Genomics and the Impact of New Technologies on the Management of Colorectal Cancer

D. Paul Harkin

Queen's University, Belfast, Northern Ireland, United Kingdom, and ArraDx Ltd, Craigavon, Northern Ireland, United Kingdom

Key Words. Genomics • Array-based comparative genomic hybridization • Single-nucleotide polymorphism arrays • Microarrays • RNA Later • Formalin-fixed paraffin-embedded

LEARNING OBJECTIVES

After completing this course, the reader will be able to:

1. Describe the genomic techniques that can determine gene expression in clinical samples.
2. Explain how these genomic techniques can be used in preclinical drug development.
3. Discuss the evolving role of genomic techniques in clinical medicine.

Abstract

High-throughput genomic technologies have the potential to have a major impact on preclinical and clinical drug development and the selection and stratification of patients in clinical trials. These technologies, which are at varying stages of commercialization, include array-based comparative genomic hybridization, single-nucleotide polymorphism arrays, and (the most mature example) expression-based arrays. One of the rate-limiting steps in the routine clinical application of expression array-based technology is the need for suitable clinical samples. One of the major challenges moving forward, therefore, relates to the ability to use formalin-fixed, paraffin-embedded–derived tissue in expression profiling-based approaches. The Oncologist 2006;11:988–991

Introduction

Although proteomics is fast evolving, only genomic platforms are currently ready for high throughput analysis. This paper therefore focuses on genomics and, more specifically, on the impact of gene expression microarrays and their use in preclinical and clinical drug development. This spans the entire process from the identification of novel targets to the selection of patients most likely to benefit from molecules designed to hit those targets. It is hoped that microarray technologies will both speed drug development and improve the rate of success.

Major Genomic Technologies

For many years, three major genomic technologies have been available: comparative genomic hybridization (array-based CGH), single-nucleotide polymorphism analysis (SNP...
arrays), and gene expression analysis (microarrays), which emerged from the Northern blot. Over the past few years, all three have been adapted for high throughput analysis.

Array-Based CGH
Given the capacity for high throughput, array-based CGH can now measure alterations in copy number across the entire genome. The assay is based on the comparative hybridization of differently labeled samples (typically normal tissue and tumor) on an array of genomic clones. These can be large genomic clones such as bacterial artificial chromosomes (BACs) and phage artificial chromosomes (PACs) and now also oligonucleotides [1].

The technique detects regions of the genome that are either amplified or deleted, in the hope that these represent overexpression of oncogenes or loss of tumor suppressor genes and that these gains or losses can be correlated with different disease settings. Array-based CGH is far more quantitative than the metaphase chromosome spreads that preceded it and also has superior resolution and dynamic range. The trend is now toward using oligonucleotide arrays, and these are proving a more reproducible and adaptable platform than the previous BAC and PAC arrays (Fig. 1) [1].

However, while array-based CGH is reproducible, quick, based on genomic DNA (which has the advantage of stability), and relatively cost-effective, it is unable to detect abnormalities (such as translocations) that do not result in changes in copy number. Nor can it detect single copy changes, so it is not useful in investigating loss of heterozygosity in chromosome 15q, for example. Until recently, no commercially validated platforms have been available, and there is still a dearth of standardized software for data analysis.

SNP Arrays
These arrays are used to map regions of the genome linked to defined phenotypes [2–5]. The technology is based on the immobilization of oligonucleotide probes corresponding to both alleles of a specific SNP. Hybridizing a target sequence (from tumor or normal tissue) to the array can then determine if a specific SNP is AA, AB, or BB; and array-based profiling is now able to interrogate thousands of SNPs in a single experiment (Fig. 2) [6].

As with array-based CGH, the technology is based on analysis of genomic DNA, and the protocol is relatively straightforward. However, the human genome contains approximately 10 million SNPs [7]. The largest array in prospect will contain only one million. The problem could be overcome by looking for SNPs that fall within regions of the genome which are conserved within haplotypes, assuming that one or two SNPs in that region will be representative of the haplotype block. There is again a lack of reliable bioinformatic tools, and SNP arrays are more expensive than their CGH counterparts.

Microarrays
Gene expression microarrays represent the most mature example of high throughput technology and have been validated in many basic and applied clinical studies across a range of tumor types [8–14]. Within colorectal cancer, this is illustrated by the retrospective study of Wang et al., who identified a 23-gene signature that predicted recurrence of Dukes’ B tumors [15]. The signature was validated in 36 patients and had an overall performance accuracy of 78%. The study demonstrated that it was possible to identify a poor-prognosis group of Dukes’ B patients who could benefit from adjuvant chemotherapy.

The available platforms are all reliable (with Affymetrix [Santa Clara, CA] the market leader, other competing platforms include Agilent Technologies, Inc. [Palo Alto, CA] and Amersham [Little Chalfont, Buckinghamshire, U.K.]) [16–18], but their different content means that standardization across platforms is not good. Researchers are best advised to choose a platform and then stick with it.

---

**Figure 1.** Comparison of regions of loss and gain on chromosome 17q using both oligo and BAC-based CGH arrays. From [1], with permission. Abbreviation: BAC, bacterial artificial chromosome; CGH, comparative genomic hybridization.

**Figure 2.** Schematic of the steps involved in preparing genomic DNA for Affymetrix array-based SNP analysis. From [6]. Image courtesy of Affymetrix, Inc. Abbreviations: PCR, polymerase chain reaction; RE, restriction enzyme; SNP, single-nucleotide polymorphism.
We have seen the development of the international MIAMI (minimum information about a microarray experiment) guidelines, which should ensure that data generated in one laboratory can be duplicated in another, provided that the same platform is used. Within a platform, data analysis software is now reliable.

The arrays measure changes in RNA (rather than DNA) expression, and RNA is notoriously unstable. However, microarrays provide the necessary genome-wide view of expression changes, although they do not measure entire transcriptomes.

Concerns center on the availability and quality of clinical samples used for expression profiling. Traditionally, samples are fresh frozen in liquid nitrogen. However, a new preservative, “RNA Later,” is excellent at preserving RNA quality.

**Requirement for Appropriate Clinical Samples**

Whatever the means of preservation, it is essential that the tissue samples be representative of the populations being studied (i.e., in ethnic background) and that samples be accompanied by clear information on diagnosis, staging, and history of treatment and response. If these criteria are met, valid genetic signatures can be generated using as few as 20 tumor samples from clear responders and 20 from patients with clear progression. Once the initial signature has been developed, the data can be enhanced by including patients who fall in the middle ground. In general, signatures need to be validated using tumor samples derived from a different center whose patients are treated with the same agents. Given that fresh frozen samples are often not available, one of the main issues in applying microarray technology has been whether it can be applied to the formalin-fixed paraffin-embedded (FFPE) tissues collected in large numbers during the course of many key cooperative group studies. Key issues relate to the reproducibility of the RNA extraction process and whether RNA derived from FFPE tissue is of sufficient quality to allow the generation of robust prognostic or predictive signatures.

**A Model for Drug Development**

Initial preclinical work will use cell lines and xenografts to profile a drug, generating biological data that allow building of an expression profile that predicts sensitivity or resistance. The predictive signature is then used to stratify patients included in phase II trials. Access to tumor samples and phase II response data then allows an in vivo gene signature to be established, which feeds back and enhances the predictive signature developed preclinically. That enhanced signature can then be used to select and stratify patients for phase III trials (Fig. 3).

That, at least, is a model of the ideal. Whether genomic technologies will actually come into routine use in preclinical and clinical drug development remains to be seen.

**Disclosure of Potential Conflicts of Interest**

The author indicates no potential conflicts of interest.

**REFERENCES**


13 Workman P. The opportunities and challenges of personalized genome-
based molecular therapies for cancer: targets, technologies, and molecular
and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol
15 Wang Y, Jatkoe T, Zhang Y et al. Gene expression profiles and molecu-
lar markers to predict recurrence of Dukes’ B colon cancer. J Clin Oncol
16 Kreiner T, Buck KT. Moving toward whole-genome analysis: a technol-
17 Hardiman G. Microarray platforms—comparisons and contrasts. Pharma-
18 Harkin DP. Uncovering functionally relevant signaling pathways using