Journal of the Korean Statistical Society 45 (2016) 149–155

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

A comparison between DASL and Affymetrix on probing the whole-transcriptome

Jaesik Jeong [a], Robert Audet [b], Jenny Chang [c], Helen Wong [c], Scooter Willis [d], Brandon Young [e], Susan Edgerton [f], Ann Thor [f], George Sledge [g], Renata Duchnowska [h], Jacek Jassem [i], Krzysztof Adamowicz [j], Brian Leyland-Jones [d], Changyu Shen [k,*]

[a] *Department of Statistics, Chonnam National University, Republic of Korea*
[b] *VM Institute of Research, Montreal, Canada*
[c] *The Methodist Hospital Research Institute, Houston, USA*
[d] *Edith Sanford Breast Cancer Research, Siuox Falls, USA*
[e] *Genomics and Cell based Screening cores, Jupiter, USA*
[f] *Department of Pathology, University of Colorado, Aurora, USA*
[g] *Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, USA*
[h] *Military Institute of Medicine, Warsaw, Poland*
[i] *Medical University of Gdansk, Gdansk, Poland*
[j] *Regional Oncology Center, Gdansk, Poland*
[k] *Department of Biostatistics, Indiana University 46202, USA*

ARTICLE INFO

Article history:
Received 11 February 2015
Accepted 3 September 2015
Available online 9 October 2015

AMS 2000 subject classifications:
62P10
62P99

Keywords:
DASL
Affymetrix
Spearman rank correlation
PAM50 classifier
Cohen's kappa

ABSTRACT

Whole-transcriptome microarray analysis has become a popular strategy to study gene-expression in cancer, exploiting the largely available formalin-fixed paraffin-embedded (FFPE) sample resources. However, there have been relatively few comparative studies evaluating the performance of the different gene-expression array platforms.

We compared two commonly used whole-transcriptome microarray platforms: Illumina human whole genome cDNA-mediated annealing, selection extension and ligation (DASL) beadchip and Affymetrix U133 Plus2 GeneChip (Affymetrix). Gene expression data based on both platforms were collected on the same total RNA extracted from FFPE tissue samples of 221 advanced breast cancer patients. Correlations between two platforms were assessed using Pearson and Spearman correlation coefficients (CCs). For both platforms we also assessed coefficient of variation, which measures relative dispersion. Finally, we compared the applicability of DASL and Affymetrix for classification of breast cancer molecular subtypes using the PAM50 classifiers.

Overall, there was a statistically significant, positive gene- and patient-wise correlation between the two platforms, with stronger relationship for patient-wise CC. The relative dispersion was smaller in DASL compared to Affymetrix. The consistency in subtype classification for both microarray platforms was weak (63%).

We observed weak, yet positive correlation between two platforms and different magnitudes of correlations were observed according to the metrics used.

* Corresponding author.
E-mail address: chashen@iupui.edu (C. Shen).

http://dx.doi.org/10.1016/j.jkss.2015.09.001
1226-3192/

1. Introduction

Formalin-fixed paraffin-embedded (FFPE) tissue samples are typically the only specimens available for large cancer patient populations in clinical setting, and are increasingly used in whole-transcriptome microarray analysis (Waldron et al., 2012). These tools are now being used to evaluate gene-expression signatures that identify differentially expressed genes from a wide variety of cell types and conditions. These genome wide expression tools are evolving and are increasingly included in standard diagnostic procedures for invasive breast cancer that normally include only histological assessment, anatomic staging and ER, PR and HER2 protein expression levels. It is now widely accepted that breast cancer can be further classified into five major subtypes namely: Luminal A, Luminal B, HER2-enriched, Basal-like and Normal-like. The RT-qPCR expression of a 50-gene set (PAM50) has been proposed to standardize this subtype classification and is now routinely used as an independent predictor of survival for breast cancer (Parker et al., 2009).

Although genome wide expression tools have been widely used, there have been relatively few comparative studies evaluating the performance of the different gene-expression array platforms and most of them studied the correlation between Affymetrix and cDNA microarray (Barnes, Freudenberg, Thompson, Aronow, & Pavlidis, 2005; Carter, Eklund, Mecham, Kohane, & Szallasi, 2005; Gry et al., 2009; Kuo, Jenssen, Butte, Ohno-Machado, & Kohane, 2002; Lee et al., 2003; Shankavaram et al., 2007; Woo et al., 2004). In this study, we compared the expression measures of two commonly used microarray platforms: Illumina human whole genome cDNA-mediated annealing, selection extension and ligation beadchip (DASL) and Affymetrix U133 Plus2 GeneChip (Affymetrix). For both platforms, gene expression was analyzed using the same total RNA extracted from FFPE primary tumor samples from 221 advanced breast cancer patients. Hybridization differences between the two whole-transcriptome analysis platforms were evaluated using Pearson and Spearman correlation coefficients. For both platforms we also assessed the coefficient of variation (CV), which measures relative dispersion. Finally, we compared the applicability of DASL and Affymetrix for the molecular subtyping of breast cancer using the PAM50 intrinsic classifier.

2. Methods

Data

Two commonly used microarray platform (DASL and Affymetrix) datasets were obtained using archival primary tumor samples of 221 advanced breast cancer patients. Demographic information for those patients is summarized in Table 1. The Affymetrix dataset included 54,675 probes and 20,026 genes. For the genes with multiple probes, the average of probes was calculated. The DASL dataset included 18,049 genes at the gene level. For the purpose of comparing two platforms, we selected common genes and patients between two platforms and ended up with 15,581 genes from 221 patients available for analysis. More details are given in the Supplementary materials (Table S1).

Table 1
Summary of demographic information.

Age	Min	Mean	Median	Max
Freq	26	52.41	52	86
Race	White	Black	Asian	Unknown
Freq	181(85%)	23(10.8%)	8(3.7%)	1(0.5%)
Ethnic	Hispanic	Non-Hispanic	Not reported	Unknown
Freq	18(8.5%)	142(66.7%)	28(13.1%)	25(11.7%)

RNA extraction

Total RNA was extracted from at least three 5 μm thick FFPE sections placed in an RNAse-free low-binding plastic tube (Abramovitz et al., 2008). Sections were deparaffinized with 100% xylene for 3 min at 50 °C and centrifuged, twice washed with ethanol and air dried before being digested with Proteinase K at 50 °C overnight. RNA isolation was performed using Ambion's RecoverAll kits and the resulting RNA was quantified using a Nanodrop spectrophotometer.

Statistical measures

Pearson/Spearman correlation coefficients: The correlation between two platform data was analyzed using Pearson and Spearman correlation coefficients (CCs). Pearson CC represents a measure of the strength of linear dependence between two variables. Pearson CC (ρ) between two data vectors, $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ is defined: