nature biotechnology

Making the most of microarrays

A major, multicenter study of microarray performance is a first step in translating the technology from bench to bedside.

No technology embodies the rise of 'omic' science more than the DNA microarray. First reduced to practice in the early 1990s, it has since undergone numerous iterations, adaptations and refinements to achieve its present status as the platform of choice for massively parallel gene expression profiling. Today, several thousand papers describing data from microarrays are published each year. Sales of arrayers, array scanners and microarray kits to the academic and industrial R&D community represent a multi-billion-dollar business. The microarray has even made its first forays into the clinic, with the US Food and Drug Administration's approval of the 'AmpliChip' to help physicians tailor patient dosages of drugs that are metabolized differentially by cytochrome P450 enzyme variants.

And yet doubts linger about the reproducibility of microarray experiments at different sites, the comparability of results on different platforms and even the variability of microarray results in the same laboratory. After 15 years of research and development, broad consensus is still lacking concerning best practice not only for experimental design and sample preparation, but also for data acquisition, statistical analysis and interpretation.

Though problematic for bench research, lack of resolution of these issues continues to even more seriously hamper translation of microarray technology into the regulatory and clinical settings. Indeed, several regulatory authorities have been wrestling with the problem of how and when (and indeed whether) to implement microarray expression profiling data as part of their decision-making processes. The move in the past two years to accept voluntary genomic data submissions by regulatory agencies overseeing human and environmental safety was the first in a long series of steps that will be needed.

One of the next steps can be found in this issue, which presents the first formal results of the MicroArray Quality Control (MAQC) Consortium—an unprecedented, community-wide effort, spearheaded by FDA scientists, that seeks to experimentally address the key issues surrounding the reliability of DNA microarray data. MAQC brings together more than a hundred researchers at 51 academic, government and commercial institutions to assess the performance of seven microarray platforms in profiling the expression of two commercially available RNA sample types. Results are compared not only at different locations and between different microarray formats but also in relation to three more traditional quantitative gene expression assays.

Although the direct comparison of microarray platforms and the establishment of common controls for microarray experiments is nothing new—several cross-format studies have already been published, and other groups, such as the External RNA Controls Consortium's (ERCC), are developing standardized RNA controls—it is the size and comprehensiveness of the data set generated by the MAQC effort that is unique. In the main study, ~60 hybridizations were carried out on each of the seven platforms; >1,300 microarrays were used during the entire project.

MAQC's main conclusions confirm that, with careful experimental design and appropriate data transformation and analysis, microarray data can indeed be reproducible and comparable among different formats and laboratories, irrespective of sample labeling format. The data also demonstrate that fold change results from microarray experiments correlate closely with results from assays like quantitative reverse transcription PCR.

The levels of variation observed between microarray runs by MAQC were relatively low and largely attributable to cross-platform differences in probe binding to alternatively spliced transcripts or to transcripts that show a high degree of cross-hybridization to probes other than their own. Thus, although factors as diverse as day-to-day fluctuations in atmospheric ozone levels (which effect cyanidine 5 fluorescence), nuclease levels in sample tissues and the quality of microarray production between batches have all been cited as influencing array performance, on the basis of the data presented here, experimental variability appears manageable.

Another clear finding is that the days of the simple two-sample t-test as a means of ranking differentially expressed genes are surely numbered. A key take-home message is that statistical analysis in regulatory submissions and clinical diagnostics is likely to be different from that used in basic research and discovery. In the case of the MAQC study—where the goal was to optimize intra- and inter-platform reproducibility—the approach was to limit the number of transcripts identified and to sort differentially expressed genes using fold-change ranking with a nonstringent P-value cutoff. But for experiments that seek to identify differentially expressed transcripts at or near the lower limits of detection, this tradeoff between reproducibility on the one hand and precision and sensitivity on the other is likely to shift, and a different type of statistical analysis will be required. There is no one-size-fits-all statistical solution.

Overall, the MAQC study represents a landmark in DNA microarray research because it provides the community with a thoroughly characterized reference data set against which new refinements in platforms and probe sets can be compared. It complements other initiatives, such as the ERCC, in providing the community with two commercially available human reference RNA samples that can be used to calibrate arrays in ongoing quality control and performance validation efforts. It can be used as the foundation for combining other microarray studies, thereby realizing the true cumulative potential of microarray data, which will undoubtedly lead to new insights. And from a clinical perspective, it validates the DNA microarray as a tool that is sufficiently robust and reliable to be embraced for use on hard-to-obtain human tissue samples.

Clearly, microarrays have a long way to go before they can be used to support regulatory decision-making or accurate and consistent prediction of patient outcomes in the clinic. But the MAQC study has given us a solid foundation from which to build.