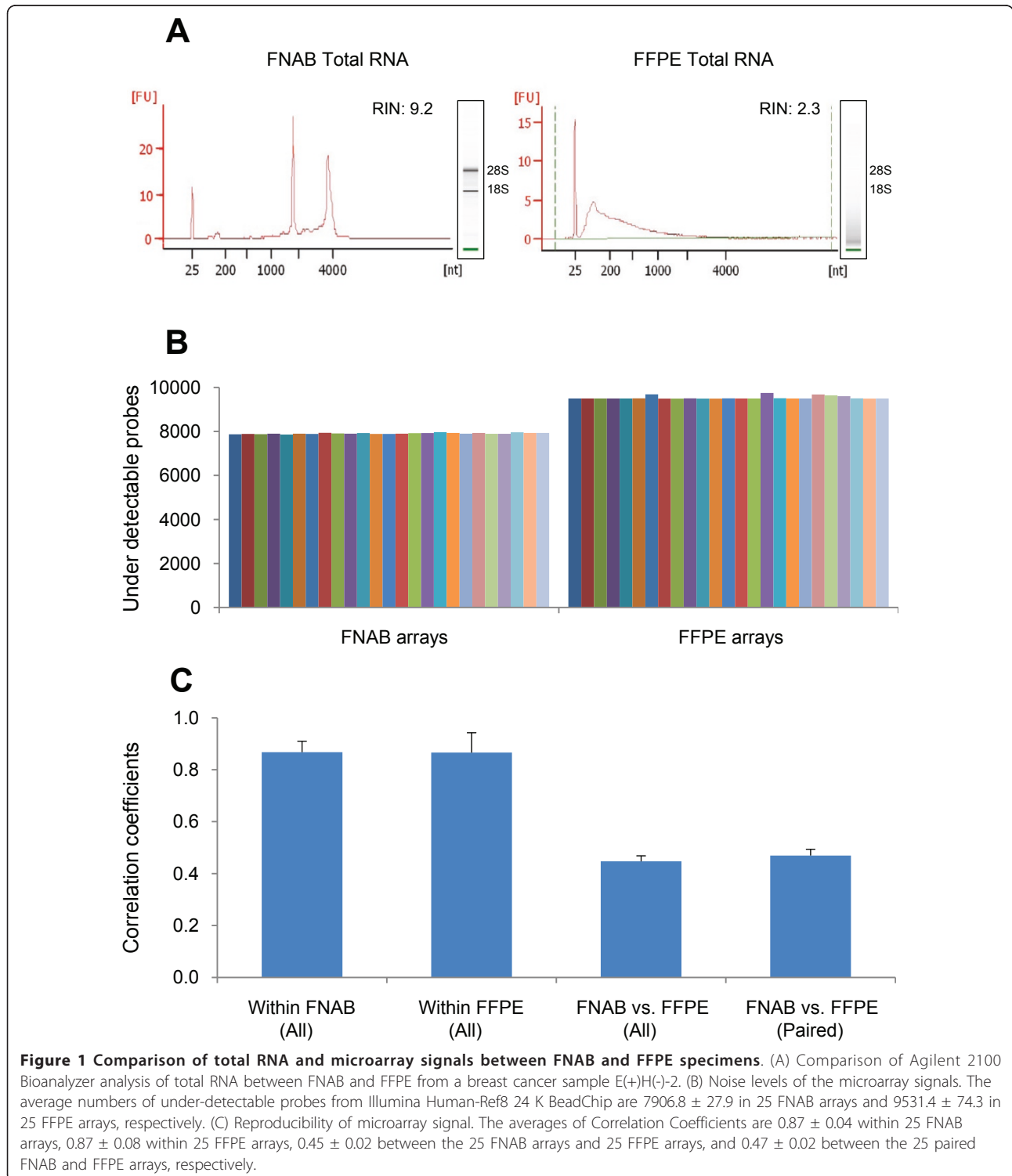
Initially, as a pilot study, we extracted RNA from FFPE specimens of eight human mammary reduction mammaplasty cases and generated DNA microarray data by using the DASL assay and Illumina Genome Wide HumanRef-8 BeadChip. When we compared the gene expression signals between the replicates, the average correlation coefficient was as high as  $0.96 \pm 0.02$  (data not shown). In this study, we evaluated 50 FFPE blocks from the clinical surgical pathology, and total RNA was extracted from 39 FFPE blocks. Eleven cases were excluded due to small tumor size or unavailability of a suitable FFPE block. Average total RNA yielded from FFPE samples was 4.3 µg (range 0.5 - 10 µg). Spectrophotometric 260/280 ratio of extracted RNA samples ranged from 1.5 - 2. As expected, the landmark ribosomal peaks (18S and 28S) were not well detected in the FFPE samples using Agilent 2100 Bioanalyzer (Figure 1A). However, samples from 32 cases (82%) had RIN values above 1.5 and were considered suitable for subsequent DNA microarray analyses [17].

### Comparison of RNA and array signal between FFPE and FNAB

Out of the 32 FFPE samples, 25 had corresponding FNAB specimens. We performed DASL assay and subsequent DNA microarray hybridization on the 25 paired FFPE samples. The 25 paired FNAB RNAs were used for direct hybridization assays to the same Illumina HumanRef-8 V3 platform. As expected, RNA from FFPE samples had lower RIN values than that of RNA extracted from

fresh FNAB (Table 1 and Figure 1A). In all 25 breast cancers, when we compared between the two different specimen types, gene expression status from FFPE specimens tended to have higher noise levels defined by a higher number of gene probes not detectable during the DNA

microarray analyses (Figure 1B). When compared within the two specimen types, the expression signals within both FFPE samples and FNAB had similarly high average correlation coefficients ( $0.87 \pm 0.04$  and  $0.86 \pm 0.08$ , respectively) supporting the reproducibility of the DNA



microarray data generated within each specimen type. They were less similar between specimen types with the average correlation coefficients of 0.45, which was likely due to the dominant effects on the RNA related to the tissue processing for FFPE specimens and/or the different hybridization techniques during DNA microarray analyses (Figure 1C).

#### Comparison of ER and Her2 expression between FFPE samples and FNAB

To determine whether FFPE samples could yield the same biologically and clinically relevant information as the FNAB, ER and Her2 clinical status detected by immunohistochemistry or fluorescence in situ hybridization were compared to the expression levels of *ESR1* and *ERBB2* in microarray expression and RT-qPCR analyses (Figure 2). The two clinical receptors' status remained reproducible as illustrated by a very strong correlation within the two specimen types FFPE ( $P = 0.0005$ ) and FNAB ( $P < 0.0001$ ) in the microarray expression level. As well, the two clinical receptors were also validated by using RT-qPCR in FFPE specimens ( $P = 0.0002$ ).

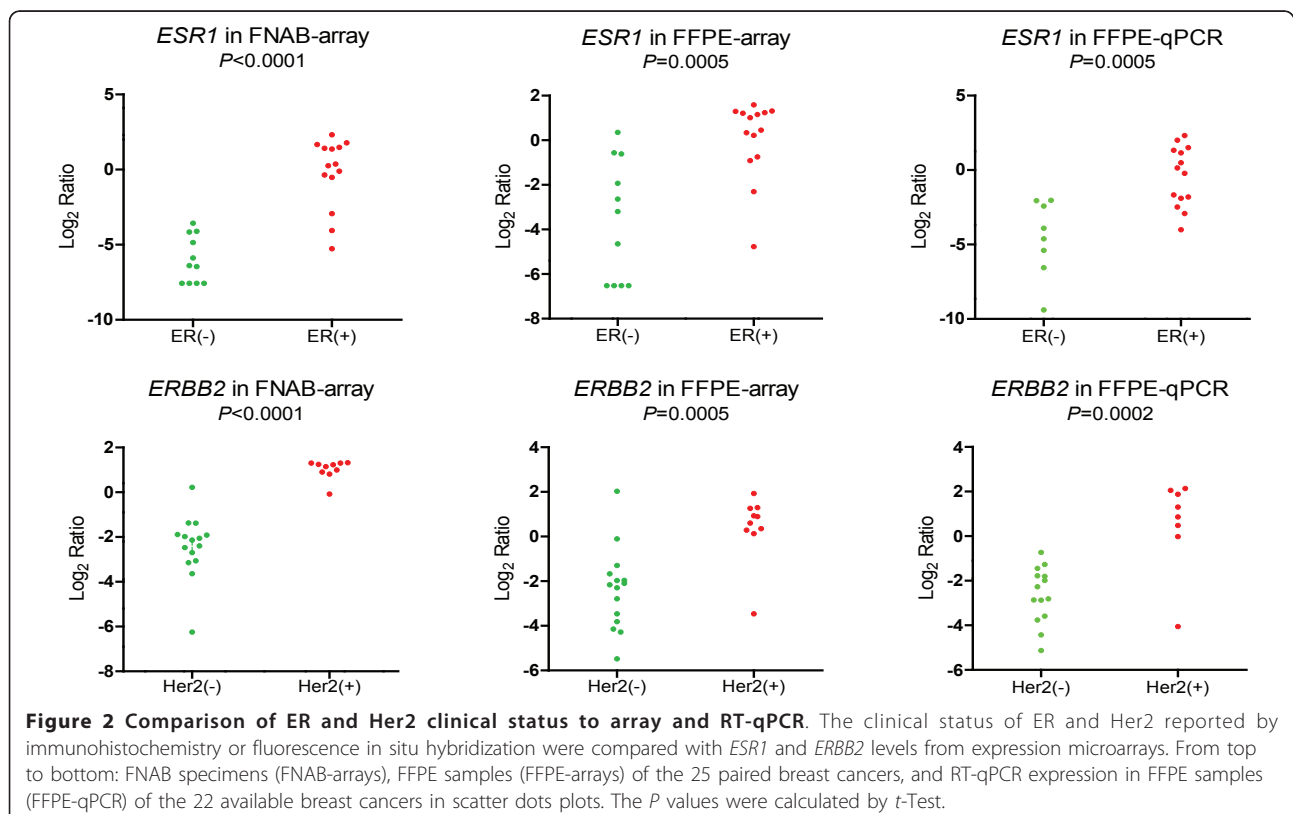
#### Comparison of gene profiling between FFPE samples and FNAB

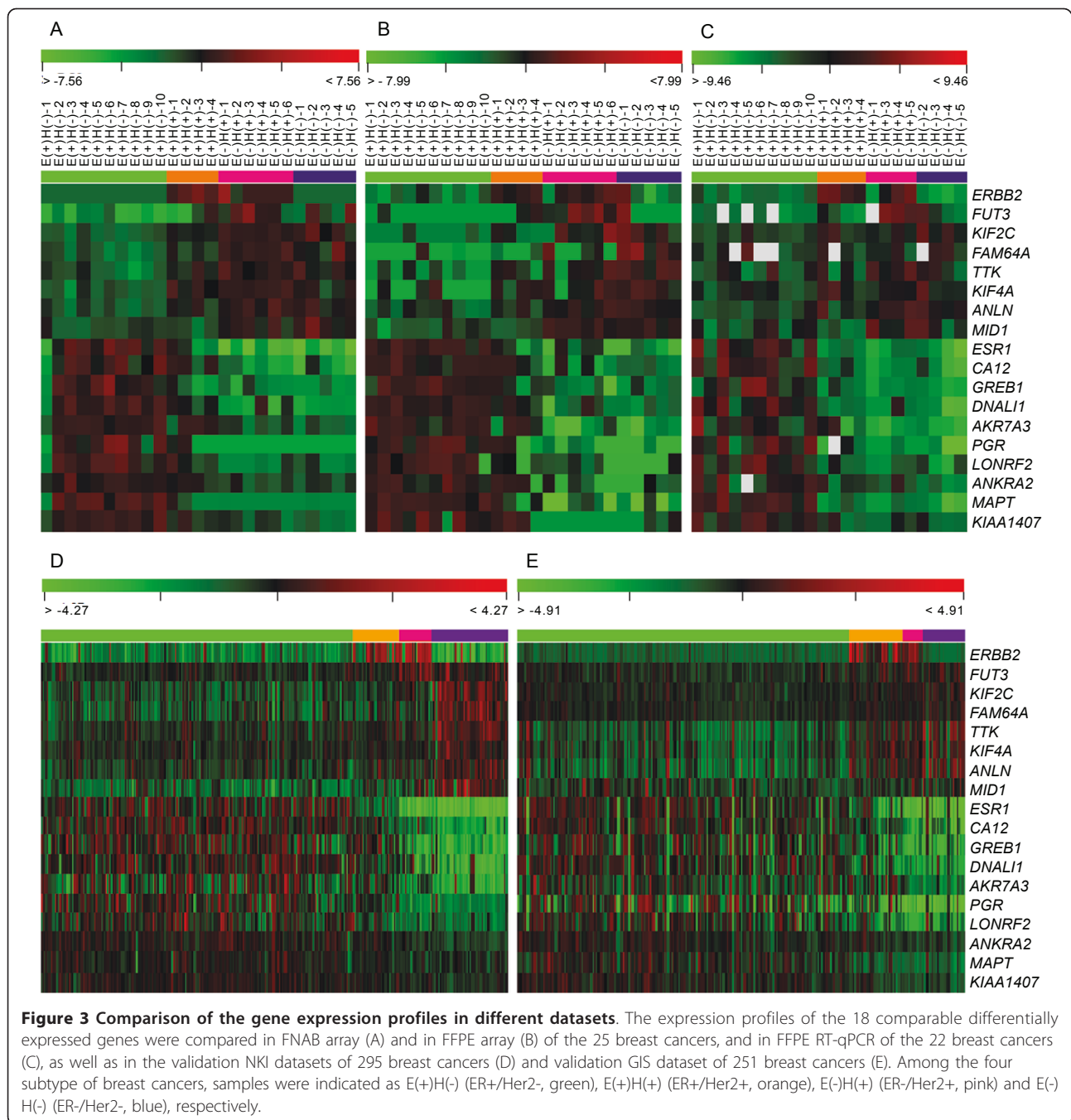
To further demonstrate whether gene profiles generated by FFPE and FNAB arrays can differentiate tumors into

their different ER/Her2 subtypes, we performed an ANOVA test to determine the genes that were most differentially expressed among the four distinct subtypes. As a result, 485 differentially expressed probes from FNAB arrays ( $P < 0.001$ ) and 258 probes from FFPE arrays ( $P < 0.01$ ) were obtained, of which 39 probes (representing 38 genes) overlapped between the two specimen types. The fact that all these 38 overlapped genes were differentially expressed in the same manner in both specimen types allowed us to directly compare the expression patterns between the two. A hierarchical clustering analysis using the 39 overlapped probes (data not shown) and a subset of 18 genes for both array and RT-qPCR analyses (see the section below for details) showed an identical expression pattern differentiating the four subtypes using FNAB arrays (Figure 3A) and FFPE arrays (Figure 3B).

#### Validation of gene profiling

Since most of the publicly available microarray data were generated using fresh tumor specimens, we tested the 38 genes that were reproducible between fresh and FFPE specimens to see if their expression pattern could be correlated with long-term clinical outcome data available in these public databases. Out of the 38 genes, 28 genes were reported in the two public databases NKI-295 [7] and GIS-251 [25]. Interestingly, these genes





were differentially expressed in a similar fashion when we divided the tumors into the same four tumor subtypes (ANOVA test,  $P < 0.001$ ). In addition, out of 28 genes, 21 were significantly correlated to patients' survivals (Log-rank test,  $P < 0.05$ ). Consequently, the 21 genes were chosen for further RT-qPCR analysis for validation of the array expression on FFPE specimens. Unfortunately, by using the commercial RT-qPCR reactions (TagMan, Applied Biosystems), three RT-qPCR reactions out of the 21 genes did not produce any

readable signals, and three samples out of 25 tumors failed to amplify in most of the RT-qPCR reactions. As the result, the remaining 18 genes and 22 FFPE specimens that were successful in RT-qPCR were used for subsequent analyses. When comparing the differential expression of the 18 genes, we found a near-identical gene expression pattern in FNAB arrays (Figure 3A), FFPE arrays (Figure 3B) and FFPE RT-qPCR (Figure 3C), as well as NKI-295 arrays (Figure 3D) and GIS-251 dataset (Figure 3E). Out of the 18 genes detected by

RT-qPCR in the 22 FFPE specimens, 14 genes were significantly differentially expressed among the four clinical subtypes (ANOVA test,  $P < 0.05$ ) as well as their array expression by using both FNAB and FFPE specimens, and in the two validation datasets. The final 14 genes are listed in Table 2.

#### The biomarkers for prognosis of breast cancers

When we divided the tumors into the four clinical subtypes, all 14 genes were differentially expressed in a similar fashion in all three tests analyses (FNAB, FFPE and RT-qPCR) in the training sets and the two independent published microarray datasets, served as validation sets. When looking at the survival data from the two validation datasets, these 14 genes showed a significant association between their expression levels in the breast cancers and the clinical outcome of the patients: overall survival and distant metastasis free survival in 295 patients from the NKI dataset, and overall survival and relapse free survival in 251 patients from the GIS dataset (Table 2). These genes could be used as potential biomarkers for predicting the clinical outcomes in breast cancer patients, in both fresh FNAB specimens and archival FFPE specimens. Specifically, two genes, *ANLN* (anillin, actin binding protein) and *KIF2C* (kinesin family member 2C) were under-expressed in ER+ tumors and over-expressed in Her2+ and ER-/Her2- tumors; whereas *MAPT* (microtubule-associated protein tau) was over-expressed in ER+ tumors and under-expressed in Her2+ or ER-/Her2- tumors (Figure 4). The over-expression of *ANLN* and *KIF2C*, and the under-expression of *MAPT* consistently showed a strong correlation with poor survival in the breast cancer patients from both validation datasets (Figure 5). These findings demonstrated that we can generate informative microarray data from FFPE specimens, and the expression levels of a subset of genes are reproducible and informative when compared to FNAB specimens. The prognostic information of these genes is preserved in FFPE specimens.

#### Discussion

Since Rupp and Locker [26] reported their first successful RNA extraction from FFPE specimens in 1988, significant strides have been made to enable RNA profiling from FFPE tissues, including efforts to standardize tissue handling and fixation procedures and improving RNA extraction methodologies [27,28]. Very few technologies have emerged despite these advancements that are capable of whole transcriptome profiling from archived FFPE material [24,29,30]. Initial attempts at DNA microarray analyses using FFPE samples yielded poor reproducibility [31], or loss of detection of gene signatures when compared with matched fresh samples [32].

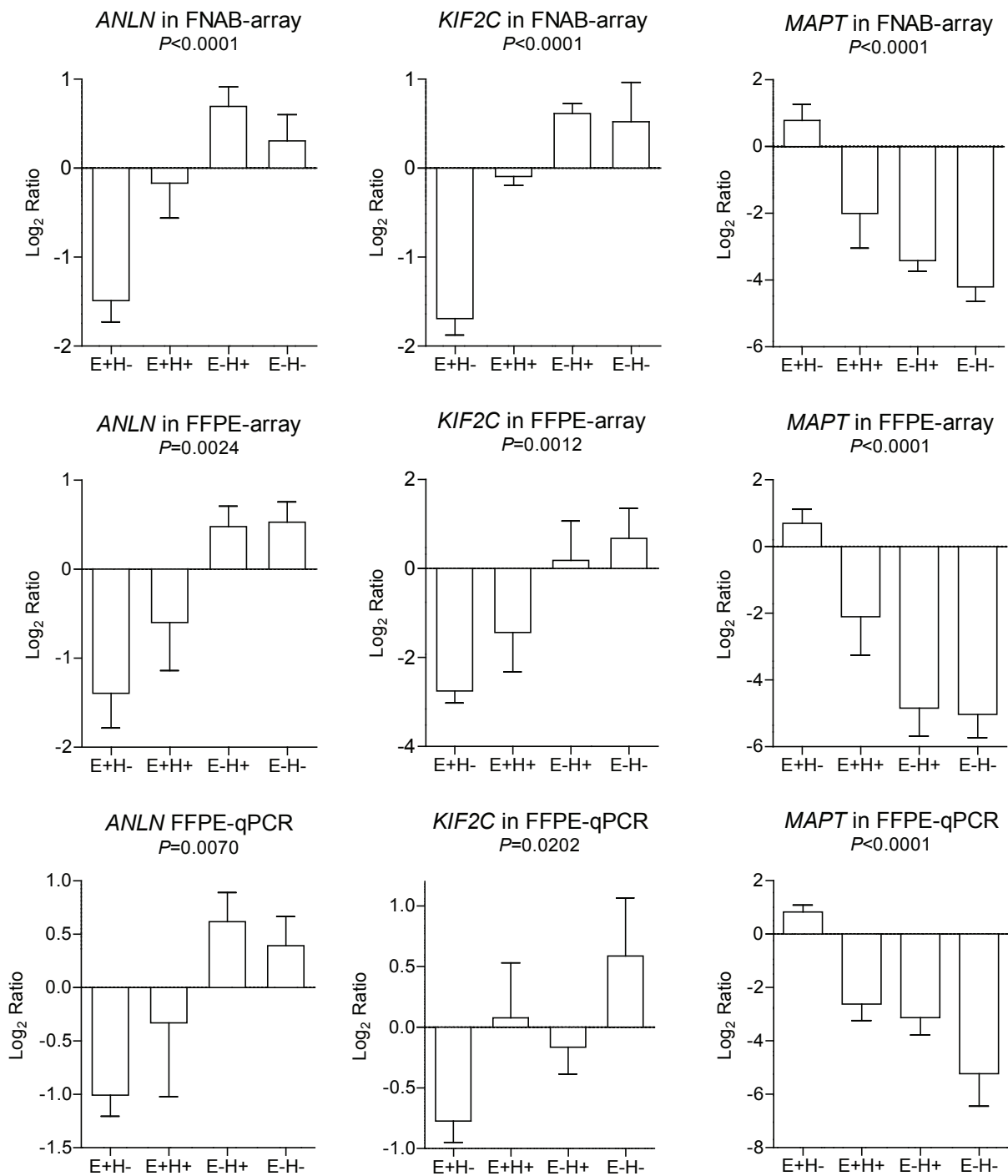
Several commercial and academic endeavours have recently become more successful in their ability to retrieve meaningful biological information from degraded FFPE-derived RNAs [15,21,22,24,33-36].

The DASL assay incorporates random priming during cDNA synthesis, and therefore does not depend solely on the polyA/oligo-dT based priming process used in conventional DNA microarray methods. In addition, the assay requires a relatively short target sequence of about 50 nucleotides to query oligonucleotide annealing; which improved its ability to quantify fragmented RNA species [37]. Using these technologies, we evaluated the feasibility of using FFPE samples for DNA microarray analyses. Initially, as a pilot study, RNA from FFPE blocks of human mammary reduction mammoplasty tissues were extracted and used to generate DNA microarray data using DASL assay and Illumina HumanRef8 BeadChip, a genome-wide gene panel that contained over 24 thousands genes. Within eight replicates of the same type of specimens, the average correlation coefficient of array expression signals was very high at  $0.96 \pm 0.02$  which showed an impressive technological reproducibility. This is similar to a recent study by Ton et al who reported a high correlation ( $r = 0.98$ ) among 12 technical replicates by using the DASL technology [21].

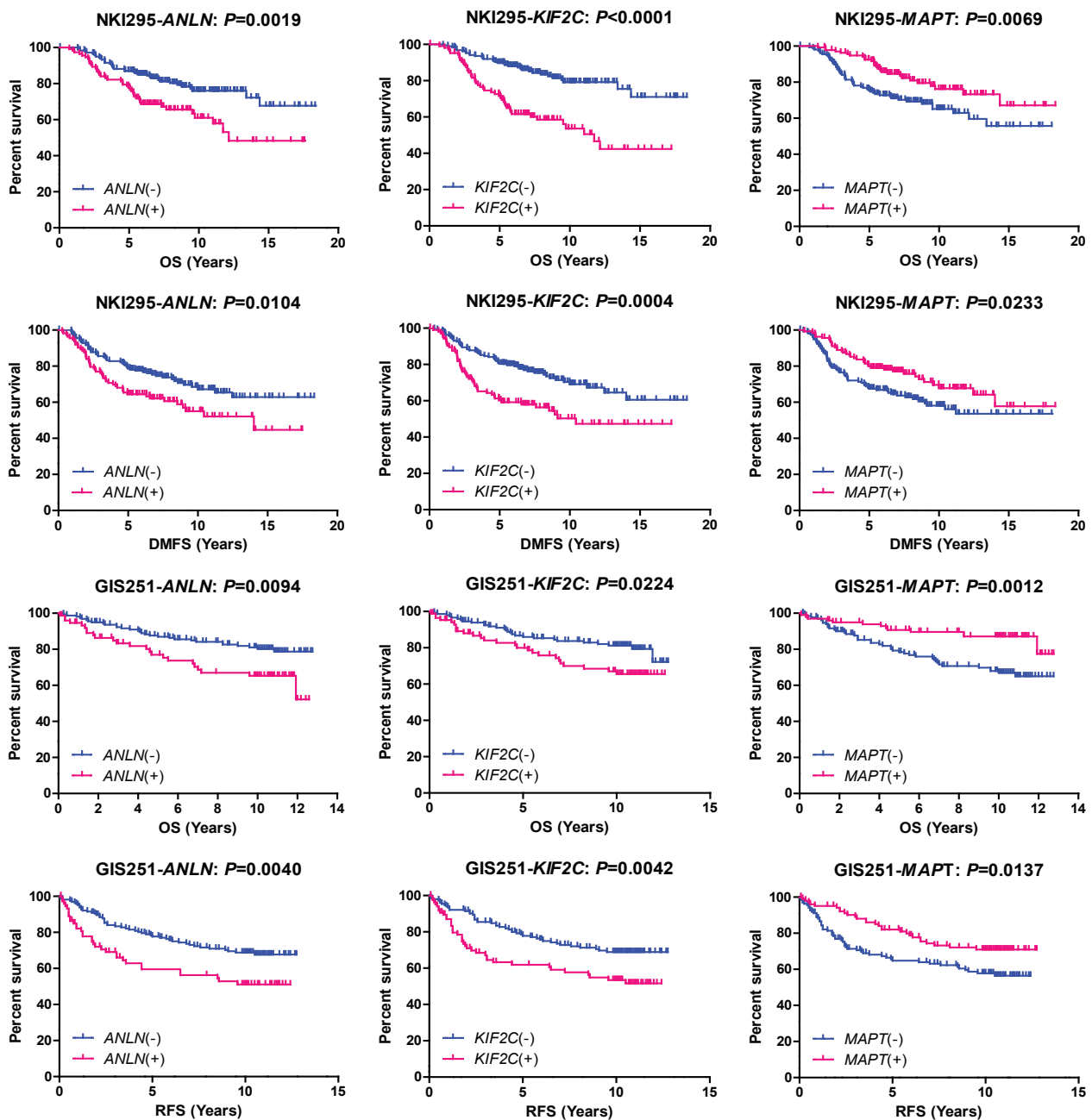
We carried out our current study using human breast cancer specimens to test the feasibility of performing such experiments with a future plan to extend the technology to a large clinical library of breast cancers. Using RNA extraction methods described recently [23], we found that over 80% of the RNA extracted from archival FFPE tumor samples could be used for subsequent DASL analyses and produced DNA microarray data that were informative and could be validated and compared to that obtained from fresh FNAB specimens collected from the same patients.

In this study, we demonstrated a high reproducibility of the DNA microarray data when comparing the array signals within the FNAB or FFPE tumor specimens ( $r = 0.87$ ). The comparison between the two specimen types yielded lower correlation coefficients ( $r = 0.45$ ), which was likely due to dominant effects on the RNA related to the tissue processing for FFPE specimens and the different hybridization techniques in DNA microarray protocols (Figure 1C). Nevertheless, some clinically relevant microarray data remained reproducible as illustrated by a very strong correlation between the microarray expression levels of *ESR1* and *ERBB2* to the clinical status of ER and Her2 respectively in both FFPE and FNAB samples, which were also validated with RT-qPCR in FFPE specimens. This result was consistent with two recent studies by Ton et al [21] and Mulligan et al [22]. The gene profiling of FFPE samples suffered from a higher ratio of noise-to-signal (Figure 1B) and





**Figure 4 Differential expression of the biomarkers.** The histograms show the comparison of the average expression values of the three prognostic genes *ANLN*, *KIF2C* and *MAPT* from left to right respectively, which were shown to be significantly different among the four ER/Her2 clinical status groups [ER+/Her2- (E+H-), ER+/Her2+ (E+H+), ER-/Her2+ (E-H+), and ER-/Her2- (E-H-)], but reproducible among the microarray data from FNAB (FNAB-array) and FFPE samples (FFPE-array) of the 25 paired breast cancers, and RT-qPCR expression in FFPE samples (FFPE-qPCR) of the 22 available breast cancers. The *P* values were calculated by ANOVA test.



**Figure 5 Kaplan-Meier survival curves.** The association between the gene expression of *ANLN*, *KIF2C* and *MAPT*, from left to right respectively, and the clinical outcomes in patients' overall survival (OS) or distant metastasis free survival (DMFS) from the 295 patients in the NKI validation dataset (NKI295), and OS or relapse free survival (RFS) from the 251 patients in the GIS validation dataset (GIS251), from top to bottom respectively. *P* values were calculated by log-rank test.

thus detected a smaller number of differentially expressed genes compared to FNAB samples (258 vs. 485). Therefore, we anticipate that the identification of subtle changes in expression levels in FFPE samples will remain challenging using current technologies.

To date, high-throughput gene expression profiling has demonstrated the potential uses of gene profiling as

molecular subtype classifiers [6], prognostic indicators [7,8], and treatment predictors [9] by using fresh breast cancer specimens. If gene profiling becomes a standard part of the pathological assessment of a tumor in the future, the use of FFPE material would be advantageous, as it obviates the need for fresh bio-specimens that can be very time sensitive and often impractical to collect in

routine clinical settings. In certain clinical situations, including recurrent diseases and evaluation of long-term endocrine therapy, when there is no fresh bio-specimen from the original primary tumors; gene profiling using FFPE may provide an alternative to allow evaluation of the prognostic or predictive gene profiles.

The use of FFPE samples for gene profiling studies has been considered suboptimal in the past due to concerns about RNA degradation and less than 5% of all microarray studies to date have been reported using archival tumor tissues that were formalin-fixed and paraffin-embedded. Very few microarray studies have been conducted to investigate the gene profiling of FFPE samples and compared the results with their matched fresh frozen tissue. Some data is available in carcinomas of colon, liver and breast [4,5,15,35]. In breast cancer, Da Silva et al [20] used FFPE samples from invasive lobular carcinoma only to check the expression profile of E-cadherin. Ravo et al. [17] used RNA from cultured cell lines, cryopreserved tumors and FFPE samples from breast cancers. Waddell et al. [19] and Bibikova et al. [16] included both fresh frozen and FFPE samples in their studies to compare the gene profiling of breast cancers in both type of tissues. However, their studies focused only on the technical aspect without any clinical correlations. We have correlated our gene profiling results to clinical receptor status, and demonstrated the ability to use gene expression profiling from FFPE to differentiate tumors from 4 distinct tumor subtypes. We identified a set of 38 genes that were reproducibly able to identify the different clinical groups in both FFPE and FNAB specimens. These findings suggested that there are subsets of genes in FFPE specimens that could maintain their clinical relevance despite the tissue processing effects. Interestingly, subsets of these genes were correlated to clinical outcomes as demonstrated in our study using publicly available microarray validation databases.

Although we are the first to show that the over-expression of *ANLN* and *KIF2C*, and the under-expression of *MAPT* predict for poor survival in patients with breast cancer (Figures 4 and 5), there is some evidence that supports the correlation of these three genes with prognosis and carcinogenesis in other cancers, and treatment in breast cancers. The over-expression of *ANLN* has been reported to be a biomarker for pancreatic carcinoma [38], and predicted for poor survival in early lung cancers [39]. Shimo et al reported that the over-expression of *KIF2C* might be involved in breast carcinogenesis and is a therapeutic target for breast cancers [40]. The expression of *MAPT* has been correlated to the sensitivity to chemotherapies in breast cancer [41,42]. These results strongly support that *ANLN*, *KIF2C*, and *MAPT* could play a role in the carcinogenesis, treatment and prognosis of breast cancers.

FFPE samples are widely available and can be linked to clinical outcome information, often available through institutional or provincial cancer registries. The availability of large libraries of archival FFPE tissue samples could potentially solve some of the most critical challenges that investigators face when using prospectively collected specimens, namely, patient recruitment and expensive long-term follow-up. Validation studies can also be done fairly quickly by selecting an independent patient population annotated with critical long-term clinical outcome data. Once the technology for gene profiling using FFPE matures, it will likely play an important role in the clinical management of breast cancer patients.

## Conclusion

There is tremendous potential in using FFPE specimens for gene profiling, especially in breast cancer, in which there are existing data to support the prognostic and predictive implications of tumor gene profiles. Based on the comparison of gene profiling on FFPE breast cancer specimens and matched fresh specimens, we can conclude that important clinically relevant information can be identified using FFPE specimens and we even demonstrated the potential for using the microarray data to be used as a prognostic tool. Further improvements in current DNA microarray technologies will likely to bring gene profiling of breast cancer into routine clinical practice as we move towards personalized breast cancer care.

## Acknowledgements

This study was partially funded by funds from Genome Canada and Clinical Research Program at University Health Network. We thank Vietty Wong for technical assistance.

## Author details

<sup>1</sup>Ontario Cancer Institute and Campbell Family Institute for Breast Cancer Research, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada. <sup>2</sup>Laboratory Medicine Program, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada. <sup>3</sup>Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, 1 King's College Circle, Toronto, Ontario, M5S 1A8, Canada. <sup>4</sup>Department of Medical Biophysics, University of Toronto, Ontario Cancer Institute, Princess Margaret Hospital, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada. <sup>5</sup>Surgical Oncology, Princess Margaret Hospital, University Health Network, 610 University Avenue, Toronto, Ontario, M5G 2M9, Canada.

## Authors' contributions

The study was designed by WLL and SJD and conducted by AMS. DYW performed data analyses, and manuscript submission. Manuscript was prepared by AMS, edited and modified by WLL, DYW and SJD, and all authors proofread the manuscript. SJD, BJJ, NM and SB performed the pathology for the study.

## Competing interests

The authors declare that they have no competing interests.

Received: 22 December 2010 Accepted: 16 June 2011  
Published: 16 June 2011

## References

- van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH: **Gene expression profiling predicts clinical outcome of breast cancer.** *Nature* 2002, **415**:530-536.
- Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J, Wolmark N: **A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer.** *N Engl J Med* 2004, **351**:2817-2826.
- van't Veer LJ, Bernards R: **Enabling personalized cancer medicine through analysis of gene-expression patterns.** *Nature* 2008, **452**:564-570.
- Lassmann S, Kreutz C, Schoepflin A, Hopt U, Timmer J, Werner M: **A novel approach for reliable microarray analysis of microdissected tumor cells from formalin-fixed and paraffin-embedded colorectal cancer resection specimens.** *J Mol Med* 2009, **87**:211-224.
- Hoshida Y, Villanueva A, Kobayashi M, Peix J, Chiang DY, Camargo A, Gupta S, Moore J, Wrobel MJ, Lerner J, Reich M, Chan JA, Glickman JN, Ikeda K, Hashimoto M, Watanabe G, Daidone MG, Roayaie S, Schwartz M, Thung S, Salvesen HB, Gabriel S, Mazzaferro V, Bruix J, Friedman SL, Kumada H, Llovet JM, Golub TR: **Gene expression in fixed tissues and outcome in hepatocellular carcinoma.** *N Engl J Med* 2008, **359**:1995-2004.
- Glas AM, Floore A, Delahaye LJ, Witteveen AT, Pover RC, Bakx N, Lahti-Domenici JS, Bruinsma TJ, Warmoes MO, Bernards R, Wessels LF, Van't Veer LJ: **Converting a breast cancer microarray signature into a high-throughput diagnostic test.** *BMC Genomics* 2006, **7**:278.
- van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R: **A gene-expression signature as a predictor of survival in breast cancer.** *N Engl J Med* 2002, **347**:1999-2009.
- Chang HY, Nuyten DS, Sneddon JB, Hastie T, Tibshirani R, Sørlie T, Dai H, He YD, van't Veer LJ, Bartelink H, van de Rijn M, Brown PO, van de Vijver MJ: **Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival.** *Proc Natl Acad Sci USA* 2005, **102**:3738-3743.
- Troester MA, Hoadley KA, Sørlie T, Herbert BS, Børresen-Dale AL, Lønning PE, Shay JW, Kaufmann WK, Perou CM: **Cell-type-specific responses to chemotherapeutics in breast cancer.** *Cancer Res* 2004, **64**:4218-4226.
- 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.
- Balogh GA, Russo J, Mailo DA, Heulings R, Russo PA, Morrison P, Sheriff F, Russo IH: **The breast of parous women without cancer has a different genomic profile compared to those with cancer.** *Int J Oncol* 2007, **31**:1165-1175.
- 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.
- Madabusi LV, Latham GJ, Andrus BF: **RNA extraction for arrays.** *Methods Enzymol* 2006, **411**:1-14.
- Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ: **Clinical application of the 70-gene profile: The MINDACT trial.** *J Clin Oncol* 2008, **26**:729-735.
- Duenwald S, Zhou M, Wang Y, Lejnine S, Kulkarni A, Graves J, Smith R, Castle J, Tokiwa G, Fine B, Dai H, Fare T, Marton M: **Development of a microarray platform for FFPET profiling: Application to the classification of human tumors.** *J Transl Med* 2009, **7**:65.
- 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.
- Ravo M, Mutarelli M, Ferraro L, Grober OM, Paris O, Tarallo R, Vigilante A, Cimino D, De Bortoli M, Nola E, Cicatiello L, Weisz A: **Quantitative expression profiling of highly degraded RNA from formalin-fixed, paraffin-embedded breast tumor biopsies by oligonucleotide microarrays.** *Lab Invest* 2008, **88**:430-440.
- Lee J, Hever A, Willhite D, Zlotnik A, Hevezi P: **Effects of RNA degradation on gene expression analysis of human postmortem tissues.** *FASEB J* 2005, **19**:1356-1358.
- Waddell N, Cocciardi S, Johnson J, Healey S, Marsh A, Riley J, da Silva L, Vargas AC, Reid L, kConFab Investigators, Simpson PT, Lakhani SR, Chenevix-Trench G: **Gene expression profiling of formalin-fixed, paraffin-embedded familial breast tumours using the whole genome-DASL assay.** *J Pathol* 2010, **221**:452-461.
- Da Silva L, Parry S, Reid L, Keith P, Waddell N, Kossai M, Clarke C, Lakhani SR, Simpson PT: **Aberrant expression of E-cadherin in lobular carcinomas of the breast.** *Am J Surg Pathol* 2008, **32**:773-783.
- Ton CC, Vartanian N, Chai X, Lin MG, Yuan X, Malone KE, Li CI, Dawson A, Sather C, Delrow J, Hsu L, Porter PL: **Gene expression array testing of FFPE archival breast tumor samples: an optimized protocol for WG-DASL® sample preparation.** *Breast Cancer Res Treat* .
- Mulligan JM, McDyer F, Deharo S, Farztdinov V, Halfpenny I, Delaney T, Couch F, Quinn JE, Harkin P, Kennedy R: **Identification of molecular subtypes within a formalin-fixed, paraffin-embedded breast cancer tumour cohort [abstract].** *Breast Cancer Res* 2010, **12**(Suppl 1):33.
- Wong V, Wang DY, Warren K, Kulkarni S, Boerner S, Done SJ, Leong WL: **The effects of timing of fine needle aspiration biopsies on gene expression profiles in breast cancers.** *BMC Cancer* 2008, **8**:277.
- Abramovitz M, Ordanic-Kodani M, Wang Y, Li Z, Catzavelos C, Bouzyk M, Sledge GW Jr, Moreno CS, Leyland-Jones B: **Optimization of RNA extraction from FFPE tissues for expression profiling in the DASL assay.** *Biotechniques* 2008, **44**:417-423.
- Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J: **An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival.** *Proc Natl Acad Sci USA* 2005, **102**:13550-13555.
- Rupp GM, Locker J: **Purification and analysis of RNA from paraffin-embedded tissues.** *Biotechniques* 1988, **6**:56-60.
- Medeiros F, Rigl CT, Anderson GG, Becker SH, Halling KC: **Tissue handling for genome-wide expression analysis: A review of the issues, evidence, and opportunities.** *Arch Pathol Lab Med* 2007, **131**:1805-1816.
- Hewitt SM, Lewis FA, Cao Y, Conrad RC, Cronin M, Danenberg KD, Goralski TJ, Langmore JP, Raja RG, Williams PM, Palma JF, Warrington JA: **Tissue handling and specimen preparation in surgical pathology: Issues concerning the recovery of nucleic acids from formalin-fixed, paraffin-embedded tissue.** *Arch Pathol Lab Med* 2008, **132**:1929-1935.
- Ribeiro-Silva A, Zhang H, Jeffrey SS: **RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: A comparison of column purification and magnetic bead-based technologies.** *BMC Mol Biol* 2007, **8**:118.
- Linton K, Hey Y, Dibben S, Miller C, Freemont A, Radford J, Pepper S: **Methods comparison for high-resolution transcriptional analysis of archival material on affymetrix plus 2.0 and exon 1.0 microarrays.** *Biotechniques* 2009, **47**:587-596.
- 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.
- Penland SK, Keku TO, Torrice C, He X, Krishnamurthy J, Hoadley KA, Woosley JT, Thomas NE, Perou CM, Sandler RS, Sharpless NE: **RNA expression analysis of formalin-fixed paraffin-embedded tumors.** *Lab Invest* 2007, **87**:383-391.
- Linton KM, Hey Y, Saunders E, Jeziorska M, Denton J, Wilson CL, Swindell R, Dibben S, Miller CJ, Pepper SD, Radford JA, Freemont AJ: **Acquisition of biologically relevant gene expression data by affymetrix microarray analysis of archival formalin-fixed paraffin-embedded tumours.** *Br J Cancer* 2008, **98**:1403-1414.
- Paik S, Kim CY, Song YK, Kim WS: **Technology insight: Application of molecular techniques to formalin-fixed paraffin-embedded tissues from breast cancer.** *Nat Clin Pract Oncol* 2005, **2**:246-254.
- Frank M, Doring C, Metzler D, Eckerle S, Hansmann ML: **Global gene expression profiling of formalin-fixed paraffin-embedded tumor samples: A comparison to snap-frozen material using oligonucleotide microarrays.** *Virchows Arch* 2007, **450**:699-711.
- Farragher SM, Tanney A, Kennedy RD, Paul Harkin D: **RNA expression analysis from formalin fixed paraffin embedded tissues.** *Histochem Cell Biol* 2008, **130**:435-445.
- 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.

38. Olakowski M, Tyszkiewicz T, Jarzab M, Król R, Oczko-Wojciechowska M, Kowalska M, Kowal M, Gala GM, Kajor M, Lange D, Chmielik E, Gubala E, Lampe P, Jarzab B: **NBL1 and anillin (ANLN) genes over-expression in pancreatic carcinoma.** *Folia Histochem Cytobiol* 2009, **47**:249-255.
39. Skrzypski M, Jassem E, Taron M, Sanchez JJ, Mendez P, Rzyman W, Gulida G, Raz D, Jablons D, Provencio M, Massuti B, Chaib I, Perez-Roca L, Jassem J, Rosell R: **Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung.** *Clin Cancer Res* 2008, **14**:4794-4799.
40. Shimo A, Tanikawa C, Nishidate T, Lin ML, Matsuda K, Park JH, Ueki T, Ohta T, Hirata K, Fukuda M, Nakamura Y, Katagiri T: **Involvement of kinesin family member 2C/mitotic centromere-associated kinesin overexpression in mammary carcinogenesis.** *Cancer Sci* 2008, **99**:62-70.
41. Tanaka S, Nohara T, Iwamoto M, Sumiyoshi K, Kimura K, Takahashi Y, Tanigawa N: **Tau expression and efficacy of paclitaxel treatment in metastatic breast cancer.** *Cancer Chemother Pharmacol* 2009, **64**:341-346.
42. Ikeda H, Taira N, Hara F, Fujita T, Yamamoto H, Soh J, Toyooka S, Nogami T, Shien T, Doihara H, Miyoshi S: **The estrogen receptor influences microtubule-associated protein tau (MAPT) expression and the selective estrogen receptor inhibitor fulvestrant downregulates MAPT and increases the sensitivity to taxane in breast cancer cells.** *Breast Cancer Res* 2010, **12**:R43.

#### Pre-publication history

The pre-publication history for this paper can be accessed here:  
<http://www.biomedcentral.com/1471-2407/11/253/prepub>

doi:10.1186/1471-2407-11-253

**Cite this article as:** Sadi et al.: Clinical relevance of DNA microarray analyses using archival formalin-fixed paraffin-embedded breast cancer specimens. *BMC Cancer* 2011 **11**:253.

**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](http://www.biomedcentral.com/submit)

