

## **GLOBAL CORRELATION ANALYSIS BETWEEN REDUNDANT PROBE SETS USING A LARGE COLLECTION OF ARABIDOPSIS ATH1 EXPRESSION PROFILING DATA**

Xiangqin Cui<sup>1,3</sup> and Ann Loraine<sup>1,2</sup>

<sup>1</sup>Section on Statistical Genetics, Department of Biostatistics

<sup>2</sup>Department of Genetics

<sup>3</sup>Department of Medicine

University of Alabama at Birmingham  
Birmingham, AL 35294

Oligo-based expression microarrays from Affymetrix typically contain thousands of redundant probe sets that match different regions of the same gene. We used linear regression and correlation to survey redundant probe set behavior across nearly 500 quality-screened experiments from the Arabidopsis ATH1 array manufactured by Affymetrix. We found that expression values from redundant probe set pairs were often poorly correlated. Pre-filtering expression results using MAS5.0 “present-absent” calls increased the overall percentage of well-correlated probe sets. However, poor correlation was still observed for a substantial number of probe set pairs. Visual inspection of non-correlated probe sets’ target genes suggests that some may be inappropriately merged gene models and represent independently expressed, but neighboring loci. Others may reflect differential regulation of alternative 3-prime end processing. Results are on-line at [http://www.transvar.org/exp\\_cor/analysis](http://www.transvar.org/exp_cor/analysis).