

Microarrays and Clinical Investigations

Edison T. Liu, M.D., and Krishna R. Karuturi, Ph.D.

Two articles in this issue of the *Journal* tell a similar story: primary acute myeloid leukemia (AML) may be divided into subclasses according to gene-expression profiles. Given the number of articles that describe the correlative and predictive usefulness of array-based molecular classifications, especially in leukemias, these outcomes no longer elicit surprise. But perhaps they should — not only because of their clinical usefulness, as outlined in the editorial by Grimwade and Haferlach (pages 1676–1678), but also because of the rapid pace at which expression genomics is changing the conduct of clinical research.

Not too long ago, a single molecular marker, such as *N-myc* amplification, would be studied alone in the clinical setting, and its individual prognostic capabilities extracted from the biologic din (e.g., the French–American–British [FAB] classification scheme and blast count) with the use of logistic-regression analysis. Large clinical studies involving 400 to 1000 patients were often deemed necessary for statistical validation. A high level of quantitative precision was needed in that single laboratory test in order to uncover associations — not to mention satisfying downstream regulatory requirements.

By contrast, the two current articles — one by Bullinger et al. (pages 1605–1616) and one by Valk et al. (pages 1617–1628) — present what might be considered the genomic approach to the discovery of biomarkers in clinical trials. In each case, the number of variables assessed (approximately 26,000 genes and 13,000 genes, respectively) far exceeded the number of patients studied (116 and 285, respectively). The specific behavior of individual genes was less informative than the composite movement of defined clusters of genes, and therefore a defined cluster was used as a biomarker; thus, quantitative variations in individual markers of 20 to 30 percent are easily tolerated. The power of microarray technology is its ability to use somewhat imprecise patterns of gene expression rather than exact thresholds of individual markers. Most intriguingly, the

clinical and biologic findings are remarkably robust. The results of several independent studies appear to be consistent, even though relatively small samples were used for each study. This new genomic approach to the discovery of biomarkers suggests that a succession of smaller studies, performed quickly and possibly with the use of progressively better technology, will outperform larger studies that are locked into outdated approaches. In other words, agility trumps size.^{1,2}

During this period of dramatic transition, students of clinical investigation will inevitably be challenged by completely new analytic approaches. In microarray experimentation, the difficulty no longer lies with the technology, but rather with the conceptual heterogeneity of the data analysis. Many investigators are troubled by the “voodoo” nature of array analysis and their impression that each study uses different analytic algorithms. Given these concerns, a more common-sense explanation of array-directed, genome-based analyses seems to be in order.^{1,3}

Expression microarrays are devices containing tens of thousands of short DNA probes of specified sequences, arrayed in an orderly fashion and tethered to a flat surface. Each spot on the array corresponds to a particular probe, and each probe corresponds to a short section of a gene. (Often, an individual spot represents a gene; however, in some formats, several spots cover one gene.) RNA derived from that gene may bind to the probe (or spot); this is the core reaction of microarray analysis. In studies like those described by Bullinger et al. and Valk et al., leukemia RNA is labeled with fluorescent dye. After the labeled RNA is incubated with the microarray and the unbound RNA is washed away, the amount of bound, labeled RNA is measured by the intensity of the fluorescent signal of the spot, which is indicative of the relative expression of the corresponding gene.

Both groups of investigators applied a standard filter to the expression data in order to eliminate as

much “noise” as possible before any analysis was performed. Only those genes whose expression actually varied among the tumors were included in the analysis. Both groups of investigators used supervised and unsupervised approaches to the analysis of array data. A supervised analysis starts with a gold-standard class definition — for example, a leukemia either has or does not have a t(8;21) translocation. Then the algorithm may ask for all genes that are statistically associated with the translocation status. An unsupervised analysis does not use any a priori class definitions, instead simply seeking to determine what structure is inherent in the data. Such an analysis is based on the question “which genes organize the leukemias into groups, with no reference to biologic outcomes?” A supervised analysis is more likely to uncover putative associations between genes and the cytogenetic class but may bias the outcome by forcing a model onto the data.

In the current studies, the two groups of investigators sought to identify a core or minimal group of genes whose differential expression can be used to assign AMLs to distinct molecular and prognostic classes. Both groups used what may be described as an orthogonal approach to data analysis: they used one method to derive a list of biologically important genes, another to classify tumor clusters, and yet another to extract the genes that were the most effective classifiers; finally, they attempted to associate these clusters with the clinical outcome or to validate their results. Such an approach grows out of the inherent belief that no single analytic tool is better than the others and that each tool alone may, in fact, introduce unrecognized bias.

Bullinger et al. first established, using unsupervised clustering, that the 6283 genes whose expression varied the most among the leukemic samples could be used to identify specific subgroups of AML. Subsequent analysis showed that these array-based clusters correlated with known cytogenetic and molecular aberrations. Using a supervised approach, called significance analysis of microarrays (SAM), they further identified genes whose individual expression correlated statistically with survival time. On the basis of these SAM-selected genes, the leukemias were then partitioned into two groups with the use of a different clustering approach (k-means cluster analysis). Standard Kaplan–Meier survival analysis confirmed the prognostic significance of these two clusters: one is associated with a good outcome, and the other with a poor outcome.

This dichotomization of the training samples into a good-prognosis group and a bad-prognosis group was important for the next step. Since one of their goals was to identify the minimal set of genes that can predict the prognosis of each AML, the investigators applied a supervised method called prediction analysis of microarrays (PAM) on this training set of samples. This approach identifies a set of genes whose expression is solidly associated with one prognostic class (in this case, a good or poor outcome). In this manner, all genes were reclassified according to their ability to separate the AMLs into good-prognosis and poor-prognosis groups. Those genes that were less useful in discriminating between these classes were eliminated.

The authors then applied a cross-validation process whereby the gene-prediction model is fit on the basis of 90 percent of the samples and then tested on the remaining 10 percent. When this process is executed multiple times with adjustments, the output is the smallest list of genes that can be used to predict the class of an individual sample with the minimal error rate. This analysis resulted in a minimal list of 133 prognostically relevant genes that predicted the outcome independently of known risk factors. The authors went one step further and tested their prognostic-gene signature on a separate validation set of samples.

Valk et al. also used a prefiltered set of genes arising from their array experiments and clustered, in an unsupervised manner, 285 AML samples. Sixteen array-based subgroups could be identified with the use of 2856 gene markers that were strongly correlated with different cytogenetic and molecular aberrations. The investigators then used SAM to narrow the set of genes down to the 599 that best discriminate the cytogenetic configurations. Applying PAM to this gene set, they then also identified the minimal number of genes that could be effectively used as a molecular classifier in individual patients. Their predictors of cytogenetic subgroups also performed well on a validation sample set.

What have we learned from these two array studies? First, despite the use of different microarray platforms, when such studies are performed properly, their results are surprisingly robust. Second, the distinct expression profiles of leukemias arise from their associated genetic mutations (e.g., tumors with an *AML1-ETO* translocation have a different profile than those with a *PML-RAR α* translocation). Third, there seems to be a hierarchy of effects of different genetic mutations on the gene-express-

sion signature of cancers (e.g., the *AML1-ETO* translocation has a greater effect than *ras* mutations). Fourth, expression profiling with the use of thousands of probes can uncover linkages between molecular subclasses and clinical outcomes that cannot be identified by standard cytogenetic analysis or the use of clinical variables.

What is the take-home message? Our usual thinking about biomarker discovery in clinical trials is about to change dramatically. In the future, clinical investigations will consist of small trials with a high density of data, precise patient stratification

according to the expression profile, and highly tailored analysis of microarray data.

From the Genome Institute of Singapore, Singapore.

1. Miller LD, Long PM, Wong L, Mukherjee S, McShane LM, Liu ET. Optimal gene expression analysis by microarrays. *Cancer Cell* 2002;2:353-61.
2. Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 2002;1:133-43.
3. Simon R. Diagnostic and prognostic prediction using gene expression profiles in high-dimensional microarray data. *Br J Cancer* 2003;89:1599-604.

Cytogenetics — In Color and Digitized

Thomas Ried, M.D.

Cytogenetics is the genetic analysis of cells, a discipline that has flourished since the chromosome-banding techniques introduced in 1969 by Torbjörn Caspersson and Lore Zech first provided a simple and inexpensive way to gauge the number and assess the structural integrity of chromosomes. Chromosome banding (see Figure 1A) is probably the most commonly performed genetic test: it is used about 500,000 times each year in North America (most laboratories use G-banding, named after the German chemist Gustav Giemsa). It is used prenatally and postnatally to screen for the presence of constitutional chromosomal aberrations associated with birth defects, as well as for the diagnosis and differential diagnosis of cancer — hematologic cancers in particular.

The application of cytogenetics to the screening of the genomes of patients has led to the linking of specific chromosomal abnormalities (and consequently genetic alterations) with specific diseases; such findings carry enormous implications for medicine. The first such specific correlation was made when Jérôme Lejeune, in Paris, detected extra copies of a small chromosome in patients with Down's syndrome. Other descriptions of constitutional abnormalities, involving both autosomes and the sex chromosomes, followed in rapid sequence.

Such associations were also established in cancer cells. The first report on the so-called Philadelphia chromosome (described by Peter Nowell and David Hungerford in 1960 and characterized by

Janet Rowley in 1973), created by an exchange of genetic material between chromosomes 9 and 22, launched the highly prolific field of cancer cytogenetics. (The Philadelphia chromosome reflects the specific target of imatinib mesylate in the treatment of chronic myeloid leukemia.) This work not only confirmed the relevance of cytogenetic abnormalities as disease-initiating events, but also led to the molecular cloning of numerous cancer-associated genes, many of which sit at the junctions of chromosomal translocations.

Cytogeneticists then adapted and combined molecular cloning and hybridization and so arrived at fluorescence in situ hybridization, which has had profound effects on cytogenetic diagnostic testing and research. Specific fluorescently tagged DNA probes could now be used to visualize, with dramatically increased resolution, small deletions and other genetic aberrations, as well as to map the chromosomal locations of genes. Complex clone libraries were developed for the "painting" of entire chromosomes or chromosome arms (see Figure 2), a technique that is useful not only for the confirmation of suspected aberrations in clinical diagnosis, but also for basic research efforts aimed at elucidating the evolution and structure of genomes, relationships between structure and function, mechanisms of DNA repair, and chromosome segregation.

Fluorescence in situ hybridization extended the application of cytogenetic diagnosis from chromosomes in metaphase to all stages of the cell cy-