

Of genomics and bioinformatics

The Pharmacogenomics Journal (2010) 10, 245–246;
doi:10.1038/tpj.2010.59

It is evident that, over the past 15 years, genomics (including genetics) has changed the way we conduct biomedical research. The genomic application involves a wide range of scientific disciplines aimed at understanding the underlying mechanisms of biological systems, biomarker discovery, predictive toxicology and personalized medicine. There is not a single passing day without a news report on the progress in genomics and its effect on public health. For example, a genetic test to personalize treatment with warfarin reduced the hospitalization rate by almost a third. In 2007, MammaPrint became the first Food and Drug Administration (FDA)-approved microarray-based prognostic tool for recurrence testing of breast cancer. Molecular diagnostic tests on variants of drug-metabolizing enzymes have also shown benefits towards improving drug efficacy, while minimizing drug toxicity.

Genomic technology has evolved rapidly, due, in part, to the technological advances in manufacturing and sample handling. Since gene expression microarray technology was introduced in 1995, the scientific community has also seen the demand for genome-wide association studies (GWASs) and next-generation sequencing, which are promising tools to identify biomarkers for discovery and personalized medicine. Although these high-throughput genomic platforms are different in design and application, they share a common characteristic, that is, hundreds of thousands of hypotheses are simultaneously tested in a single experiment. This poses several challenges to the field of bioinformatics—in particular, data analysis for extracting biologically instructive information from the massive data sets. This challenge is largely due to the fact that the inter-dependency of multiple hypotheses tested in a single experiment is unknown, and thus the choice of statistical methods may have a significant effect on the ability to distinguish the true discovery from the false discovery. In most cases, different methods may lead to different results and, unfortunately, even to a different understanding of the underlying biology. The authors of some statistical studies have claimed that the promise of genomics in clinical settings is over-optimistic. However, rebuttals have suggested that these negative studies suffer technical flaws in statistical analysis. Arguments and debates surrounding genomics are often centered on the choice of bioinformatics approaches and on how to apply these approaches to analyze genomic data.

A major role of the US FDA is to protect and promote public health. Over the past 10 years, we have seen a decline in the number of new molecular entities and biologics license applications submitted to the FDA for approval, whereas the research expenditures in both pharmaceutical companies and the National Institutes of Health research grants have increased. Consequently, the FDA has become an active participant in working with the research community and sponsors to evaluate new technologies and inventions in drug discovery and development, in efforts to improve public health. For example, the FDA established the Voluntary eXploratory Data Submission program and the Critical Path Initiative (<http://www.fda.gov/oc/initiatives/criticalpath/>). An important goal of both of these programs is to assess genomic technology and its perspective use in drug discovery and development, as well as in the regulatory process. The programs recognize that there remains considerable work to be completed and controversies to be resolved before a consensus can be reached for the appropriate use of these approaches in the regulatory setting. Thus, an FDA endeavor is being undertaken in parallel with ongoing research aimed at determining the best scientific practices for the use of genomics data. An important parallel research effort is the MicroArray Quality Control (MAQC) project. To aid in data collection, analysis and visualization, FDA scientists also developed a public genomic tool, ArrayTrack, to support the review of the genomic data.¹

The MAQC is an FDA-led community-wide effort to assess the technical performance, practical utilities and bioinformatics approaches for genomic technologies in biomedical research, drug development and safety assessment. The first phase of the MAQC project (MAQC-I) evaluated various issues related to microarray technology, including, but not limited to, quality control, data analysis, cross-laboratory and cross-platform consistency, for determining differentially expressed genes and the reliability of the technology. The project involved 137 scientists from 51 organizations, and was published in six research papers in the September 2006 special issue of *Nature Biotechnology*.² To ensure transparency, the MAQC-I data are publicly available through several sources. In addition, the MAQC-I reference RNA samples were made available commercially to offer an opportunity for the MAQC-I results and conclusions to be verified by third parties.

The field of microarray and GWAS has rapidly advanced owing to the increased interest and application in biomarkers for risk assessment and clinical application. The microarray-based genomic biomarkers are particularly

promising in the clinical setting, whereas GWAS shows potential for identifying specific single-nucleotide polymorphisms (SNPs) relating to complex diseases such as diabetes. However, the debate remains as to whether a reliable and reproducible genomic biomarker can be derived from these technologies. Some of the confusion is based on the use of different bioinformatics approaches that may result in the identification of qualitatively different genomic biomarkers for the same data set. The question remains as to whether there exists a best practice for determining a reliable and reproducible genomic biomarker from a single genomic data set, and if so, what is the recommended practice for clinical application, safety evaluation and the regulatory review process. These questions have led to the second phase of MAQC (MAQC-II) that was centered on the development and validation of genomics biomarkers. This large effort involving approximately 200 participants from 10 countries and over 80 organizations included individual scientists from government, academic and private sectors. More specifically, the MAQC-II evaluated methodologies for developing and validating microarray-based models aimed at predicting toxicological and clinical end points (that is, diagnostic, prognostic and treatment selection). In addition, the technical performance of GWAS platforms and the impact of different methods for analyzing GWAS data were also evaluated. The results of these efforts are published in the compendium of papers that follows along with those published in *Nature Biotechnology*. The results of these carefully designed and integrated studies are expected to substantially affect the clinical and regulatory use of genomic data.

Genomic technology continues to advance and evolve. Quality control and appropriate bioinformatics processing

will remain a challenge for any new high-throughput molecular technology. For example, next-generation sequencing represents a new wave in genomics and promises to replace some existing methods, such as microarray analysis, by delivering more robust and accurate results with a broader range of application. However, its potential in biomarker discovery, drug development and clinical application has yet to be fully assessed. Therefore, a third phase of the MAQC project (MAQC-III) is in progress to address these needs. Such an undertaking requires a continued collaborative effort among government agencies, industry and the academic research community to objectively evaluate the approaches in depth and build a consensus on the best way to move the genomics field forward. The MAQC projects represent models of how government agencies can have a leading role in translating research into applications that benefit regulatory science and public health.

W Slikker Jr

National Center for Toxicological Research, The US Food and Drug Administration (FDA), Jefferson, AR, USA
E-mail: william.slikker@fda.hhs.gov

REFERENCES

- 1 Tong W, Harris SC, Fang H, Shi L, Perkins R, Goodsaid F *et al*. An integrated bioinformatics infrastructure essential for advancing pharmacogenomics and personalized medicine in the context of the FDA's critical path initiative. *Drug Discov Today* 2007; **4**: 3–8.
- 2 The MAQC Consortium. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006; **24**, (9) and 24(10s) 1151–1161.