Clinical Chemistry 56:12
1845–1853 (2010)

Automated Extraction of DNA and RNA from a Single Formalin-Fixed Paraffin-Embedded Tissue Section for Analysis of Both Single-Nucleotide Polymorphisms and mRNA Expression

Guido Hennig,[1,2] Mathias Gehrmann,[1] Udo Stropp,[1] Hiltrud Brauch,[3] Peter Fritz,[3] Michel Eichelbaum,[3] Matthias Schwab,[3,4] and Werner Schroth[3*]

BACKGROUND: There is an increasing need for the identification of both DNA and RNA biomarkers from pathodiagnostic formalin-fixed paraffin-embedded (FFPE) tissue samples for the exploration of individualized therapy strategies in cancer. We investigated a fully automated, xylene-free nucleic acid extraction method for the simultaneous analysis of RNA and DNA biomarkers related to breast cancer.

METHODS: We copurified both RNA and DNA from a single 10-μm section of 210 paired samples of FFPE tumor and adjacent normal tissues (1–25 years of archival time) using a fully automated extraction method. Half of the eluate was DNase I digested for mRNA expression analysis performed by using reverse-transcription quantitative PCR for the genes estrogen receptor 1 (*ESR1*), progesterone receptor (*PGR*), v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) (*ERBB2*), epoxide hydrolase 1 (*EPHX1*), baculoviral IAP repeat-containing 5 (*BIRC5*), matrix metallopeptidase 7 (*MMP7*), vascular endothelial growth factor A (*VEGFA*), and topoisomerase (DNA) II alpha 170kDa (*TOP2A*). The remaining undigested aliquot was used for the analysis of 7 single-nucleotide polymorphisms (SNPs) by MALDI-TOF mass spectrometry.

RESULTS: In 208 of 210 samples (99.0%) the protocol yielded robust quantification-cycle values for both RNA and DNA normalization. Expression of the 8 breast cancer genes was detected in 81%–100% of tumor tissues and 21%–100% of normal tissues. The 7 SNPs were successfully genotyped in 91%–97% of tumor and 94%–97% of normal tissues. Allele concordance between tumor and normal tissue was 98.9%–99.5%.

CONCLUSIONS: This fully automated process allowed an efficient simultaneous extraction of both RNA and DNA from a single FFPE section and subsequent dual analysis of selected genes. High gene expression and genotyping detection rates demonstrate the feasibility of molecular profiling from limited archival patient samples.

The identification and validation of molecular markers in formalin-fixed paraffin-embedded (FFPE)[5] tissue has been a field of intensive research in recent years because expression profiling, genotyping, and mutation analysis have been increasingly shown to aid in cancer diagnosis and to guide cancer treatment. As yet, however, only a limited number of marker panels or individual markers developed from histopathological tissue specimens have been introduced into routine clinical practice [e.g., OncotypeDX®: *NCCN Clinical Practice Guidelines in Oncology (1)* and v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (*KRAS*)[6] mutation analysis for

[1] Siemens Healthcare Diagnostics Products, Molecular Research Germany, Cologne, Germany; [2] Siemens Healthcare Diagnostics, Eschborn, Germany; [3] Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, and University Tübingen, Germany; [4] Department of Clinical Pharmacology, University Hospital Tübingen, Germany.
* Address correspondence to this author at: Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Stuttgart, Auerbachstrasse 112, 70376 Stuttgart, Germany. Fax +49-711-859295; e-mail werner.schroth@ikp-stuttgart.de.
Received May 28, 2010; accepted September 20, 2010.
Previously published online at DOI: 10.1373/clinchem.2010.151233
[5] Nonstandard abbreviations: FFPE, formalin-fixed, paraffin-embedded; SNP, single-nucleotide polymorphism; GOI, genes of interest; MS, mass spectrometry; HWE, Hardy-Weinberg equilibrium; RT-qPCR, reverse transcription–quantitative PCR; Cq, quantification cycle; MAF, minor allele frequency.
[6] Human genes: *KRAS*, v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog; *ESR1*, estrogen receptor 1; *PGR*, progesterone receptor; *ERBB2*, v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian); *EPHX1*, epoxide hydrolase 1; *BIRC5*, baculoviral IAP repeat-containing 5; *MMP7*, matrix metallopeptidase 7; *VEGFA*, vascular endothelial growth factor A; *TOP2A*, topoisomerase (DNA) II alpha 170kDa; *RPL37A*, ribosomal protein L37a; *PAEP*, progestagen-associated endometrial protein; *ESR2*, estrogen receptor 2; *NCOA3*, nuclear receptor coactivator 3; *PPARGC1B*, peroxisome proliferator-activated receptor gamma, coactivator 1 beta; *RRAS2*, related RAS viral (r-ras) oncogene homolog 2; *CYP2D6*, cytochrome P450, family 2, subfamily D, polypeptide 6; *HOXB13*, homeobox B13; *IL17RB*, interleukin 17 receptor B.

targeted therapies based on epidermal growth factor receptor *(2)*]. FFPE tissue seems to be an invaluable sample source in translational research related to biomarkers based on DNA, RNA, and microRNA because FFPE tissue samples are readily available through their use in routine clinical diagnosis. This material is attractive for use in research because nucleic acids can be extracted from decades-old FFPE samples and large archives of FFPE tissue from various disease entities linked to clinical and follow-up databases can be easily accessed *(3)*.

Although analysis of protein expression in pathological specimens by immunohistochemical detection is well established and part of routine diagnostic practice, mRNA analysis from FFPE tissue remains difficult primarily owing to degradation and chemical modification of nucleic acids during tissue processing and storage. These barriers require individual adaptations of standardized protocols *(4)* and therefore have been incompatible with high-throughput formats. At the level of single nucleotide polymorphisms (SNPs), high-throughput MALDI-TOF genotyping has been shown to be applicable to DNA derived from FFPE tissue by use of commercial DNA extraction kits *(5)*. The rates of successful allele assignment in samples >15 years old range between 70.9% and 99.6%, depending on assay-specific characteristics and extent of degradation of genomic DNA.

Various manual and semiautomated protocols for nucleic acid extraction from FFPE tissue have been developed and are widely available *(6–9)* but no standardized protocol exists to address the automation of the deparaffinization step. Furthermore, most commercially available protocols focus on the extraction of either DNA or RNA (including microRNA) and require 2 independent nucleic acid extraction procedures that employ different kits and work flows. Thus, DNA and RNA isolation from the same tissue has required a minimum of 2 FFPE tissue sections. The introduction of an automated simultaneous extraction method for high-throughput application in routine laboratories would facilitate genomic and transcriptomic investigations and reduce consumption of precious tissues and costs.

In this study we applied a fully automated, xylene-free nucleic acid extraction method to a single FFPE section to investigate its routine applicability for the measurement of relative expression levels and genotyping of genes of interest. We compared both readouts between paired tumor and adjacent normal (peritumoral) breast tissue with the goal of providing simultaneous DNA and RNA fractions for biomarker research.

Material and Methods

FFPE TISSUE SAMPLES

This study included 210 formalin-fixed paraffin-embedded breast cancer and paired noncancer tissue blocks from 105 patients whose disease was diagnosed and tissue samples were archived between 1984 and 2007 at the Pathology Department of the Robert-Bosch Hospital, Stuttgart. The samples had been irreversibly anonymized such that no clinical or patient-related information was available for this study. Hematoxylin-eosin staining was performed for each block to verify the absence of tumor in normal and the presence of at least 20% tumor cells in tumor sections. From each block 10-μm sections were cut on a standard microtome (Leica-Microsystems), placed individually in Sarstedt tubes (P/N 72.692.005, Sarstedt), and stored at 4 °C until extraction.

AUTOMATED DNA AND RNA EXTRACTION

We performed the fully automated method to isolate total nucleic acids from FFPE tissue using iron oxide beads coated with a nanolayer of silica on a modified VERSANT® kPCR Molecular System (Siemens Healthcare Diagnostics) as previously described *(10)*. Briefly, FFPE sections were heat lysed for 30 min at 80 °C followed by 30 min at 65 °C in the presence of proteinase K and detergent. Residual debris was removed from the lysis fluid through unspecific binding to silica-coated iron oxide beads. Beads were subsequently separated on a magnet, and lysates were transferred to a 2-mL deep-well plate. During magnetization, the melted paraffin separated and formed a ring around the tube wall via hydrophobic interactions. Total RNA and DNA were bound to a fresh volume of beads under chaotropic conditions in the deep-well plate. Then, beads were magnetically separated and supernatants were discarded. Surface-bound nucleic acids were washed 3 times and eluted by incubation of the beads with 100 μL of elution buffer for 10 min at 70 °C with shaking. Subsequently, a modified automated pipetting protocol (Fig. 1) was programmed, which allowed splitting of the 100 μL of eluate into 2 aliquots of 50 μL each. One aliquot containing total nucleic acid was separated from the beads and collected into a 96-place rack of 0.75-mL round-bottom tubes (P/N 4274; Thermo Scientific). The second 50 μL was incubated in 2-mL deep-well plates with 12 μL DNase I Mix (1 μL DNase I, RNase free; 5 μL 10× DNase I buffer; and 5 μL water; Ambion/Applied Biosystems) to remove genomic DNA for subsequent mRNA expression analysis. After incubation for 30 min at 37 °C, DNA-free total RNA solution was obtained and collected in the same collection plate as was used for the undigested

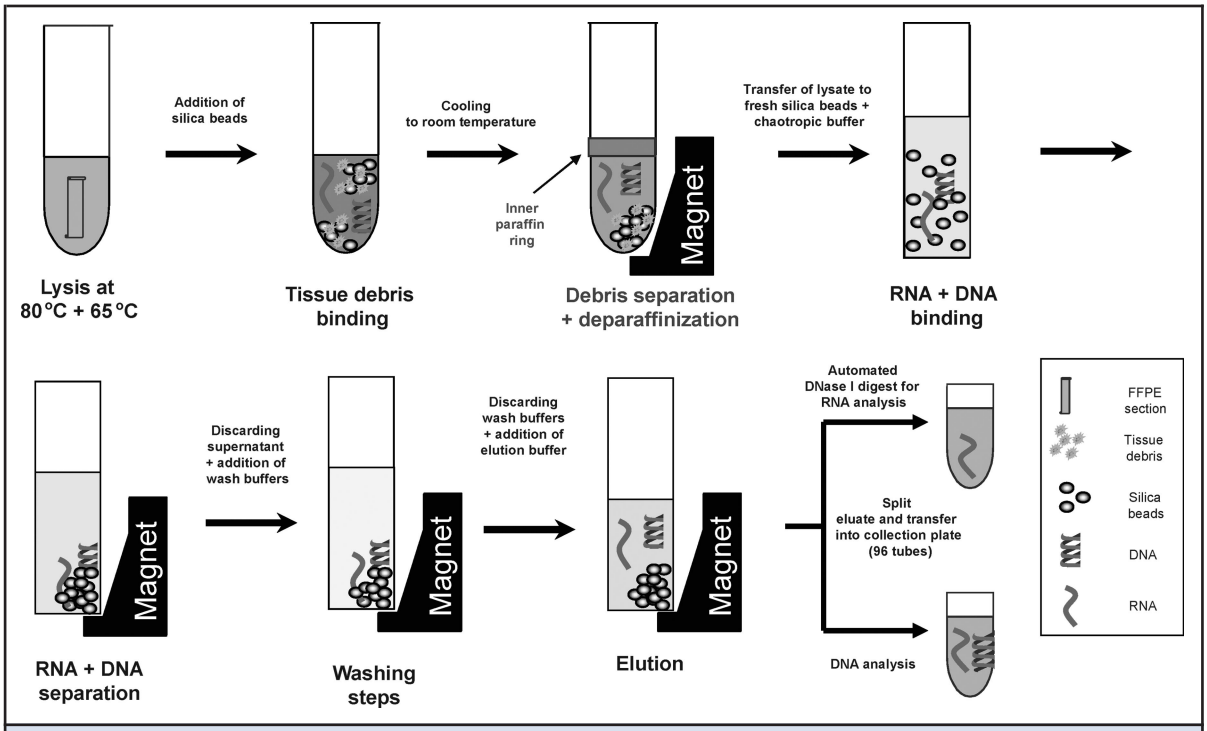

Fig. 1. Schematic work flow of the automated DNA and RNA isolation procedure for FFPE tissue sections.

fraction (second 48 wells) and stored at −80 °C until analyses.

The total time for the extraction of 48 samples, including the hands-on time of 30 min at the beginning and end of the run, was 4 hours and 20 min. Overall, 5 extraction runs were performed with 41 study samples each.

GENE EXPRESSION ANALYSIS BY USING QUANTITATIVE PCR

We assessed expression of estrogen receptor 1 (*ESR1*), progesterone receptor (*PGR*), v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian) (*ERBB2*, alias *HER2*), epoxide hydrolase 1 (*EPHX1*), baculoviral IAP repeat-containing 5 (*BIRC5*), matrix metallopeptidase 7 (*MMP7*), vascular endothelial growth factor A (*VEGFA*), topoisomerase (DNA) II alpha 170kDa (*TOP2A*), and ribosomal protein L37a (*RPL37A*) by quantitative 1-step reverse transcription PCR (RT-qPCR). We used a SuperScript® III Platinum® One-Step Quantitative RT-PCR System with ROX (Invitrogen) according to the manufacturer's instructions. Experiments were performed on an ABI PRISM® 7900HT (Applied Biosystems) with 30 min at 50 °C and 2 min at 95 °C followed by 40 cycles of 15 s at 95 °C and 30 s at 60 °C. Relative expression levels of genes of interest (GOI) were calculated as ΔCq value ($\Delta Cq = 20 - [Cq_{GOI} - Cq_{RPL37A}]$) to yield positively correlated numbers and to facilitate comparisons *(10–12)*.

For the assessment of DNA amounts in the total nucleic acid extract as well as DNA contamination in the respective RNA extract, we performed a progestagen-associated endometrial protein (*PAEP*)-gene–specific qPCR without preceding reverse transcription using the reagents from the SuperScript III Platinum One-Step Quantitative RT-PCR System with ROX and *Taq*DNA polymerase.

The human breast adenocarcinoma cell line MCF-7 total RNA (Ambion/Applied Biosystems) was used as a positive control for RT-qPCR, and human genomic DNA (Roche Diagnostics) as a positive control for qPCR. All PCR assays were performed in triplicate, and the mean of triplicates was reported. Primer and probe sequences are shown in Supplemental Table 1 and RTqPCR amplification efficiencies in Supplemental Table 2 in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol56/issue12.

As an internal method control we tested RNA yields by taking RPL37A Cq values as a surrogate for overall RNA amount for correlation with tissue input. The latter was measured either as tissue area (cm^2) or as

Table 1. Investigated SNPs in tumor and normal breast tissue and the proportion of congruent genotypes.

Gene	Chromosome location	rs No.	Tumors genotyped, n (%)	Normal tissues genotyped, n (%)	Tumor-normal pairs, n	Identical/total chromosomes, n/N, (%)
ESR1	6q25.1	rs9340799	97 (92.3)	101 (96.2)	94	187/188 (99.4)
ESR2	14q23.2	rs4986938	97 (92.3)	100 (95.2)	92	183/184 (99.4)
ESR2	14q23.2	rs944052	101 (96.2)	102 (97.1)	98	194/196 (98.9)
NCOA3	20q12	rs2230782	98 (93.3)	101 (96.2)	95	189/190 (99.5)
NCOA3	20q12	rs2076546	98 (93.3)	101 (96.2)	95	188/190 (98.9)
PPARGC1B	5q32	rs7732671	102 (97.1)	103 (98.1)	100	198/200 (99.0)
RRAS2	11p15.2	rs11023197	96 (91.4)	99 (94.3)	91	181/182 (99.4)

cell count per defined area, scored as 1–10 within the range of <50 cells (score 1) to >4000 cells (score 10).

SNP ANALYSIS

Seven SNP loci from 5 genes, *ESR1*, estrogen receptor 2 (*ESR2*), nuclear receptor coactivator 3 (*NCOA3*), peroxisome proliferator-activated receptor gamma, coactivator 1 beta (*PPARGC1B*), and related RAS viral (r-ras) oncogene homolog 2 (*RRAS2*, alias *TC21*) with different minor allele frequencies were genotyped (Table 1 and Table 2) by using MALDI-TOF mass spectrometry (MS) according to manufacturer's instructions (Sequenom MassARRAY platform). Briefly, specific primers were used to amplify short amplicons (80–120 nucleotides) spanning the SNP of interest from 2 μL of the 50-μL DNase I–undigested nucleic acid aliquot. After removal of unincorporated nucleotides with shrimp alkaline phosphatase, we used a MassExtend reaction to generate short oligonucleotide stretches with different masses according to the allelic state of the SNP. Data acquisition from the oligonucleotides spotted on SpectroCHIP was performed with a Sequenom Compact MALDI-TOF MS, and genotyping calls were made with MASS ARRAY RT software V 3.0.0.4 (Sequenom). PCR primers and primers used for homogenous MassExtend reactions are available on request.

STATISTICAL EVALUATION

For statistical analyses we used the software Prism® 4 (GraphPad Software) and Microsoft Excel 2003 (Microsoft). Medians, Pearson correlation coefficients, and *P* values from paired *t*-tests were calculated. We used the Mann–Whitney test for comparisons of ΔCq values between normal and tumor tissues. *RPL37A* Cq values were analyzed in relation to tissue input by using linear curve fitting and Spearman correlation estimates. SNP allele frequencies were tested for deviation from Hardy-Weinberg equilibrium (HWE) in normal and tumor tissues separately, and for allele frequency differences to a reference breast cancer patient cohort genotype sample (n = 920) by using χ^2 tests. All *P* values were 2-sided and considered significant when <0.05.

Table 2. MAF of investigated SNPs, test of HWE, and MAFs from a reference cohort.

SNP	Tumor MAF	HWE, *P*	Normal tissue MAF	HWE, *P*	MAF reference[a]	Differences paraffin vs reference[b]
rs9340799	0.289	0.58	0.312	0.69	0.349	NS
rs4986938	0.397	0.95	0.405	0.98	0.369	NS
rs944052	0.139	0.67	0.137	0.99	0.154	NS
rs2230782	0.107	0.99	0.104	0.99	0.099	NS
rs2076546	0.071	0.74	0.069	0.77	0.098	NS
rs7732671	0.059	0.51	0.063	0.61	0.073	NS
rs11023197	0.328	0.74	0.333	0.99	0.379	NS

[a] Reference MAF refers to a cohort of breast cancer individuals (N = 920) genotyped from blood DNA (H. Brauch, P. Fritz, M. Eichelbaum, M. Schwab, and W. Schroth, unpublished data).
[b] Two-sided χ^2 test for differences in genotype frequencies between either tumor or normal paraffin tissue and the reference frequencies derived from blood DNA. NS indicates not significant.

A

Measured genes on RNA eluate (RT-qPCR)	Numbers of adjacent normal tissues	Numbers of tumor tissues
RPL37A	105 (100%)	105 (100%)
ESR1	105 (100%)	104 (99%)
PGR	99 (94%)	102 (97%)
ERBB2	104 (99%)	105 (100%)
EPHX1	104 (99%)	102 (97%)
BIRC5	22 (21%)	85 (81%)
MMP7	76 (72%)	101 (96%)
VEGFa	105 (100%)	105 (100%)
TOP2a	89 (85%)	100 (95%)
Detected gene on DNA eluate (PCR)		
PAEP	104 (99%)	104 (99%)

B

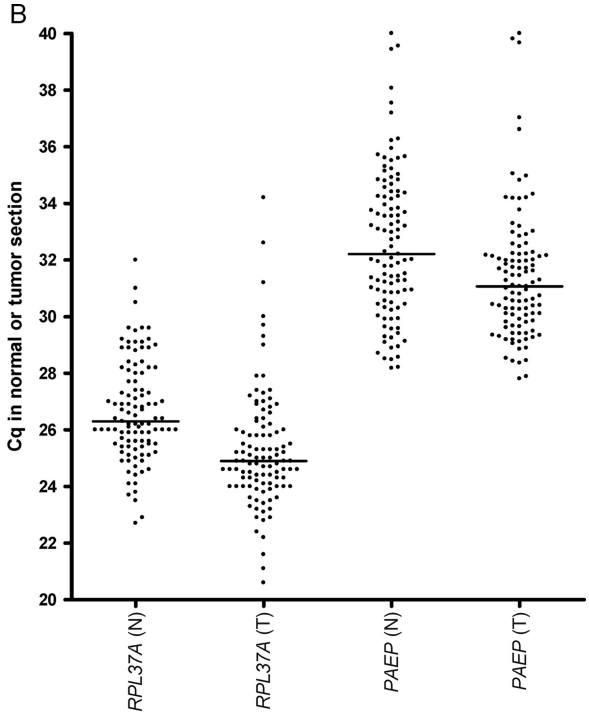

Fig. 2. Amplified genes and yields.

(A), Absolute and relative number (%) of measured RT-qPCR results for 9 genes obtained from DNase I–treated eluate and qPCR results for the *PAEP* gene without prior reverse transcription from paired tumor (T) and adjacent normal (N) tissue block sections. (B), Plots of Cq values for the normalization gene *RPL37A* and the *PAEP* gene used as surrogate markers for mRNA or DNA yield, respectively. Cq values are shown for adjacent normal (N) or paired tumor (T) FFPE sections. Black bars indicate median Cq values of 26.3; 24.9; 32.2, and 31.1, respectively.

Results

RNA AND DNA EXTRACTION PERFORMANCE

A total of 210 independent 10-μm FFPE sections from paired tumor and adjacent normal tissue were processed by using a fully automated extraction procedure without conventional xylene deparaffinization before tissue lysis (Fig. 1). The storage time of the FFPE blocks before section preparation, extraction, and analysis ranged between 1 and 25 years. From each section an eluate for RNA analysis, i.e., free of DNA after DNase I digestion, and a second eluate for DNA analysis containing total nucleic acid were generated. RNA and DNA extraction performance were controlled by a detectable RT-qPCR signal for the normalization gene *RPL37A* (RNA) or qPCR for the *PAEP* gene (DNA) as surrogate markers for nucleic acid yield and integrity. Amplifiable RNA and DNA were detected in 208 of 210 samples. Two sections showed no DNA signal [quantification cycle (Cq) = 40], yet contained detectable RNA (Fig. 2A). In Fig. 2B, distributions of Cq values for *RPL37A* and *PAEP* are plotted for the tumor and adjacent normal sample groups. The median RNA yield was significantly higher in the tumor than in the adjacent normal samples, with Cq of 24.9 vs 26.3 ($P < 0.0001$). The same pattern was observed for the median DNA yield, which showed significantly lower Cq values (31.07) for tumors compared to normal tissue (32.21; $P = 0.0003$). We confirmed higher RNA yields in tumors relative to normal sections by measuring total RNA using the Ribogreen method (Invitrogen, data not shown) for the same eluates.

For internal validation of the extraction method, we tested for a relationship between RNA concentrations, taking *RPL37A* Cq values as a surrogate, and tissue input. There was no distinct relationship with the size of the tissue area ($r = -0.13$; $P = 0.07$); however, the number of input cells was correlated with resulting RNA levels ($r = -0.56$; $P < 0.001$; Fig. 3).

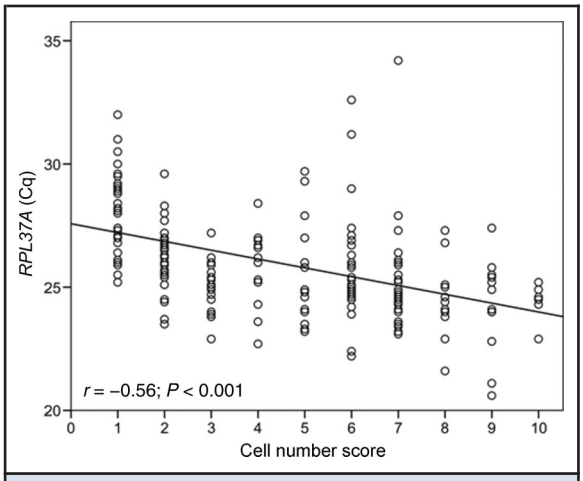

Fig. 3. Relationship between tissue cell count and RNA yield per tissue section.

Cell numbers per defined tissue area were scored as 1–10, ranging between <50 cells [Harris et al. *(1)*] to >4000 cells [Bohmann et al. *(10)*]. Expression of the housekeeping gene *RPL37A* (RT-PCR Cq value) was used as a surrogate for overall RNA content per tissue section. The plot shows an inverse correlation between Cq values and cell numbers within a tissue section (Spearman $r = -0.56$), suggesting an unbiased tissue processing of the method with respect to RNA yield and downstream applications.

NORMALIZED EXPRESSION OF 8 BREAST CANCER–RELATED GENES IN PAIRED TUMOR AND ADJACENT NORMAL FFPE SECTIONS

An automated DNase I digestion step was performed for half of the total nucleic acid eluates. Complete DNase I digestion was successful in 195 of 210 RNA eluates (93%) resulting in a PAEP Cq of 40, i.e., with no detectable genomic DNA left. A very weak but yet detectable PAEP signal between 37 and 39 Cq was observed in the remaining 15 eluates which did not affect subsequent mRNA detection analyses based on RT-qPCR with bispecific primer/probe sets. For the latter, we analyzed the mRNA expression of 8 breast cancer–related genes using RT-qPCR and normalization to *RPL37A*. The genes were chosen to represent different biological pathways relevant for breast cancer biology, including hormone receptor status (*ESR1*, *PGR*), HER2 (human epidermal growth factor receptor 2) status for targeted antibody therapy (*ERBB2*), tamoxifen metabolism/detoxification (*EPHX1*), *WNT* signaling pathway (*MMP7*), angiogenesis (*VEGFA*), and proliferation (*BIRC5, TOP2A*).

The detection rates (Cq <40) for each gene ranging from 21% (*BIRC5*) to 100% (*ESR1* and *VEGFA*) in normal breast tissue and from 81% (*BIRC5*) to 100% (*ERBB2* and *VEGFA*) in tumor tissue sections are summarized in Fig. 2A. For 5 of 8 genes (*ESR1*, *PGR*, *ERBB2*, *BIRC5*, and *TOP2A*) the median expression was at least 2-fold higher, i.e., 1 Cq cycle difference in the tumor compared to the corresponding normal tissue ($P < 0.0002$; Fig. 4). The maximum difference in gene expression was observed for the invasion marker *BIRC5*, which was hardly detectable in normal tissue and was found at an 8-fold increased tumor vs normal tissue expression ($P < 0.0001$; ΔCq difference of 3.1). Two genes, *EPHX1* and *VEGFA*, showed a similar median expression level (12.3 vs 12.4 and 12.3 vs 13.0, respectively; both $P = 0.99$) between both tissue types. The correlation for mRNA expression per gene between paired tumor and adjacent normal sections, as indicated by Pearson correlation coefficient calculations, was very low, ranging between $r = -0.11$ and $r = 0.33$, which suggested that these genes are differently expressed in tumor and neighboring normal tissue. In addition, we did not observe a significant correlation between the median quantitative RNA expression of each target gene and the storage time of blocks (data not shown).

MALDI-TOF SNP GENOTYPING IN FFPE SECTIONS FROM PAIRED TUMOR AND ADJACENT NORMAL FFPE SECTIONS

Five genes related to breast cancer biology were selected to represent estrogen receptor signaling (*ESR1, ESR2*), estrogen receptor activation (*NCOA3, PPARGC1B*), or RAS/MAPK signal transduction (*RRAS2*). Genotyping call rates for the 7 investigated SNP loci ranged between 91.4% and 97.1% in tumor, and between 94.3% and 98.1% in normal tissue (Table 1). There was a high number of concordant allele calls between paired tumor/normal tissue sections ranging from at least 98.9% to a maximum of 99.5% (Table 1). Allele frequencies were low [minor allele frequency (MAF) = 0.06, *PPARGC1B* rs7732671], moderate (MAF = 0.14, *ESR2* rs944052), or abundant (MAF = 0.4, *ESR2* rs4986938) and did not show deviation from HWE irrespective of tissue type (Table 2). Genotype and allele frequencies did not differ between tumor and normal tissue sections and no frequency differences were observed compared to a reference sample based on peripheral blood DNA genotyping (Table 2). With respect to drop-out rates in genotyping and rate of discordant allele calls among tissue types, we did not observe a significant correlation with the storage time of blocks (data not shown). There was a slightly increased allele discordance rate within very old specimens (collected on or before 1986); however, only 2 of 10 tissue pairs in this age class accounted for the majority of discordant allele calls for the 7-SNP panel.

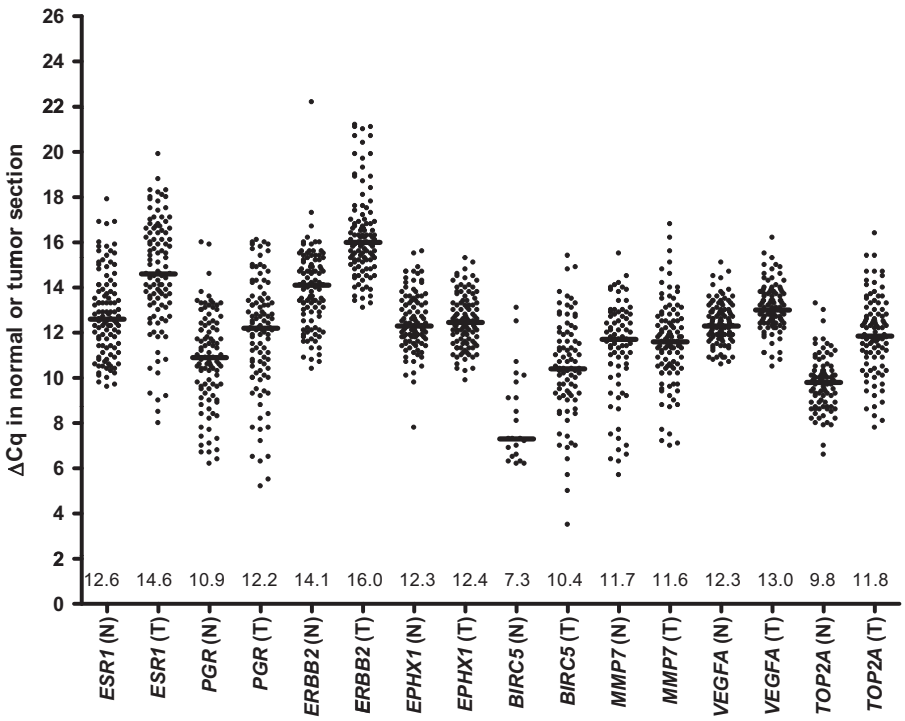

Fig. 4. Overview of quantitative mRNA expression.

RPL37A-normalized quantitative expression values (DCq) of *ESR1, PGR, HER2, EPHX1, BIRC5, MMP7, VEGFA,* and *TOP2A* in FFPE sections from paired tumor (T) and adjacent normal (N) tissue block sections. Median Cq values are indicated by black bars and as numbers on top of the *x* axis.

Discussion

Here we describe a novel xylene-free, fully automated nucleic acid extraction method that allows the simultaneous isolation of RNA and DNA from a single FFPE breast cancer tissue section. We successfully applied downstream real-time expression analyses of mRNA and MALDI-TOF MS genotyping for DNA and demonstrated that both a standard expression measurement method like RT-qPCR as well as a high-throughput genotyping technology can be successfully applied to crude nucleic acid eluates. Moreover, molecular analytical technologies like sequencing (data not shown) and microarray analysis *(11)* also have been successfully performed. The protocol thus provides an efficient and economical dual analysis of genes of interest (e.g., molecular biomarkers), allowing not only the conservation of limited pathological source material but also a significant reduction in hands-on time and laboratory costs per sample processing. These issues are of particular interest within the context of genetic variation and gene expression research within clinical studies.

Although various methods for the isolation of nucleic acids from FFPE samples are commercially available and widely used, a number of limitations may be inherent to all of them, including total processing time, involvement of toxic chemicals such as xylene for manual deparaffinization, and high tissue consumption. Although these issues relate to the nucleic extraction process, the nucleic acid degradation due to tissue fixation in FFPE blocks presents an additional inevitable difficulty. To account for this nucleic acid degradation we have chosen quantitative Taqman (RT-) PCR and MALDI-TOF MS methods targeting small fragments of nucleic acids of interest for downstream applications.

The automated extraction protocol presented here demonstrated high success rates regarding the number of analyzable specimens for both the normalization and control genes used for RNA (100%) and DNA (99%), respectively. We demonstrated that overall RNA yields depended on the number of cells subjected to lysis, rather than on the tissue area (size). This observation suggests that the extraction performance is not influenced by the excess fat that frequently pre-

vents the isolation of sufficient nucleic acid amounts from breast tissue. There was no observed effect of age/storage time on amplification signals for either DNA or mRNA in older paraffin material, which indicates that the extraction procedure allows for unbiased analyses of patient materials up to 25 years old. Thus it is possible to cover various types of retrospective marker studies, including studies of prognostic or therapy-predictive biomarkers in tumor sections.

We chose 8 different genes for expression analysis to encompass various biological disease pathways implicated in breast cancer. The finding of an increased expression level in tumors relative to adjacent normal tissues for the majority of targets is in line with their potential function as mitogens (*ESR1, PGR*, and *ERBB2*) or dysregulated oncogenes (*BIRC5, VEGFA*, and *TOP2A*) in breast cancer. Interestingly, the correlation of mRNA expression of these genes in paired tumor and corresponding normal tissue sections was very low (−0.11 to 0.33), suggesting that these genes underwent a shift in expression regulation during tumorigenesis.

Retrospective research based on FFPE tissue, such as pharmacogenetic association studies and the identification of prognostic DNA signatures based on genetic variability, require the use of reliable, high-throughput genotyping methods to investigate large numbers of patient samples. The genotypes we identified by MALDI-TOF MS clearly indicate that FFPE DNA isolated with the automated DNA/RNA isolation platform is well suited for this analytical method, with previously proven accurate allelic discrimination of 99.9% from blood DNA *(13)*. The sample dropout rate in this investigation was <10% across all assays, which indicates that the automated extraction method provides genomic DNA of a quality comparable to that in similar studies in which archived FFPE tissue samples were used with manual DNA extraction procedures *(5)*. We selected 5 genes known to harbor germline polymorphisms that have been reported to be related to breast cancer risk or disease progression *(14, 15)*, to functionally influence estrogen-receptor activity (*NCOA3*), or to be predictive for breast cancer tamoxifen outcome *(16)*. Although our data showed very high concordance rates of genotypes between tumor and corresponding normal tissue, a few samples (1.1%) showed discordant alleles between paired samples. Cancer cells usually display a wide range of genetic alterations, including mutations, DNA gene copy number variations caused by chromosomal amplifications, and deletions, which can also be observed in premalignant cancer-surrounding breast tissue *(17)*. Moreover, breast cancer is genetically heterogenous, showing clonal divergences that potentially result in discordant loss of heterozygosity among different areas during tumor progression *(18)*. Although the tumor genome may therefore differ from the corresponding host genome, the discrepancies between tumor and normal tissue pairs we observed were consistent across all 7 SNPs, which comprised 5 different chromosomal regions. With the exception of the *ESR1* gene, none of the remaining loci have been listed as hot spots of genomic alterations, including those reported by Li et al. *(17)* for the comparison between breast tumor and normal adjacent tissue. Because allele frequencies of the investigated SNPs were in HWE and did not differ from a reference blood cohort (no blood was available from the patients under study), the possibility of tumor-specific discordant loss of heterozygosities and allele dropouts can be excluded. We suggest that the small fraction of differing genotypes may rather be attributable to an intrinsic failure rate due to differential DNA degradation in FFPE material, or other yet unknown factors. In addition, technical reasons may account for genotyping incongruencies between tumor and corresponding normal tissue, especially because such incongruencies were predominantly observed for larger amplicon sizes, e.g., more than 150 nucleotides and/or higher-plex levels in the MALDI-TOF MS assay design (data not shown). Overall, our data for SNP genotyping by mass spectrometry demonstrated that tumor biopsies processed with the automated RNA/DNA extraction method can be used as suitable source material for the assessment of the underlying host genome, with an accuracy ≥98.9%.

Currently, extensive biomarker research programs are in operation and depend heavily on archived FFPE tissues linked to clinical follow-up. Our projection for the future is that one focus of investigation will be validated RNA/DNA algorithms that incorporate genetic variation of the host combined with dysregulated gene expression pathways to improve predictive power compared with algorithms that use single-marker information. As an example, Goetz et al. *(19)* demonstrated that compared to single-marker analyses, determination of cytochrome P450, family 2, subfamily D, polypeptide 6 (*CYP2D6*) genotype and homeobox B13 (*HOXB13*)/interleukin 17 receptor B (*IL17RB*) gene-expression ratio improved the prediction of tamoxifen resistance in breast cancer patients. Translation of these algorithms into clinical practice is fundamentally based on method accuracy. For routine clinical implementation of disease-relevant molecular profiles in large-scale retrospective biomarker studies, a standardized and automated nuclear acid extraction and detection platform for the simultaneous handling of DNA and RNA is a prerequisite. We therefore consider the complete automation of simultaneous RNA and DNA extraction from a single tissue section to be a major step toward linking sample preparation with assay technology. This improvement will simplify the development

of combined DNA and RNA biomarker assay platforms, including real-time PCR cycler instrumentation and fully automated validated work flows for the clinical research environment.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures or Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: M. Gehrmann, Siemens Healthcare Diagnostics Products; U. Stropp, Siemens Healthcare Diagnostics.

Consultant or Advisory Role: None declared.
Stock Ownership: G. Hennig, Siemens; U. Stropp, Siemens Healthcare Diagnostics.
Honoraria: H. Brauch, AstraZeneca; M. Eichelbaum, Roche Molecular Diagnostics; M. Schwab, Roche Molecular Diagnostics.
Research Funding: M. Schwab, Bundesministerium fuer Bildung und Forschung (O1ZP0502, 03IS2061C, 0315755). H. Brauch and M. Schwab, Roche Molecular Diagnostics and Siemens Healthcare Diagnostics Products; H. Brauch, M. Schwab, W. Schroth, M. Eichelbaum, and P. Fritz, Robert Bosch Foundation, Stuttgart.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Acknowledgments: We thank T. Acht, R. Koehler, P. Maaß, and A. Frehn for excellent technical assistance; K. Weber for advice about statistical analyses; and S. Schwers for critical review of the manuscript.

References

1. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 2007;25:5287–312.
2. Wang HL, Lopategui J, Amin MB, Patterson S. KRAS mutation testing in human cancers: the pathologist's role in the era of personalized medicine. Adv Anat Pathol 2010;17:23–32.
3. Mizuno T, Nagamura H, Iwamoto KS, Ito T, Fukuhara T, Tokunaga M, et al. RNA from decades-old archival tissue blocks for retrospective studies. Diagn Mol Pathol 1998;7:202–8.
4. 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;27:4436–43.
5. Jaremko M, Justenhoven C, Abraham BK, Schroth W, Fritz P, Brod S, et al. MALDI-TOF MS and TaqMan assisted SNP genotyping of DNA isolated from formalin-fixed and paraffin-embedded tissues (FFPET). Hum Mutat 2005;25:232–8.
6. 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;8:118–28.
7. Specht K, Richter T, Müller U, Walch A, Werner M, Höfler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. Am J Pathol 2001;158:419–29.
8. Rupp GM, Locker J. Purification and analysis of RNA from paraffin-embedded tissues. Biotechniques 1988;6:56–60.
9. Stanta G, Schneider C. RNA extracted from paraffin-embedded human tissues is amenable to analysis by PCR amplification. Biotechniques 1991;11:304–8.
10. Bohmann K, Hennig G, Rogel U, Poremba C, Mueller BM, Fritz P, et al. RNA extraction from archival formalin-fixed paraffin-embedded tissue: a comparison of manual, semiautomated, and fully automated purification methods. Clin Chem 2009;55:1719–27.
11. Schwers S, Reifenberger E, Gehrmann M, Izmailov A, Bohmann K. A high-sensitivity, medium-density, and target amplification-free planar waveguide microarray system for gene expression analysis of formalin-fixed and paraffin-embedded tissue. Clin Chem 2009;55:1995–2003.
12. Mueller BM, Kronenwett R, Hennig G, Euting H, Weber K, Bohmann K, et al. Quantitative determination of estrogen receptor, progesterone receptor, and HER2 mRNA in formalin-fixed paraffin-embedded tissue—a new option for predictive biomarker assessment in breast cancer. Diagn Mol Path. Forthcoming.
13. Oeth P, Beaulieu M, Park C, Kosman, D, del Mistro G, van den Boom H, Jurinke C. iPLEX™ assay: increased plexing efficiency and flexibility for MassARRAY® system through single base primer extension with mass-modified terminators. 2007 Sep 27. Sequenom application note. http://www.sequenom.com/getdoc/52ad1d40-ff33-421f-a26f-58ca91b86364/iPLEX_Application_Note (Accessed November 2010).
14. Tan SH, Lee SC, Goh BC, Wong J. Pharmacogenetics in breast cancer therapy. Clin Cancer Res 2008;14:8027–41.
15. Wirtenberger M, Tchatchou S, Hemminki K, Schmutzhard J, Sutter C, Schmutzler RK, et al. Associations of genetic variants in the estrogen receptor coactivators PPAGC1A, PPARGC1B and EP300 with familial breast cancer. Carcinogenesis 2006;27:2201–8.
16. Rokavec M, Schroth W, Amaral SM, Fritz P, Antoniadou L, Glavac D, et al. A polymrphism in the TC21 promoter associates with an unfavorable tamoxifen treatment outcome in breast cancer. Cancer Res 2008;68:9799–808.
17. Li J, Wang K, Li S, Timmermans-Wielenga V, Rank F, Wiuf C, et al. DNA copy number aberrations in breast cancer by array comparative genomic hybridization. Genomics Proteomics Bioinformatics 2009;7:13–24.
18. Lichy JH, Dalbegue F, Zavar M, Washington C, Tsai MM, Sheng ZM, Taubenberger JK. Genetic heterogeneity in ductal carcinoma of the breast. Lab Invest 2000;80:291–301.
19. Goetz MP, Suman VJ, Couch FJ, Ames MM, Rae JM, Erlander MG, et al. Cytochrome P450 2D6 and homeobox 13/interleukin-17B receptor: combining inherited and tumor gene markers for prediction of tamoxifen resistance. Clin Cancer Res 2008;14:5864–8.