Comparing Platforms for Messenger RNA Expression Profiling of Archival Formalin-Fixed, Paraffin-Embedded Tissues

Svitlana Tyekucheva, Neil E. Martin, Edward C. Stack, Wei Wei, Vinod Vathipadiekal, Levi Waldron, Michelangelo Fiorentino, Rosina T. Lis, Meir J. Stampfer, Massimo Loda, Giovanni Parmigiani, Lorelei A. Mucci, and Michael Birrer

Archival formalin-fixed, paraffin-embedded (FFPE) tissue specimens represent a readily available but largely untapped resource for gene expression profiling—based biomarker discovery. Several technologies have been proposed to cope with the bias from RNA cross-linking and degradation associated with archival specimens to generate data comparable with RNA from fresh-frozen materials. Direct comparison studies of these RNA expression platforms remain rare. We compared two commercially available platforms for RNA expression profiling of archival FFPE specimens from clinical studies of prostate and ovarian cancer: the Affymetrix Human Gene 1.0ST Array following whole-transcriptome amplification using the NuGen WT-Ovation FFPE System V2, and the NanoString nCounter without amplification. For each assay, we profiled 7 prostate and 11 ovarian cancer specimens, with a block age of 4 to 21 years. Both platforms produced gene expression profiles with high sensitivity and reproducibility through technical repeats from FFPE materials. Sensitivity and reproducibility remained high across block age within each cohort. A strong concordance was shown for the transcript expression values for genes detected by both platforms. We showed the biological validity of specific gene signatures generated by both platforms for both cohorts. Our study supports the feasibility of gene expression profiling and large-scale signature validation on archival FFPE prostate and ovarian tumor specimens using commercial platforms. These approaches have the potential to aid precision medicine with biomarker discovery and validation. (J Mol Diagn 2015, 1–8; http://dx.doi.org/10.1016/j.jmoldx.2015.02.002)

Q1 From the Departments of Biostatistics and Computational Biology, Radiation Oncology, and Medical Oncology, and the Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, Massachusetts; the Departments of Biostatistics, Epidemiology, and Nutrition, Harvard School of Public Health, Boston, Massachusetts; the Department of Pathology, Brigham and Women’s Hospital, Boston, Massachusetts; the Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts; the City University of New York School of Public Health, Hunter College, New York, New York; the Pathology Unit, Addarit Institute, S. Orsola-Malpighi Hospital, Bologna, Italy; the Channing Division of Network Medicine, Department of Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts; the Broad Institute; Cambridge, Massachusetts; and the Division of Cancer Studies, King’s College London, London, United Kingdom.

Q2 Accepted for publication February 24, 2015.

Address correspondence to Svitlana Tyekucheva, Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, 450 Brookline Ave., Boston, MA 02115. E-mail: svitlana@jimmy.harvard.edu.

Q4 With the advent of precision medicine, there is a growing interest in developing prognostic and predictive molecular signatures to help guide the care of patients with cancer. Among approaches taken to realize the goal of identifying the ideal treatment for the patient at the correct time are mutation analysis, immunohistochemical staining, and mRNA expression analysis. Such studies require access to tissue specimens with detailed clinical annotation and often long-term follow-up.
evaluation. These clinical samples, including biopsy and surgical specimens, typically are formalin-fixed and embedded in paraffin to allow for morphologic assessment by a pathologist. Although this process preserves morphologic features of the tumors, it also makes RNA expression profiling more challenging relative to fresh-frozen tissue. RNA in formalin-fixed, paraffin-embedded (FFPE) samples is subject to degradation, fragmentation, and cross-linking, which typically limits library preparation and gene expression assessment. Importantly, traditional quality control measurements for RNA, such as RNA integrity number, are not necessarily predictive of the success of corresponding gene expression assays.

The large numbers of well-annotated FFPE tumor tissue samples currently archived remain a vast and underused resource in the genomic study of cancer. Notably, most large clinical and epidemiologic cohorts only collect FFPE samples. Given this wealth of archival material from patients with known outcomes and the continued FFPE processing of new clinical specimens, there is a need to develop and test reliable methods for profiling mRNA expression in FFPE materials.

Several platforms have been developed in recent years to assess mRNA expression from FFPE tissue including whole-transcriptome amplification and direct assessment using multiplexed color-coded probes. As an initial step to developing prognostic and predictive mRNA signatures from archival tumor specimens, we performed head-to-head comparisons of gene expression profiles from prostate and ovarian cancer FFPE specimens from large-scale epidemiologic studies and clinical trials representative of a wide variety of fixation times, block ages, and block storage conditions using two platforms: the NuGen WT-Ovation FFPE System (NuGen, NuGen, Inc., San Carlos, CA) + Affymetrix (Affymetrix, Santa Clara, CA) and the NanoString nCounter Cancer panel (NanoString, NanoString Technologies, Seattle, WA).

Materials and Methods

Prostate Tissue Samples

Archival FFPE radical prostatectomy specimens were collected from treating institutions for men diagnosed with prostate cancer who had enrolled in the prospective Physicians’ Health Study. Participants provided informed consent to collect biospecimens, and the study was approved by the Institutional Review Boards at the Harvard School of Public Health and Partners Health Care. Pathology was reviewed centrally to confirm the cancer diagnosis and to provide consistent histopathologic review including Gleason scores. Areas of high-density tumor were identified, and two to three 0.6-mm punches were taken from both tumor and adjacent nontumor prostatic tissue for RNA extraction. The cores were deparaffinized using 800 μL Citrisolv (Fisher Scientific, Pittsburgh, PA) at 60°C for 20 minutes followed by 1.2 mL Citrisolv:absolute alcohol (2:1) at room temperature for 10 minutes. Cores subsequently were washed with absolute alcohol, dried at 55°C, and incubated overnight at 45°C in 300 μL lysis buffer [10 mmol/L NaCl, 500 mmol/L Tris (pH 7.6), 20 mmol/L EDTA, 1% sodium dodecyl sulfate] containing 1 mg/mL proteinase K (Ambion, Austin, TX). RNA was extracted using the RecoverAll Total Nucleic Acid Isolation kit (Ambion). After tissue digestion, following the manufacturer’s protocol, samples were incubated in 10× DNase (Ambion) and column-purified to elute RNA. The concentration was determined using the Nanodrop 1000 (Fisher Scientific) and RiboGreen RNA Assay Kit (Ambion). Samples and technical replicates were randomized and assigned study identifications to blind laboratory personnel.

Ovarian Tissue Samples

Ovarian cancer specimens were obtained from the Gynecologic Oncology Group tissue bank (Columbus, OH). All specimens were archival FFPE tissue samples obtained from women with advanced-stage epithelial ovarian cancer in clinical trial GOG218, a randomized phase III trial testing the impact of the addition of bevacizumab to standard chemotherapy for the up-front treatment of advanced-stage epithelial ovarian cancer. The specimens were obtained at the primary debulking surgery.

The pathology was confirmed by the central Gynecologic Oncology Group review to ensure the correct histology and percentage of tumor. Ten-micron sections were made on positively charged slides. RNA was extracted from paraffin scrolls using the RNaseasy FFPE kit (Qiagen) with modifications. In brief, 1 mL of xylene was added to a 25-mm paraffin scroll, vortexed vigorously for 10 seconds, and centrifuged at full speed for 2 minutes. Supernatant then was removed and 1 mL of ethanol (100%) was added to the pellet, mixed by vortexing, and centrifuged at full speed for 2 minutes. Supernatant was removed and the tube was incubated at room temperature for 10 minutes with the lid opened. RNA subsequently was extracted using reagents provided in the RNaseasy FFPE extraction kit according to the manufacturer’s instructions. Finally, purified RNA was eluted from the RNasey MinElute spin column using 30 mL RNase-free water.

Experimental Design

We profiled seven paired tumor and adjacent normal prostate tissue samples from three patients with a Gleason score of 8, one patient with a Gleason score of 7, and three patients with a Gleason score of 6 disease on the NuGen + Affymetrix platform, and a subset of five pairs (two with a Gleason score of 6 and three with a Gleason score of 8) on NanoString. Block ages for prostate cancer specimens ranged from 11 to 21 years. We profiled two to three technical replicates for each sample with varying RNA input amounts. For ovarian cancer we selected five serous
carcinoma and six clear cell carcinoma samples, and block ages for this cohort ranged from 4 to 7 years. All of these samples were profiled on both NuGen + Affymetrix and NanoString, with either two or three technical replicates. Complete information on all samples and technical replicates that were analyzed are presented in Supplemental Table S1.

**Messenger RNA Expression Profiling**

**Array**

For array-based mRNA profiling, we first performed whole-transcriptome amplification using the WT-Ovation FFPE System V2 (NuGen). This approach initiates amplification at the 3' end as well as randomly throughout the transcriptome, improving the performance in severely degraded FFPE samples. After isothermal amplification, 50 or 100 ng of total RNA was amplified to 4 to 7 µg of biotinylated cDNA complementary to the original mRNA. The amplification step has been optimized for RNA extracted from FFPE specimens15,16 and has shown comparable differential expression profiles to corresponding fresh-frozen tissues.17 After amplification, we hybridized 3.75 µg of amplified cDNA to the GeneChip Human Gene 1.0 ST Array (Affymetrix). The 1.0 ST Array profiles expression of >28,000 genes with an average of 26 probes per gene.

**NanoString**

We used the NanoString nCounter platform18 to capture and count 230 cancer-related human genes using a prebuilt kit supplied by the manufacturer. Following the manufacturer’s instructions, we aliquoted 100 to 200 ng total RNA in 5 µL to initiate analysis.

**Data Preprocessing**

We normalized the Affymetrix data across samples and batches using RMA.19,20 For the NanoString platform the data were processed according to the manufacturer’s recommendations. Briefly, background subtracted counts were multiplied by scaling factors proportional to the sum of counts for spiked-in positive control probes to account for individual assay efficiency variation, and to the geometric average of the housekeeping gene probes (CLTC, GAPDH, GUSB, HPRT1, PGK1, and TUBB) to account for variability in the mRNA content. Background signal was calculated as a median value of the negative hybridization control probes. Normalized counts were log-transformed for downstream analysis.

Raw and preprocessed data were submitted to the GEO database [http://www.ncbi.nlm.nih.gov/geo; Accession number GSE54809].

Data preprocessing and statistical analysis were performed using statistical software packages R21 and Bioconductor.22 Power analysis was performed using G∗Power version 3.1.3.23

---

**Table 1**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Percent present</th>
<th>Correlation with block age</th>
<th>Correlations between technical replicates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ovarian samples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NuGen + Affymetrix</td>
<td>0.17–0.58</td>
<td>0.02</td>
<td>0.91–0.98</td>
</tr>
<tr>
<td>NanoString</td>
<td>0.40–0.79</td>
<td>-0.11</td>
<td>0.88–0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prostate samples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NuGen + Affymetrix</td>
<td>0.40–0.68</td>
<td>0.06</td>
<td>0.94–0.98</td>
</tr>
<tr>
<td>NanoString</td>
<td>0.83–0.93</td>
<td>-0.04</td>
<td>0.95–0.99</td>
</tr>
</tbody>
</table>

**Results**

We sought to evaluate several potential variables related to the performance of RNA expression assays. We assessed the sensitivity and reproducibility of the gene expression measurements across technical replicates of RNA extracted from the same tissue with respect to the following: i) RNA input levels, ii) FFPE block age, and iii) tumor grade and histology. We also used tumor grade and histology to assess the biological validity of the expression profiles.

**Sensitivity and Reproducibility Analysis**

For each platform and each cohort we calculated the percentages of probes detectable above the background. In each platform, background probes comprise sequences not found in the human genome. We observed no significant association between the proportion of detectable probes and the block age, as quantified by the Pearson correlation coefficients (Table 1). The wider range in the percentage of probes present for the NuGen + Affymetrix data on the prostate cohort likely can be explained by differences in the instrument calibrations between the two batches in which the samples were assayed. One batch had higher intensities, which were corrected analytically by RMA normalization (Supplemental Figure S1). With sample sizes of 24 (Ovarian and prostate NanoString studies, and Ovarian NuGen + Affymetrix) and 30 (Prostate NuGen + Affymetrix), we had a power of 0.86 and 0.93, respectively, to detect whether 30% or more of the variance in the percentage of the probes present was explained by the block age. Therefore, failure to find a significant correlation between the percentage of the probes present and block age was suggestive of the fact that block age indeed does not explain much of the variation in the percentage of the probes within each study.

In the prostate samples with sufficient replicates at each RNA quantity we did not observe a statistically significant difference between the percentage of probes present at 50 versus 100 ng of input RNA for the NuGen + Affymetrix platform (analysis of variance power of 0.85 to detect whether
40% of variance in percentage of the present probes is explained by RNA quantity. A lower percentage of probes in the samples with smaller RNA input quantity was present in the NanoString data (Tukey HSD for 200 ng versus 100 ng, $P = 0.04$). For this analysis we had an analysis of variance power of 0.85 to detect whether 45% of variance in the percentage of probes present above the background is explained by RNA quantity. For ovarian samples only two replicates were of lower quantity, therefore the formal testing was not performed.

The Pearson correlation coefficients between all pairs of technical replicates within each platform ranged from 0.88 to 0.99 (Table 1 and Figure 1, A–D). The consistency of gene expression levels, as measured by the magnitude of the correlation, between technical replicates did not change with block age. We had a power of 0.85 to detect the following percentages of the concordance explained by the block age: 32% and 35% for prostate NuGen + Affymetrix and NanoString, respectively, and 41% for ovarian samples on both platforms. The consistency of gene expression levels depended on the input RNA amounts for the NanoString platform (Supplemental Figure S2). This can be explained by higher input amounts corresponding to a higher percentage of detectable probes. We had a power of 0.85 to detect 45% of the concordance explained by RNA input amounts. We also found lower correlations between replicates among the ovarian samples that presented with more extensive tissue necrosis.
and 33 (13%) in ovarian were less than 0.3. Among annotations, we mapped all 236 genes Q15 and the size is proportional to the absolute value of the correlations. Salmon represents positive correlations and blue represents negative correlations. C and D: Density plots of the observed correlations between the gene expression values on the NuGen + Affymetrix and NanoString platforms (solid lines) and the correlations under the null hypothesis of no association between the expression values measured on the NuGen + Affymetrix and NanoString platforms obtained by permuting sample labels (dashed lines).

Concordance between NuGen + Affymetrix and NanoString

By using NetAffx annotations, we mapped all 236 genes surveyed on the NanoString Cancer panel assay to a total of 256 transcript clusters on the Affymetrix GeneChip Human Gene 1.0 ST array. For these genes we calculated correlations between the samples assayed on both platforms. For this analysis, we averaged transcript expression values for technical replicates within each platform. Most of the genes had high positive correlations, with lower correlations predominantly found for the genes with lower expression levels (Figure 2, A and B). Of the 256 correlations, 77 (30%) in prostate and 33 (13%) in ovarian were less than 0.3. Among these genes with lower correlation, 18 mapping pairs, corresponding to 15 unique gene symbols, were common in both diseases: BCR, CASP10, CEBPA, CSF3, CYP1A1, FLT3, GATA1, HRAS, LMO2, MLH1, MLL, MPL, TFE3, WEE1, and WNT10B. We performed a permutation analysis and obtained distributions of the correlations between pairs of genes measured on two platforms under the null scenario of spurious correlations between the unrelated measurements, which can be high for relatively small sample sizes. Null scenario correlations were computed for 100 permutations of the sample labels for one of the platforms. From this analysis we show that for both prostate and ovarian cohorts the observed correlations between the platforms were higher than expected by chance (Figure 2, C and D).

Biological Validation

To provide evidence that both platforms provide biologically useful information we considered established gene signatures that distinguish prostate cancer versus noncancerous tissue and high versus low Gleason grade in prostate cancer, and clear cell versus serous adenocarcinoma for the ovarian cancer cohort. The prostate tumor versus normal tissue signature was obtained from Oncomine, and was defined by genes that were significantly up- or down-regulated in tumor versus normal comparison in four prostate cancer studies\textsuperscript{24–27} and had a median rank of the differential expression of less than 150 across these four studies. An mRNA signature related to Gleason grade was taken from Penney et al.\textsuperscript{28} For ovarian cancer, the signature was retrieved as the union of three gene signatures describing differences between serous and clear cell carcinomas from GeneSigDb.\textsuperscript{29–32} Gene signatures used for validation are presented in Supplemental Table S2.

We performed principal components analysis using genes from each signature that were represented on Affymetrix and NanoString platforms (Figure 3). For each comparison we observed meaningful separation of the classes defined by the signatures on the first two principal components.

In addition, for the Gleason signature genes, we calculated the log-fold changes in the gene expression values between high and low Gleason grade tumors. We compared our results with the log-fold changes observed in the Gleason signature. By using the NetAffx annotations implemented in the Bioconductor packages oligo and pd.hugene.1.0.st.v1, we mapped 150 of 157 genes from the signature to 157 Affymetrix transcript clusters. The correlation between the log-fold changes from two studies was 0.57. Only 10 genes from the 157-gene Gleason signature were represented on the NanoString Cancer Panel. The correlation between the log-fold changes observed on NanoString and previously published data were 0.89.

Discussion

The reliable use of archival FFPE samples for mRNA signature development will provide enormous opportunities
for precision medicine in oncology by unlocking banked tissue collected over decades with mature associated outcomes. Previous work has established that new approaches for RNA expression profiling in FFPE correlate well with corresponding fresh-frozen samples. In this study, we sought to determine the reliability of two approaches for mRNA expression profiling in archival FFPE samples from two representative research cohorts. We selected samples with block ages most typically used for biomarker discovery studies in prostate and ovarian cancer. Prostate cancer has a long natural course and a follow-up period of more than 10 years typically is necessary to study progression and lethality. Ovarian cancer is a much more aggressive disease, and a follow-up period of 5 years is common for lethality and progression studies. The NanoString platform allows assessment using low RNA input quantity and provides expression data on a prespecified set of genes. Whole-genome expression data can be obtained using an array-based approach after amplification.

We found that both the NanoString platform and whole-genome amplification followed by array profiling produced highly concordant results across technical replicates. These findings were independent of FFPE block ages up to 21 years old, and held for the suboptimal RNA input amounts down to 50 ng, highlighting the clinical applicability of this approach. There was good correlation of mRNA abundances between the NanoString and NuGen + Affymetrix expression profiling platforms investigated here. As expected from experience with gene expression profiling from fresh-frozen specimens, reproducibility diminished with the expression levels, and reliable assessment of gene expression remains imperfect for low-abundance transcripts. Furthermore, although we did not intend to directly compare results at the sample level with mRNA expression data from fresh-frozen tissue, we did observe biological correlations to existing literature from fresh-frozen samples, suggesting that biologically relevant signatures can be derived from archival FFPE samples.

Concordance between microarray gene expression profiling and the NanoString nCounter platform has been shown previously for frozen specimens. To our knowledge, our study is the first to compare the performance of these two platforms on FFPE specimens across two different diseases. We have shown the feasibility of discovery using larger microarray assays and a follow-up validation using the targeted NanoString nCounter approach.

Although the study sample size was relatively modest, no sample failed to produce usable results on either platform. This is in contrast to our prior experience with the Illumina whole-genome DASL HT assay platform, in which filtering...
out low-quality samples was essential to maintain the reproducibility of the differential gene expression inference.\textsuperscript{36} This study did not compare the DASL platform with these more recent profiling approaches for FFPE.

Our comparative study and several recent studies\textsuperscript{35,37} that used NuGen + Affymetrix and NanoString platforms have shown the feasibility of these platforms for clinical applications such as discovery of the gene expression signatures for personalized medicine and its applications from FFPE biopsy material. However, the development of additional statistical methodologies specifically tailored for preprocessing of the gene expression data from FFPE tissues still might be beneficial because the reproducibility between the technical replicates observed in our study allows for further improvement.

With a growing acceptance of the utility of mRNA expression profiles in precision medicine,\textsuperscript{5,38,39} researchers will require reliable methods for assessing gene expression data from archival specimens. In this study, we found that two modern platforms, one for more targeted profiling and one looking across the genome, can produce reliable, biologically relevant signatures. Importantly, we showed this for two different cancers with very different RNA yields: challenging prostate cancer and ovarian cancer, for which RNA yields usually are abundant. Differences in the RNA extraction techniques between the ovarian and prostate cases reflected previous experience and optimization by the two groups of investigators involved. Also, it previously was shown that both the RNeasy and RecoverAll kits allow extraction of total RNA of similar integrity and suitable for microarray hybridization.\textsuperscript{4,40} Similar findings in both of our cohorts additionally suggest that extraction protocols did not have much influence on the RNA used for library preparation. From this perspective our analysis provided additional independent validation.

Although additional experience and pilot studies in other cancer types are needed, our data suggest identical conclusions for each cohort that we studied, which suggests that our conclusions are generalizable across cohorts of different diseases and block age ranges. The data presented here should help researchers use FFPE tissue to answer questions central to the care of their patients.

Acknowledgments

We thank Alex Forrest-Hay, Brian Read, and Affymetrix, who provided the 1.0 ST array plates for the prostate cancer study and contributed extensively to the scientific discussion regarding the assay; the late Dr. Edward Fox (Microarray Core Facility, Dana Farber Cancer Institute) for contributions and collaboration; Kathryn Penney, Travis Gerke, Michael Pitt, and Elizabeth Nuttall for assistance on the prostate cancer study; and the participants in the Physicians’ Health Study prostate cancer project.


Supplemental Data

Supplemental material for this article can be found at http://dx.doi.org/10.1016/j.jmoldx.2015.02.002.

References


Supplemental Figure S1  Boxplots of raw and RMA normalized NuGen + Affymetrix prostate sample data.

Supplemental Figure S2  Boxplots of the correlations between technical replicates against the input RNA amount.