Gene Expression Platforms for Global Co-Expression Analyses: A Comparison of spotted cDNA microarrays, Affymetrix microarrays, and SAGE

Obi Griffith, Erin Pleasance, Debra Fulton, Misha Bilenky, Sheldon McKay, Mehrdad Oveisi, Peter Ruzanov, Kim Wong, Scott Zuyderduyn, Martin Ester, Asim Siddiqui and Steven Jones
Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada, V5Z 4E6

Introduction: We have conducted a comprehensive comparison of three major expression platforms: cDNA microarray, oligonucleotide microarray and serial analysis of gene expression (SAGE) using large sets of available data for *Homo sapiens*. Several studies have compared two of the three platforms to evaluate the consistency of expression profiles for a single tissue or sample set but none have determined if these translate into reliable co-expression patterns for global analyses across many conditions.
Methods: We analyzed a recently published data set of 1202 cDNA microarray experiments (Stuart *et al*. 2003), 242 SAGE libraries from the Gene Expression Omnibus (GEO), and 667 Affymetrix (HG-U133A) oligonucleotide microarray experiments also from GEO. All expression data were assigned to LocusLink Ids resulting in an overlap set of 5881 unambiguously mapped genes represented in all three platforms. Using standard co-expression analysis methods, we have assessed each platform for internal consistency and performed all pairwise platform comparisons.
Results: Internal consistency was determined by randomly dividing the datasets in half and comparing the Pearson distances for each subset: Affymetrix gave an $r = 0.94$, cDNA microarray an $r = 0.89$, and SAGE an $r = 0.355$ ($p < 0.001$) when at least 100 overlapping data points were required for Affymetrix and cDNA microarray, and 10 required for SAGE. When datasets were semi-randomly divided, keeping similar or replicate experiments in the same subset, internal consistencies were reduced to 0.255 for Affymetrix, 0.287 for cDNA and 0.150 for SAGE. All pairwise comparisons found poor correlation between platforms ($r < 0.1$, $p < 0.001$). Comparison against the Gene Ontology (GO) showed that all three platforms identify more co-expressed gene pairs with common GO biological process annotations than random data. However, SAGE and Affymetrix performed equally best with microarray performing only slightly better than random.
Conclusions: The results of a global co-expression analysis differ greatly depending on the platform and dataset utilized. Each dataset identified different subsets of biologically relevant gene interactions as identified by GO. Researchers should be cautioned against using any one set of publicly available expression data for integration with or validation of other interaction types. A filtered set of the most reliable gene pairs from each platform may be the best candidates for further study.