

Role For Anatomical Tissue I in Microarray Analysis of Clinical Samples

Paul Norton, Peter Haak, Bree Berghuis, Eric Kort, James Resau

Introduction

Molecular medicine holds the promise of tailoring treatment according to the genetic or protein expression characteristics of a patient or their diseased tissue. However, modern multiplex analysis platforms, including cDNA microarray technology, can only be as good as the tissue assayed. Technologies like Laser Capture Microdissection and RNA Amplification have enabled researchers to complete microarray-based gene expression studies on samples that would have previously been unusable due to their small size.¹⁻⁷ As the amount of sampled tissue decreases, there is a need for an accompanying increase in quality control to assure that the tissue sampled is representative of the disease process.

For microarray analysis to become a reliable diagnostic tool, it is imperative that the reliability and accuracy of identification of the clinical sample be confirmed by histologic analysis and correlated with the gene expression assay. In parallel the microarray production quality must be controlled to ensure accurate and reproducible results.

By combining the strengths of microarray analysis with the long-standing knowledge of histopathology we can complement and supplement each systems value.

In this report, we describe a workflow for identifying representative tissue suitable for gene expression analysis based on its histological characteristics. Furthermore, we demonstrate the quantitative value of specific histologic features of a sample as predictors of RNA yield and quality, and subsequent accuracy and reproducibility of information gleaned from the gene expression analysis using microarrays. These methods will assist in the planning of genomic analyses by enabling clinicians and researchers to accurately determine the suitability of a given tissue sample for gene expression analysis, the number of replicates that can reliably be provided by the sample, and the appropriateness of the sample classification (e.g. normal vs. tumor) as it relates to comparing the paired samples.

This technology will improve quality control and quality assurance of microarray analysis using either cDNA or commercially available gene platforms (e.g. Affymetrix, Agilent). For research and developmental studies, cDNA genomic analysis is cost effective and accurate but only if it meets the rigorous QC and QA standards characteristic of CLIA and CAP approved laboratories.

In this study, we selected a series of nine sarcoma cases from the Van Andel Tissue Repository⁴ which were reported to contain patient matched samples of tumor and normal tissue. We report here the results of paired histologic and microarray based gene expression analysis on these samples, and propose a system for determining likelihood of successful RNA recovery based on nuclei count and tissue volume. A schematic flowchart of the proposed workflow is shown in Figure 1.

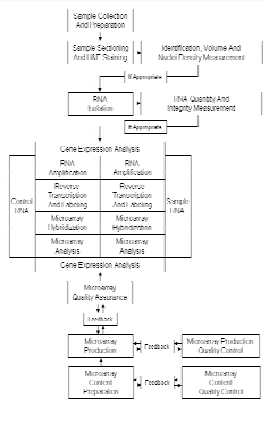


Figure 1. Schematic flowchart for microarray analysis of frozen clinical samples.

Sequential decision making is essential to successful gene expression analysis of clinical samples. The appropriateness of the analysis is driven by the quality and identity of the sample and also the quality of the microarray product.

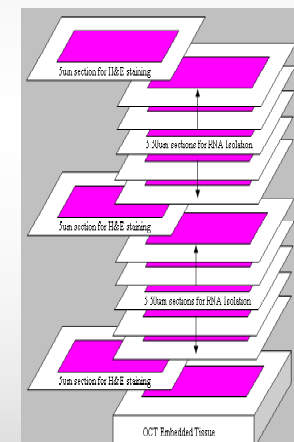


Figure 2. Paired Histology and Gene Expression Analysis Paradigm.

It is expected that differences in histology between sections within each spot was used to filter the data. The data sets were normalized by pin-dependent Lowess normalization and compared by Pearson correlation to determine degree of similarity between expected normal and tumor pairs.

Materials and Methods

Van Andel Tissue Repository (VATR)

Frozen human tissues were collected from Grand Rapids, Michigan area hospitals in accordance with HIPAA and IRB approved protocols.⁴

MIAME (Minimum Information About A Microarray Experiment)

Requirements for microarray-based gene expression experiments as outlined by MIAME have been implemented for this report.⁸

Raw microarray data is available on our website at www.microarray.vanl.org.

Tissue Embedding, Sectioning, and Homogenization

Samples were embedded in OCT Compound and sectioned using a Leica Cryostat at -20C. The sample was faced at 5um until a reproducible area of tissue was present. A 5um section was then prepared for histological analysis (Figure 2). Five sections were then cut at 50um thickness and transferred to 10ml TRIzol Reagent for total RNA isolation. Another 5um section was cut for H&E staining, followed by five more 50um sections for RNA isolation. A final 5um section was cut for histological analysis. The sample set now consisted of two aliquots for gene expression analysis bordered by three histologic sections.

Hematoxylin and Eosin Staining

Slides were post-fixed in 70% alcohol, rinsed in water and Hematoxylin and Eosin stained using a Leica Autostainer XL. The slides were differentiated in a 0.5% solution of acetic acid and rinsed in water between stains. The slides were then dehydrated through a series of graded alcohols and xylene and coverslipped using a Leica CV5030 automated cover slipper.

Image Capture and Analysis

Digital images of representative areas of the tissue were recorded using a Nikon E600 microscope equipped with a Spot Insight color CCD camera. The images were analyzed using Image Pro Plus V4.1 to determine the total area and number of nuclei in the selected region of interest. Each case had multiple images taken at 20X. For each case, five representative regions were selected for calibration. The number of nuclear pixels was automatically identified for each calibration image by software developed by our lab, applying the segmentation method of Ridler and Calvard to the blue channel of each image. The nuclei were also manually counted for the calibration images, to generate a conversion factor for converting pixel counts into number of nuclei. This scaling factor was then used with our software to automatically enumerate the nuclei in the remainder of the images to calculate a mean pixel count per mm² for each case.

RNA Isolation

RNA isolation was performed according to manufacturer's specifications (Invitrogen).

RNA Amplification

The Ambion MessageAmp aRNA Kit was used to amplify all RNA samples. One round of amplification was performed for all samples according to manufacturer's specifications. 10ug aRNA was used for all subsequent direct labeling reactions.

Microarray Production / Quality Control

Microarrays were produced using a custom contact arrayer designed by the Laboratory of Microarray Technology of the Van Andel Institute and Beta Integrated Concepts. The arrays consist of 19,968 features that are a combination of human cDNA clones from the IMAGE cloneset and positive and negative controls. The control sequences are included on the arrays for the necessary verification of a lack of nonspecific binding and to demonstrate that expected ratios can be observed when known amounts of control mRNA are spiked into the sample.

To control for print plate order and orientation errors that can potentially be made during microarray production or microtiter plate construction, we have devised a method of plate registration that allows for immediate identification of any errors. We insert a specific control cDNA into a patterned selection of wells on each of the 96-well storage plates that make up the clone set to control for potential technician errors that can be made during the preparation of the plates for microarray production and during production itself. Two unique wells of each 96-well plate contain an *Arabidopsis thaliana* clone that is visualized when the complementary mRNA is added to the labeling reaction and subsequently hybridized to the array (Figure 3).

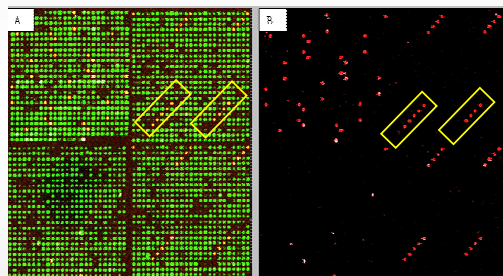


Figure 3. Microarray production quality control.

Panel A: Hybridization of a Cyanine-3 (green) labeled oligonucleotide complementary to the primer sequence common to all human cDNA spots on the array allows for visualization of microarray production efficiency and spot morphology.
Panel B: Hybridization of a Cyanine-5 (red) labeled oligonucleotide complementary to the primer sequence of the Arabidopsis thaliana plate registration control sequence allows for verification of print plate order and orientation during microarray production.

Post-Production Quality Assurance of Microarrays

To control for laboratory, technician, and sample preparation variability, concurrent hybridizations using RNA from cells with well-characterized chromosomal abnormalities are performed with each microarray experiment. Colorectal carcinoma cell lines DLD1-0746 and DLD1-1370 have duplications of chromosomes 7 and 13 respectively as confirmed by spectral karyotyping.

Generation of Fluorescently Labeled cDNA Targets

We selected for microarray analysis samples 7 and 8 because they represented the clearest cases of correctly and incorrectly identified tumor/normal pairs respectively. 10ug aRNA was used as template for direct labeling by reverse transcription using SuperScript¹¹ reverse transcriptase and cyanine5-dCTP and cyanine3-dCTP. Following reverse transcription, the aRNA template was degraded with RNase One. The labeled targets were then purified using QIAquick PCR purification kits and concentrated with Microcon YM-50 filters. Hybridizations were performed in 35% formamide at 50C for 20 hours. All samples were hybridized in color-inverted replicates against a common Universal Human Reference (Stratagene Inc.).

Stringency Washes

After hybridization, the arrays were washed in the following solutions for 5 minutes each.

1X SSC, 0.1% SDS

0.2X SSC, 0.1% SDS

0.2X SSC

0.1X SSC

The first wash was done at 50oC. Subsequent washes were done at room temperature.

All washes were performed in the dark and in an environmental control chamber to minimize photo-bleaching and oxidation of the dyes. The arrays were dried by centrifugation to avoid residual contamination associated with evaporation.

Microarray Data Acquisition and Analysis

Microarrays were scanned with a Perkin Elmer Scannarray Lite and analyzed with GenePix 5.0 software. A statistically based custom flagging script that evaluates signal above background, percent of saturated pixels and red to green regression of pixels within each spot was used to filter the data. The data sets were normalized by pin-dependent Lowess normalization and compared by Pearson correlation to determine degree of similarity between expected normal and tumor pairs.

Results

Tissue Identification

Representative H&E images of the cases are shown in figure 4. It is clear that for these cases, the identification and viability of the samples as reported by the surgeon at the time of excision is not always accurate.

RNA Isolation

The total mass of RNA isolated for each sample was compared to the volume of the tissue and to the tissue nuclei counts. Nuclei density more accurately predicted the success of RNA isolation than did tissue volume as determined by Pearson Correlation. The correlation of nuclei density and RNA mass was 0.82, whereas for tissue volume and RNA mass it was only 0.47. Of the 36 isolations performed, RNA isolation was most successful, in terms of acquiring at least the minimum amount of RNA required for reliable amplification, for those with both high volumes and high nuclei per area. But, for samples with high volume and low nuclei per area, sufficient RNA was not obtained.

Microarray Analysis

The results of microarray analysis for representative samples are shown in Figure 5. These data show positive correlations (0.85, sd 0.039) between reciprocally labeled replicates and adjacent sections (a vs. b, c vs. d) indicating reproducibility of gene expression profile within samples. This is reflected in the histological similarity of adjacent H&E sections (e.g. a vs. b, c vs. d). The average correlation of sample pairs (ab vs. cd) however, show a strong discrepancy between samples seven and eight. Sample seven, being a true normal/tumor pair showed an average correlation of 0.76 while sample eight, a misidentified normal/tumor pair had a correlation of 0.90.

Post-Production Quality Assurance

The cytogenetic changes observed by SKY in the DLD1-0746 and DLD1-1370 cell lines were recapitulated in the microarray data. As shown in figure 6, amplification of chromosomes 7 and 13 were evident in the microarray data as determined by CGMA, providing a positive control that the microarray production process resulted in arrays that could demonstrate gene expression changes in samples.

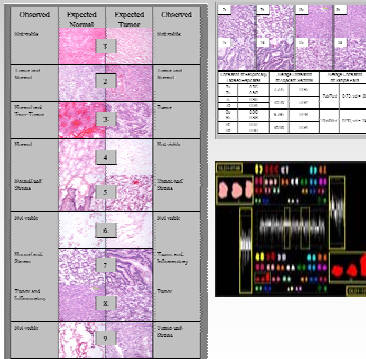


Figure 4. Expected vs. Observed Tissue Identity.

It is clear that for these cases, the identification and viability of the samples as determined by the surgeon at the time of excision is not always accurate.

Discussion

Array analysis is a comparison between two different but related samples or a sample against a well characterized standard⁷. The paired samples are usually related to each other by a common patient or tissue source. The paired samples could just as likely be infected vs. uninfected or developed vs. undeveloped. The data is dependent on the accuracy of the division of the samples between the "normal" and the "tumor". The requirement also exists for the sample to be the "equivalent" normal to match the tumor. For this reason it is optimal for a trained surgical pathologist to observe the samples and confirm that the "pairs" are correct and appropriate and that the tissues are suitable for molecular analysis. Only pathologists are trained to recognize subtle differences between histologic structures that would prove of vital importance in the interpretation of the genomic results. Through the use of the established frozen section technology, pathologists can do both a molecular and an anatomical analysis.

Specifically, the potential pitfall in paired analysis analyzing cancer is that it is possible and sometimes common for a tumor sample to contain significant numbers of normal cells and vice versa. This is because cancer by its nature invades into the surrounding and adjacent tissues. If two tissues contain equivalent volumes of normal and tumor then there will be no difference in the genomic signature but if one isolates normal from tumor using laser capture micro-dissection done by a skilled surgical pathologist then the unique signature will be correct. Similarly, mRNA analysis is dependent on viable, well preserved cells and tissue. If the sample is necrotic or contains heterogeneous cellular material, the amount of mRNA is adversely affected. A simple histologic section can often determine if the sample is viable, cellular and contains the appropriate histologic tissue.

The histologic signature of frozen clinical samples is predictive of RNA isolation success and consequently, the ability to perform gene expression analysis using microarrays. Furthermore, these data indicate that nuclei count is a more robust predictor of RNA isolation success than tissue volume. Tissue classification at the surgical level continues to be subjective and qualitative. From this study it is apparent that the histology of clinical samples must be determined because the label is not necessarily indicative of the content. This is not unreasonable or unexpected as the identifications of surgical samples are made by visual observation and gross analysis. However, if the paired samples are not coherent as to their normal and tumor designations, the results, while correct, will not be informative. This is evidenced by our gene expression analysis of sample numbers 7 and 8. By pairing the histologic and molecular signatures of the sample pairs it becomes clear that while sample number 7 represents a valid opportunity for comparison, sample number 8 does not.

The results of this study suggest that histologic analysis of all clinical samples prior to microarray analysis decreases the likelihood of artifactually biased microarray results. As biomarkers and gene expression patterns of disease continue to develop, microarrays become more capable of segregating individuals into groups based on prognosis of disease, ultimately leading to appropriate courses of treatment for each individual.^{11, 13, 14} The combination of histology and gene expression analysis has the power to improve the accuracy of clinical diagnostics and prognostics by precise segregation of individuals into meaningful groups, allowing for better informed therapeutic decision-making.^{8, 11}

References

1. Steirou, C., et al. 2002. Core Biopsies Can Be Used to Distinguish Differences in Expression Profiling by cDNA Microarrays. *Journal of Molecular Diagnostics*, 4(1): 30-36.
2. Theodor, P., et al. 2004. cDNA microarray analysis with amplified RNA after isolation of intact cellular RNA from neoplastic and non-neoplastic prostate tissue separated by laser microdissection. *International Journal of Oncology*, 24(5): 1085-92.
3. Kort, E., et al. 2003. A human tissue and data resource: an overview of opportunities, challenges, and development of a provider/researcher partnership model. *Comput Methods Programs Biomed*, 70(2): 137-50.
4. Brauma, A., et al. 2001. Minimum information a biologist should provide for microarray data. *Nature Genetics*, 29(4): 365-371.
5. Upend, MB, et al. 2004. Chromosome transfer induced aneuploidy results in chromosome degradation of the cellular transcriptome in immortalized and cancer cells. *Cancer Research*, Oct-11: 6419/6441-6449.
6. Wu, T.D. 2001. Analyzing gene expression data from DNA microarrays to identify candidate genes. *Journal of Pathology*, 195: 53-65.
7. Kern, W., et al. 2004. Gene expression profiling as a diagnostic tool in acute myeloid leukemia. *American Journal of Hematology*, 4(4): 225-37.
8. Robinson, J., et al. 2004. State of the science: molecular classification of breast cancer for clinical diagnostics. *Clinical Biochemistry*, 37(7): 572-8.
9. Fanto, T. 2004. A novel cDNA array method for detecting the chimeric genes in human leukemias. *Genes*, 5(2): 162-6.
10. Ramasamy, S., et al. 2001. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences of the United States of America*, 98(20): 15149-15154.
11. The Tumor Analysis Best Practices Working Group. 2004. Expression profiling - best practices for data generation and interpretation in clinical trials. *Nature Reviews*, 5: 229-37.
12. Wang, J., et al. 2003. Tumor classification and marker gene prediction by feature selection and fuzzy c-means clustering using microarray data. 2003. *BMC Bioinformatics*, 4(1): 60.
13. Quackenbush, J. 2004. Meeting the Challenges of Functional Genomics: From the Laboratory to the Clinic. *Pediatrics*, 2(5): 313-316.



VAN ANDEL INSTITUTE