GENOMIC BIOMARKER DISCOVERY:
BRINGING THE GENOME TO LIFE

A convergence of technological breakthroughs has taken genomic biomarker discovery to a new level. Between advances in sequencing, evolving array designs, and a more sophisticated understanding of genome architecture, simple tests for glucose, cholesterol, and human chorionic gonadotropin could soon be competing for pharmacy shelf space with gene expression– and epigenome-based diagnostics that promise to detect diseases earlier, stratify patients into treatment classes, and identify those most likely to respond to therapies. First to benefit: cancer patients. The era of personalized medicine is just around the corner. **By Jeffrey M. Perkel**

The US Food and Drug Administration approved the first gene expression microarray–based test for use in the United States on February 6, 2007. MammaPrint, from Amsterdam-based **Agendia,** is a breast cancer diagnostic that uses the combined expression of 70 genes to predict whether a woman is at high or low risk for her cancer to recur.

Based on **Agilent Technologies'** 60-mer oligonucleotide microarrays and costing about $4,200, MammaPrint helps doctors and their patients make more informed decisions, says René Bernards, Agendia's scientific director and professor of molecular carcinogenesis at the **Netherlands Cancer Institute (NCI).**

Patients classified as having a high risk of metastasis have a 50 percent 10-year survival, says Bernards, versus 96 percent for low-risk patients. A combination of adjuvant chemotherapy and hormonal therapy, however, increases the high-risk pool's survival to about 70 percent.

"If you can increase your chances of survival from 50 percent to 70 percent, that is fairly significant," he says. At the same time, Bernards continues, "you can show that the net benefit of adding chemotherapy to the hormonal therapy [for low-risk patients] is essentially zero."

Such prognostic clarity can pay both medical and economic dividends. About 90 percent of US breast cancer patients currently receive chemotherapy, Bernards estimates. If that number were reduced to 60 percent (the fraction of women whose tumors are labeled high risk), the healthcare system could save millions. Meanwhile, patients who either do not need, or cannot benefit from, chemotherapy avoid its unpleasant side effects.

It truly is individualized medicine. And MammaPrint is just the beginning. Researchers in both academia and industry are scrambling to find new genomic biomarkers based on both gene expression and epigenetic modifications, especially for cancer.

Serving Different Needs

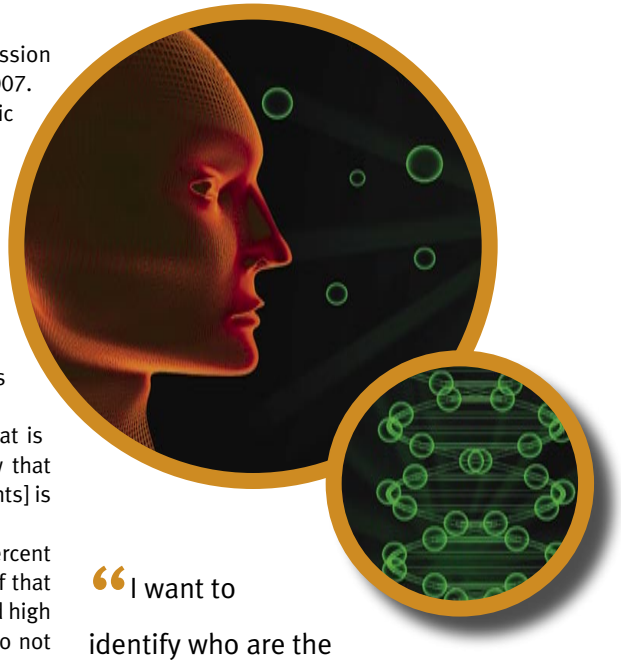
Genomic biomarkers come in several flavors, including disease detection and classification, treatment response prediction, treatment efficacy, and prognosis.

Their development follows a common theme. Typically, genomewide microarrays are used in a preliminary screen on a relatively small set of cell lines or tissue biopsies. The resulting biomarker candidates are then validated on an independent sample set, and typically migrated to a different assay platform, such as quantitative real-time PCR, which can better probe a small number of loci for a large number of samples.

Certain groups at **Wyeth** identify candidate patient-selection biomarkers using **Affymetrix's** U133 whole genome microarray and targeted signaling protein array analysis of drug-resistant and drug-sensitive cell lines, says Christina Coughlin, associate director of translational medicine efforts in the oncology pipeline at the company's Collegeville, Pennsylvania, facility. The 100 or 200 most-promising hits are then culled via more focused work using xenografts and genetically engineered mice.

"I want to identify who are the patients who will have true clinical benefit, and include them in my trial, and who will be resistant, and exclude them from the trial," says Coughlin.

Other biomarkers, like those found on MammaPrint's 70-gene panel, help continued >



CREDIT: ©ISTOCKPHOTO.COM/YAKOBCHUK

> "I want to identify who are the patients who will have true clinical benefit, and include them in my trial, and who will be resistant, and exclude them from the trial."

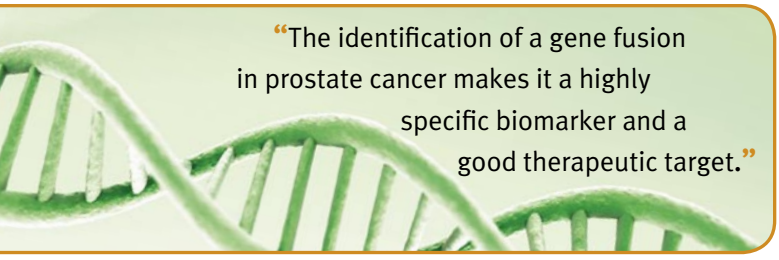
Look for these **Upcoming** Articles

Genomics 1 — April 4

RNAi — June 6

Proteomics 2 — August 1

Inclusion of companies in this article does not indicate endorsement by either AAAS or Science*, nor is it meant to imply that their products or services are superior to those of other companies.*



guide treatment decisions. Similarly, Oncotype DX, from Redwood City, California–based **Genomic Health,** uses the combined expression of just 21 genes to predict both a patient's breast cancer prognosis and her likelihood of responding to chemotherapy.

Steve Shak, Genomic Health's chief medical officer, explains that for every 100 women diagnosed with node-negative, ER-positive breast cancer, 85 "will survive without recurrence" with surgery and hormonal therapy alone. Of the 15 who will recur, chemotherapy benefits four. In other words, just 4 percent of breast cancer patients both need and can benefit from cytotoxic chemotherapy.

"So because we don't know who is more likely to recur, we 'punish' the many to benefit the few," he says.

Oncotype DX uses quantitative RT-PCR to produce a "recurrence score" between 0 and 100, with the likelihood of recurrence ranging from 4 percent for a low score to 35 percent for a high one. As with MammaPrint, its gene panel arose by retrospectively mining earlier studies correlating outcome with gene expression in breast cancer.

But key to Oncotype's development was a novel method to perform RT-PCR on RNA in formalin-fixed tissues, says Shak—an advance that took Genomic Health two years to pull off, but which opened up for the company all those archived tissue blocks left over from 30 years of clinical studies in the United States and abroad.

"The RNA was there, but it was fragmented. By optimizing the assay to look at those fragments of RNA, we were able to reliably quantify it," he says.

Expect the Unexpected

The success of Oncotype DX and MammaPrint highlights the continued utility of archived cancer array datasets. But how to locate, integrate, and analyze them all? Arul Chinnaiyan, professor of pathology and urology at the **University of Michigan School of Medicine,** and former graduate student Daniel Rhodes developed what has turned out to be an unexpectedly fruitful resource for researchers looking to do just that.

Oncomine (available through **Compendia Bioscience**), is an online database of gene expression in cancer. Its 608 million datapoints represent 342 studies on 40 different disease types —24,250 arrays in total.

According to Chinnaiyan, the project arose out of "mundane" necessity. "We were doing more gene expression analyses, and were always asked by academics whether their favorite genes were dysregulated in a specific cancer, and we had to mine the datasets to find out."

They developed Oncomine to "empower the average oncologist or biologist to ask the question themselves."

But Chinnaiyan quickly realized Oncomine was a useful hypothesis-generating tool, as well. "It's like a Google for high throughput gene expression data in cancer," he says.

His lab has plumbed Oncomine to discover several potential biomarkers, including a novel chromosome translocation (typically associated with "liquid cancers") in prostate cancer, a solid tumor. The translocation produces a gene fusion between TMPRSS2 and an Ets transcription factor, empowering Ets with androgen responsiveness, and represents a unique opportunity, he says.

"The identification of a gene fusion in prostate cancer makes it a highly specific biomarker and a good therapeutic target," he says, just as the Bcr-Abl fusion is both a biomarker for chronic myelogenous leukemia, and the target of its therapy, Gleevec.

So unexpected was the observation, Chinnaiyan recalls, he initially thought it must be wrong. Now San Diego–based **Gen-Probe** is developing a urine-based detection diagnostic for prostate cancer based on the discovery, he says.

Epigenetic Biomarkers

Not all genomic biomarkers are based on gene expression. Researchers have become increasingly aware recently of the critical role that epigenetic changes—whether involving DNA methylation or histone modifications—play in both development and disease.

Most epigenetic biomarker discovery focuses on DNA hypermethylation. One of the most common events in tumor formation, it turns out, is silencing tumor suppressors and cell cycle control genes by methylating their (otherwise unmethylated or partially methylated) promoters.

Kenneth Nephew, Professor of Cellular and Integrative Physiology at **Indiana University School of Medicine,** who studies epigenetic changes as indicators of breast and ovarian cancer, says such changes represent good disease biomarkers for three reasons: they are stable, common, and involve a gain of signal.

They also are easily detected: As tumor cells die they shed their DNA into the circulation. So, researchers are looking for ways to diagnose diseases earlier using the methylated DNA they find in blood, urine, feces, and sputum.

The process is essentially the same array-based method as with single nucleotide polymorphism and gene-expression biomarker discovery, says Art Petronis, of the **Centre for Addiction and Mental Health** in Toronto. "The difference is we have to use tiling arrays."

Rather than focusing on protein-coding or polymorphic DNA loci, tiling arrays cover the entire genome (or a large fraction of it). Petronis uses two in his search for epigenetic hallmarks of complex non-Mendelian disorders: a homemade array of some 12,000 CpG islands, and Affymetrix's higher density GeneChip Human Tiling 2.0R Array Set, which spans the entire human genome with 35-bp spacing on seven arrays.

The other difference, Petronis adds, "is we have to enrich the fraction that contains a high density or low density of methylcytosine, and then interrogate that fraction."

There are basically two ways to perform those steps, one based on differential restriction enzyme sensitivity of methylated DNA, the other on bisulfite conversion of methylcytosines to uracil.

Peter Laird, director of the **University of Southern California** Epigenome Center, uses two bisulfite conversion strategies in his search for colorectal cancer classification and ovarian cancer detection biomarkers. Laird starts with **Illumina's** Infinium assay, an array-based method that probes 27,000 CpG islands in a 12-sample format. He then validates those hits using a more sensitive technique called MethyLight, which is based on **Applied Biosystems'** TaqMan quantitative PCR chemistry and the ability of primers to distinguish minute sequence variations.

"With MethyLight we can detect 0.1 percent methylated versus 99.9 percent unmethylated DNA," he says. "With a profiling technology [like an array] you couldn't detect that."

Featured Participants

Abbott Molecular
www.abbottmolecular.com

Affymetrix
www.affymetrix.com

Agendia
www.agendia.com

Agilent Technologies
www.agilent.com

Applied Biosystems
www.appliedbiosystems.com

Broad Institute of Harvard and MIT
www.broad.mit.edu

Centre for Addiction and Mental Health
www.camh.net

Cold Spring Harbor Laboratory
www.cshl.edu

Compendia Bioscience
www.compendiabio.com

Epigenomics
www.epigenomics.com

Genomic Health
www.genomichealth.com

Gen-Probe
www.gen-probe.com

Illumina
www.illumina.com

Indiana University School of Medicine
www.medicine.iu.edu

Netherlands Cancer Institute
www.onderzoekinformatie.nl/en/oi/

Nimblegen
www.nimblegen.com

Northwestern University
www.northwestern.edu

Oncomine
www.oncomine.org

Orion Genomics
www.oriongenomics.com

Sidney Kimmel Comprehensive Cancer Center
www.hopkinskimmelcancercenter.org

University of Michigan School of Medicine
www.med.umich.edu/medschool

University of Southern California
www.usc.edu

Wyeth
www.wyeth.com

According to Victor Levenson, research associate professor at **Northwestern University,** the discovery of bisulfite conversion chemistry was a "breakthrough" in epigenetic research. "Before that, it was extremely difficult to identify sequence-specific epigenetic modifications," he says.

"But the breakthrough came at a cost," he adds. First, the conversion usually destroys 85 percent to 95 percent of input DNA, creating potential bias. And PCR on the surviving DNA is difficult, both because the two strands are no longer complementary, and because the sequence has degenerated from a tetranucleotide alphabet to one with only three bases.

Instead, Levenson uses a restriction enzyme–based approach for biomarker discovery. So does biomarker firm **Epigenomics,** says Achim Plum, vice president of corporate communications. Epigenomics' process, called differential methylation hybridization (DMH), involves fragmenting genomic DNA, attaching primers to all fragment ends, and then digesting them with a methylation-sensitive enzyme. While methylated DNA will survive the treatment, he explains, unmethylated DNA will not. Thus, it is excluded from subsequent steps, in which the fragments are amplified and used to probe a custom Affymetrix array covering some 50,000 genomic fragments—mostly promoters and CpG islands.

"The beauty of the array is we put all our knowledge of what regions to look at on the array," says Plum. "So the fragments covered by probes are highly selected based on our knowledge and experience."

Plum says the company has used DMH to drive internal efforts to develop a lung cancer detection diagnostic. The process took less than a year, he says, from discovery to clinical proof-of-concept with a set of markers that could detect lung cancer with 69 percent sensitivity and 91 percent specificity in blood. Epigenomics' most advanced diagnostic, in development with **Abbott Molecular,** is a colorectal cancer–screening test using methylated Septin-9 DNA in blood as a biomarker.

Orion Genomics uses a similar strategy. Developed by **Cold Spring Harbor Laboratory's** Robert Martienssen, Orion's MethylScope assay involves shearing genomic DNA to approximately 1-kb fragments (which enables complete genome coverage), splitting that pool into two fractions, digesting one with the methylation-dependent enzyme McrBC, size-fractionating the products to select for the uncut fragments, and hybridizing the two pools to a 2.1-million oligonucleotide array from **Nimblegen.**

With the right marker, even a single region can be informative, says Orion CEO Nathan Lakey. "We found even single-locus epigenetic changes that can discriminate breast cancer tumors from nontumor tissue with 90 percent sensitivity and 96 percent specificity."

Who Needs Arrays?

Bradley Bernstein, of the **Broad Institute of Harvard and MIT,** has developed a genomewide screening approach for histone modifications that eliminates DNA arrays.

"ChIP-Seq" combines chromatin immunoprecipitation (ChIP) and Illumina's next-generation sequencing technology, acquired from Solexa. Like ChIP-on-chip, which reads ChIP data using whole-genome microarrays, ChIP-Seq provides a genomewide view of chromatin changes. The difference, says Bernstein, is that ChIP-on-chip measures fluorescence intensity, while ChIP-Seq counts the absolute number of times a particular genomic fragment is detected.

"ChIP-on-chip is an analog readout, whereas sequencing is a digital readout," he explains.

This past August, Bernstein coauthored a study that used ChIP-Seq to map several histone modifications across the genome in mouse embryonic stem cells, neural progenitors, and embryonic fibroblasts. The data shed light on the cells' transcriptional past, present, and future, Bernstein said.

"It is a way to really precisely characterize the state of the cell," he explains. "The patterns tell you about what the cell is doing, and what it can become later."

A study, published February 17 in *Nature,* describes a similar method for probing methylated DNA with single-base resolution. Called BS-Seq, the technique combines Solexa sequencing technology with bisulfite conversion; it was used to scan some 93 percent of all possible methylation sites in the *Arabidopsis* genome.

Such data will be invaluable as the National Institutes of Health's $190 million Epigenome Project, part of the so-called "Roadmap Initiative," kicks off.

"The idea is to collect genomewide maps for some defined number of primary cell types and models, in order to understand the normal epigenome," Bernstein says. "Once you do that, you can begin to look at diseased epigenomes."

Yet Bernstein admits DNA methylation will likely be more valuable for biomarker development than histone modifications, because while tumors release DNA into the circulation, they do not shed intact chromatin. Thus, histone marks can only be assayed in intact cells and tissues.

"The information content in the histones and chromatin is much richer, but it's more difficult to assay," he says.

However these and other tests play out, says Stephen Baylin of the **Sidney Kimmel Comprehensive Cancer Center** in Baltimore, Maryland, the result will be a more individualized medical experience.

Take colon cancer, for instance. "Right now you do a blood test in the stool to screen for blood," he says. "That is suggestive, but not indicative of colon cancer. But if you can get the DNA, it will be much more diagnostic." After all, why have a colonoscopy if you don't need it?

Jeffrey Perkel is a freelance writer based in Pocatello, Idaho.

DOI: 10.1126/science.opms.p0800024