Review

UHN Microarray Centre

MAQC Project: Sparking scientific debate about the reliability of microarray technology

Review of: MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature Biotechnology, 2006, 24(9):1151

Although some publications have found excellent inter- and intraplatform reproducibility among different microarray platforms[1-3], other studies have found very little overlap in differentially expressed genes among various platforms[4,5]. Many researchers use DNA microarrays as a high-throughput screening tool for obtaining expression profiles and understand the limitations of microarray technology. It has even been suggested that using a combination of microarray platforms may overcome the inherent biases of each method and that such an experimental approach will identify robust changes in gene expression[1]. However, members of the scientific community are concerned about the lack of concordance in results obtained from different microarray platforms. As microarrays evolve, and the idea of using them for diagnostic and regulated clinical experimentation becomes reality, such concerns must be addressed. The Microarray Quality Control (MAQC) project was initiated to provide quality control tools to the microarray community and to develop guidelines for microarray data analysis by providing large reference datasets along with accessible reference RNA samples[6].

International consortiums like the External RNA Control Consortium (ERCC) and Microarray Gene Expression Data (MGED) Society have provided the microarray community with RNA controls to enable consistency and reliability of gene expression platforms and standards for annotating microarray data, respectively. The MAQC project is also an international, community-wide effort with the goal of experimentally addressing the key issues surrounding the reliability of DNA microarray data and establishing operational metrics to assess the performance of seven microarray platforms[7].

In summary, Phase I of the MAQC project involved 137 participants from 51 organisations. Gene expression levels from two commercially available RNA samples (Universal Human Reference RNA (UHRR; Stratagene) and Human Brain Reference RNA (HBRR; Ambion)) in four titration pools (100% UHRR; 100% HBRR; 75% UHRR:25% HBRR; and 25% UHRR:75% HBRR) were evaluated on seven microarray platforms (Applied Biosystems, Affymetrix, Agilent Technologies, GE Healthcare, Illumina, Eppendorf, and National Cancer Institute). Each microarray platform was evaluated at three independent test sites and five replicates for each of the four sample types (6 platforms were one-colour) were assayed at each site (total of approximately 60 hybridisations per platform). The RNA samples were also tested on three alternative gene expression platforms (TaqMan Gene Expression Assays (Applied Biosystems), StaRT-PCR (Gene Express), and QuantiGene (Panomics)) to assess the relative accuracy of each microarray platform[8].

When analysing the data, the MAQC project based most of their results on a set of 12,091 common genes that are represented on all 6 of the high-density microarray platforms, albeit using different probe sequences[8]. Analysis of the MAQC data set found that one-colour microarray platforms had a median coefficient of variation (CV) of 5-15% for quantitative signal and a concordance rate of 80-95% for the qualitative detection call between sample replicates[8]. As expected, the variation increased when data from the different test sites using the same platform were included. The results indicate that, for the sample types chosen and these test sites, microarray results were repeatable within each test site, reproducible between test sites and comparable across platforms, even considering the difference in probe sequences across platforms as well as unique protocols for labelling and expression detection[8]. Other issues that may have affected the interplatform variability include possible annotation problems and the specificity of each probe on the array[8]. This study also highlighted differences in

various performance metrics between microarray platforms. For instance, the data suggests that the Affymetrix platform had better intersite reproducibility, Illumina had better intrasite repeatability, and that some platforms (e.g., Agilent one-colour and Applied Biosystems) were more comparable to TaqMan assays[8].

Much of the debate sparked by this study has to do with the way in which that was analysed. The MAQC Consortium suggests that relying on the statistical significance (*P* value) instead of the actual measured quantity of differential expression (ratio or fold change) when identifying differentially expressed genes, may contribute to the lack of agreement between microarray platforms[9]. The MAQC group has been criticised for implying that prioritising genes by magnitude of effect is more productive than prioritising genes by the level of statistical significance[10], although the MAQC does recommend a nonstringent level of statistical significance ($P < 0.05$ or 0.01) be used in conjunction with fold change[11]. MAQC emphasises that criterion such as sensitivity and specificity should also be considered when developing rules for determining which genes are differentially expressed, but that reproducibility is also critical as results which are not reproducible are of no use in scientific environments[11]. Another concern was that differences in the normalisation methods used by each platform might have led to discrepancies[10]. Another study has suggested that the impact of normalisation methods on the reproducibility of gene lists becomes minimal when the fold change, instead of the *P* value, is used as the ranking criterion for gene selection[9]. Also, the analysis of the MAQC data set does not include biology-based performance metrics such as Gene Ontology terms or pathways[8].

Critics of this study have suggested a flawed experimental design, specifically the small size (n=5)[10] and the choice of RNA samples[12]. The MAQC defends its sample selection; that such distinct reference RNA samples were deliberately chosen so that the technical accuracy of the different platforms could be determined. As part of the MAQC project, Shippy *et al*. have found that RNA titration samples are a valuable tool for assessing microarray platform performance and different analysis methods[13]. The MAQC Consortium points out that the comparability of microarray data in this study does not necessarily mean that the same level of consistency would be achieved in experiments where more biologically similar samples were compared[8].

Initial analysis of the MAQC data set indicates that each microarray platform has made different trade-offs with respect to repeatability, sensitivity, specificity and ratio compression[8]. The MAQC project provides a framework for assessing the potential of microarray technologies as a tool to provide reliable gene expression data for clinical and regulatory purposes[8]. This study concludes that the technical performance of microarrays as assessed in the MAQC project supports their continued use for gene expression profiling in basic and applied research and may lead to their use as a clinical diagnostic tool.

Further analyses of the MAQC reference data set (Phase I) has been published[14,15]. The results from Phase II of the MAQC project are scheduled for release in September 2008[6]. These results will most likely continue to fan the flames of debate over the reliability and reproducibility of microarray technology.

References:

1. Bosotti, R. et al. Cross platform microarray analysis for robust identification of differentially expressed genes. BMC Bioinformatics, 2007, 8 Suppl 1:S5
2. Wang, Y. et al. Large scale real-time PCR validation on gene expression measurements from two commercial long-oligonucleotide microarrays. BMC Genomics, 2006, 7:59
3. Kuo, W.P. et al. A sequence-oriented comparison of gene expression measurements across different hybridization-based technologies. Nature Biotechnology, 2006, 24(7):832
4. Tan, P.K. et al. Evaluation of gene expression measurements from commercial microarray platforms. Nucleic Acids Research, 2003, 31(19):5676

5. Miller, R.M. et al. Dysregulation of gene expression in the 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine-lesioned mouse substantia nigra. J Neurosci, 2004, 24(34):7445
6. http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/
7. http://www.nature.com/nbt/focus/maqc/index.html
8. MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature Biotechnology, 2006, 24(9):1151
9. Guo, L. et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. Nat Biotechnol, 2006, 24(9):1162
10. Klebanov, L. et al. Statistical methods and microarray data. Letter to the Editor, 2007, Nature Biotechnology, 25(1):25
11. Shi, L. et al. Reply to Letter to the Editor, Nature Biotechnology, 2007, 25(1):26\
12. Liang, P. MAQC papers over the cracks. Letter to the Editor, 2007, Nature Biotechnology, 25(1):27
13. Shippy, R. et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. Nature Biotechnology, 2007, 24(9):1123
14. Canales, R.D. et al. Evaluation of DNA microarray results with quantitative gene expression platforms. Nature Biotechnology, 2007, 24(9):1115
15. Patterson, T.A., et al. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) Project. Nature Biotechnology, 2007, 24(9):1140