Characterization of Errors in Transcriptomics by Spotted DNA microarray Data

Tobias K. Karakach[1 *], and Peter D. Wentzell[1] karakach@dal.ca

1 Department of Chemistry, Dalhousie University, Halifax, Nova Scotia, CANADA B3H 4J3

Keywords: DNA Microarrays, Measurement Error Models, Bootstrapping.

The so-called '-*omics* revolution' is characterized by high throughput measurements and vast quantities of data. Unfortunately, the quality of the measurements is often over-estimated and, as such, simplistic assumptions regarding the structure of the measurement errors are made subsequent to application of routine data analysis techniques. In transcriptomics by spotted DNA microarrays, it is beginning to emerge that the conceptual simplicity of the technology belies a potentially complex measurement error structure. For example it has been shown variously that the data exhibit a non-constant variance leading to the recent development of models to 'fix' the problem. Regrettably, the extent to which these error inhomogeneities affect individual microarray data has not been systematically explored and, consequently, the structure of the measurement uncertainties after transformations are applied to the data remains uncharacterized.

This work presents the results of a systematic characterization of the measurement error structures for spotted DNA microarray data as well as a model that compartmentalizes the total variance exhibited by the measured intensity ratios into distinct components that can be measured independently. In particular, one of the ratio variance components, which is often ignored in microarray data analysis, is the uncertainty associated with the measurement of the ratio itself. In most microarray data, the ratio is determined as a mean or median of pixel intensities comprising a spot and no implicit information is provided about the accuracy with which this quantity is measured. In this work, the ratio is measured as an orthogonal regression slope of the pixel intensities comprising the spot and a bootstrap approach is employed in determining the magnitude of the uncertainty associated with determining this ratio. Those measurements for which this value dominates the total variance are eliminated in order to maintain the distribution of errors. The structure of the measurement uncertainties is then characterized empirically using replicate measurements.