

A 43-gene expression array for profiling clinical samples for development and validation of breast cancer recurrence test

David F. Englert¹, D. Alan Kerr II², Sarah A. Andres², Daniel J. Wilson¹ and James L. Wittliff²

¹Xceed Molecular, Toronto, ON, Canada M9W 1B3; ²Department of Biochemistry & Molecular Biology, University of Louisville, Louisville, KY 40202

Abstract

Background: Candidate genes identified with global expression arrays must be verified on independent patient cohorts to validate classification according to clinical outcomes. We identified expression signatures in a large body of microarray data and by independent analysis of estrogen receptor positive (ER+) breast cancers. We analyzed biopsy tissues with qPCR and a focused microarray system (Ziplex® Automated Workstation, Xceed Molecular) to verify clinical relevance of candidate genes and to define a test for predicting cancer recurrence.

Methods: Total RNA was isolated, and gene expression relative to β -actin was measured by qPCR using Power Sybr® Green PCR Master Mix (Applied Biosystems). Labeled cRNA was prepared by *in vitro* transcription (TotalPrep™ RNA Amplification Kit, Ambion) and hybridized on a custom flow-through microarray (TipChip, Xceed Molecular) containing probes for 43 transcripts and control probes. Intensities of different probes designed for the same mRNA target were compared to optimize the array for robust signal intensity and specificity.

Results: Results indicated concordance between qPCR and the focused microarray for many of 43 candidate genes. Preliminary data for many of the genes correlated well with clinical outcomes. Altered expression of 21 candidate genes identified from earlier microarray studies and 6 genes derived from ER+ breast cancers predicted early recurrence.

Conclusions: Candidate genes identified from microarrays were verified as expression markers for predicting cancer recurrence using an independent sample set and two analytical platforms. Comparisons of data from qPCR and separate microarray probes provide unambiguous identification of mRNA targets for development of a clinical test to predict breast cancer behavior.

Supported in part by Phi Beta Psi Charity Trust, UofL Office of Technology Transfer and from Xceed Molecular Corp. SAA & DAK II are recipients of IPIBS Graduate Fellowships.

Flow-through Chip Technology

Ziplex¹ uses flow-through chip technology where probes are immobilized within a highly porous silicon matrix providing a very high surface area for binding probes. During hybridization, washing and detection, samples and reagents are continually perfused through the chip within a microplate well to provide efficient mixing and transport of target molecules to the immobilized probes. Figure 1 shows an electron micrograph of the silicon chip which is fixed to a plastic tube to form a TipChip.

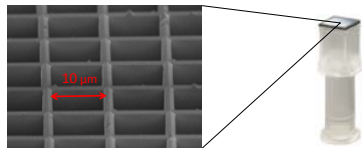


Figure 1 – Scanning EM of the porous Si substrate and an assembled TipChip.

¹Ziplex is for Research Use Only

Sample Hybridization on the Ziplex Workstation

Total RNA samples were analyzed for quality with the Agilent Bioanalyzer, amplified and biotin-labeled by oligo-dT primed *in vitro* transcription. TipChip microarrays, samples, and reagents were loaded into specific microplate wells, and then hybridization, washing, chemiluminescent imaging and data reduction were performed automatically on the Ziplex workstation (Figure 2).

The Ziplex manifold picks up the TipChips and lowers them into specific wells where solutions are repeatedly aspirated and dispensed through the chips. Eight TipChips were hybridized and analyzed simultaneously in less than three hours. Tables of mean intensities and coefficients of variation of triplicate spots for each probe were output by the instrument and analyzed on an external computer.

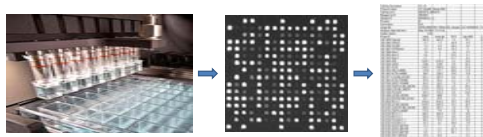


Figure 2 – Ziplex manifold and TipChips, chemiluminescent image, report from Ziplex.

Probe Design and Screening for the Breast Cancer Recurrence Array

Oligonucleotide probes were designed to hybridize to the 3' ends of target genes. Two or more probes were designed for each target, and TipChip arrays with the probes printed in triplicate were hybridized with labeled cRNA prepared from 109 total RNA samples from breast cancer tissues. The probe intensities from probes for the same nominal targets were correlated to screen them for relative signal strength and specificity. Probes that hybridize exclusively to a given target transcript are expected to correlate well, and a regression line through the data is expected to pass through the origin.

All correlations between three probe designs for a target gene are shown in Figure 3. All three probes correlated well with each other, although the three linear plots at the left indicate subtle differences in hybridization patterns with different probes. The best correlation was between probes 2 and 3 indicating that these two probes provide essentially identical information. Probe 1 produced a somewhat different pattern of expression than the other two probes, indicating that it hybridized to somewhat different molecular entities in some samples. Poorly correlating probes are flagged for critical attention in subsequent analysis.

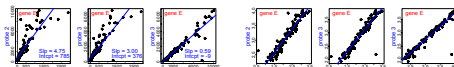


Figure 3 – Correlations between three different probes for the same target gene. Signals for 109 samples are plotted for each probe. Three correlations with linear scales are on the left side, and three correlations with logarithmic scales are on the right. Samples with very small signals are plotted with open circles.

Survival Probabilities of Breast Cancers with Known Prognostic Factors

To ascertain that the population of de-identified invasive ductal/carcinoma ER+ carcinoma specimens and associated patient follow-up selected for the development of the 43 gene TipChip array was representative of ER+ cancer biopsies, survival probabilities were assessed according to currently accepted prognostic factors. Collectively, the data show the expected relationship of patient survival to known prognostic factors (Figure 4).

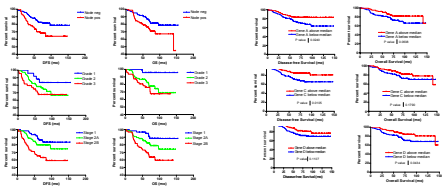


Figure 4 – Survival probabilities of breast cancer patients with three standard prognostic factors.

Figure 5 – Disease-free survival (left, N = 187) and overall survival (right, N = 200) as a function of expression of individual genes measured by qPCR in ER+ biopsies of stage 1-2 breast cancer.

Survival and Expression of Individual Genes Measured by qPCR

Individual genes from the 43-gene set were analyzed by qPCR, and breast cancer patients were classified according to the expression levels of the individual genes. The disease-free and overall survival times of the patient classes were compared in Kaplan-Meier plots (Figure 5). Statistically significant differences in survival were observed for some genes.

Concordance of Expression Levels from TipChip Probes and qPCR

We are evaluating a subset of 13 of the 43 genes as an independent predictor of patient outcome in ER+ breast cancers, and the performance of several individual probes for representative genes in a preliminary recurrence model are reported below. We have correlated expression levels measured with qPCR and Ziplex for the genes in 65 samples (examples in Figure 6). The expression levels determined by the two platforms correlate reasonably well for most genes and the majority of samples, although some measurements exhibited outliers.

We have previously observed discordance between qPCR and microarrays. In a differential expression study using the MAQC reference samples and protocol, three of 52 target genes were consistently found to be outliers between qPCR and three microarray platforms including Ziplex².

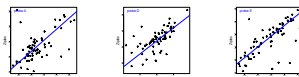


Figure 6 – Correlations between measurements of expression from Ziplex and qPCR. Data from both platforms were normalized to the signals from the β -actin gene.

²<http://www.xceedmolecular.com/whitepapers.html>

Survival and Expression of Individual Genes Measured by Ziplex

Expression levels were analyzed on the 43-gene TipChip array, and breast cancer patients were classified according to the expression levels measured by individual probes on the array. The disease-free and overall survival of the patient classes were compared in Kaplan-Meier plots (Figure 7). As with qPCR, statistically significant differences in survival were observed for some individual genes.

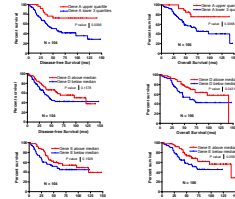


Figure 7 – Kaplan-Meier plots showing patient classification with single genes analyzed on the TipChip array.

Survival of Patients Classified with a 13-Genes Prediction Model

Gene expression results from multiple Ziplex probes for 13 genes and clinical follow-up (disease-free and overall survival probabilities) were evaluated using Partek™ Genomics Suite. A population of 52 specimens was used as a training set to generate a multivariate model. The most accurate model, determined by K-nearest neighbor clustering, was then used to classify an independent test population of 52 specimens into low and high recurrence groups. The resulting molecular signature composed of 13 genes collectively predicted breast cancer recurrence as shown in the Kaplan-Meier survival plots. (Figure 8).

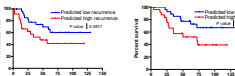


Figure 8 – Kaplan-Meier plots showing patient classification with a 13-gene recurrence model.

Conclusions

- Ziplex hybridization probes were compared to identify probes that produced maximum signals with minimal evidence of cross hybridization.
- Expression results from qPCR and the Ziplex array were generally concordant, although some samples were discordant between the platforms.
- With either analytical method, expression levels of some individual candidate genes permitted classification of breast cancer patients with statistically significant differences in disease-free and/or overall survival.
- A preliminary 13-gene recurrence prediction model was developed with a training set of clinical samples with annotated follow-up and tested with an independent test set of samples.