

Gene-Expression Analysis Exploits More Technologies

To quantify the expression of specific genes, researchers can use a variety of techniques, including arrays, PCR, and high throughput sequencing. However, getting accurate results still depends on precisely carrying out these methods, even with increasingly user-friendly technologies. In fact, as more scientists study gene expression, the standards for analysis are growing more rigorous to ensure that only accurate data are published. Likewise, software has been keeping pace, helping researchers follow protocols and analyze their results. **By Mike May**

One of the biggest challenges for researchers in today's gene-expression analysis arises from choice. "A scientist needs to decide which technology to use—qPCR [real-time polymerase chain reaction], microarrays, or something like RNA sequencing. Those are the big ones," says Heidi Kijenski, microarray marketing director at **Agilent Technologies** in Santa Clara, California.

Even the now old-school approach to analyzing gene expression by way of arrays keeps offering new twists. For example, the higher density (four-fold higher than the company's previous technology) SurePrint G3 Gene Expression Microarrays from Agilent include one million oligonucleotide probes, which can be arranged to study two samples of 400,000 features each, four samples of 180,000 features apiece, or eight samples with 60,000 features for each. "These higher-density arrays come in catalogue or custom formats," Kijenski says. "This allows a researcher the option of customizing the content for any gene-expression project." She adds, "This includes any organism where the customer has the sequence." Moreover, it takes only weeks to get a custom array, and it doesn't cost any more than a catalogue one. Even the catalogue versions offer considerable variety. "There are 31 organisms in all," Kijenski says.

To enable even better performance, Agilent developed its new, compact SureScan Microarray Scanner. "It's a smaller footprint and provides better sensitivity and image resolution," Kijenski says.

Other companies are also pushing array technology to new capabilities. As an example, Kevin Cannon, vice president of gene expression at **Affymetrix** in Santa Clara, California, points out the company's new Human Transcriptome and Splice Junction Array. "It includes almost 7 million transcripts—99 percent coverage of the human genes," he says. "A comparative study by Stanford University researchers showed that this array was as sensitive as RNA sequencing in gene-expression profiling studies, and it's more reproducible, faster, and more cost effective."

In addition, Dara Wright, vice president of clinical applications at Affymetrix, notes that gene-expression signatures are becoming increasingly useful for clinical molecular pathology applications. The Powered by Affymetrix Program, for example, enables companies to license Affymetrix's GeneChip technology to develop and clinically validate devices based on multiplex

gene-expression signatures. Wright explains that the program's current focus is primarily on developing prognostic cancer applications, and "after years of applied research and the anticipation of clinical utility, we are now seeing tests emerge for routine [clinical] use," she says.

One such clinical application, developed on the Affymetrix gene-expression platform, is the Pathwork Tissue of Origin Test, which uses RNA extracted from a tumor to assess the expression of more than 2,000 genes. That gene-expression pattern can reveal features of the tumor, such as its metastatic potential.

MULTIPLEXING PCR

"A lot of researchers have been focusing on subsets of gene markers, or biomarkers," says Handy Yowanto, global product manager at **Beckman Coulter** in Brea, California. The company's XP-PCR Process for gene expression makes biomarker multiplexing easy, according to Yowanto.

This technology consists of four steps: total RNA isolation, reverse transcription to form cDNA, multiplex PCR amplification, and fragment separation. During the PCR step, says Yowanto, "The universal tags attached to gene-specific primers are used to amplify all of the target genes collectively." He adds, "This minimizes the primer bias typically associated with traditional multiplexed applications."

UPCOMING FEATURES

Outsourcing Lab Services—November 25

Genomics: Building Clinical Models—February 24

Toxicology: Animal-free Techniques—March 2

"A scientist needs to decide which technology to use—qPCR [real-time polymerase chain reaction], microarrays, or something like RNA sequencing. Those are the big ones."

The specificity of this technology stems from the target-specific primers along with PCR-product separation using capillary electrophoresis. "This enables down to one-base fragment resolution," explains Yowanto. In addition, he points out that researchers can "include multiple housekeeping genes as controls within a single reaction, so they don't have to worry about well-to-well or run-to-run variations."

The technology is beginning to gain traction, Yowanto notes and it's being used in a range of applications. As examples, he mentions cancer research and microbial detection.

For primer design, researchers can use Beckman Coulter's eXpress Designer software. In addition, Yowanto points out a few primer-design tips: "You should make sure that the primer is intron-spanning to minimize the interference with genomic DNA. Also, make sure primers are not designed where SNPs could occur because that might reduce the quantification." He adds, "Designing primers will get easier over time as the users get more experience with this technology in different experiments."

CPR FOR PCR

Getting accurate results from PCR demands considerable attention to detail. As early as 1997, Stephen Bustin, professor of molecular science at **Queen Mary, University of London**, realized this constraint. "It became clear to me early on that the way things were being done was not sufficient if you want to quantify the results," he says.

Specifically, Bustin says, researchers often do not include enough information in their publications about how they perform quantitative PCR. "When there is information," he adds, "it's likely to be wrong. So there's a real problem with peer-reviewed literature."

In 2009, Bustin and his colleagues published the *minimum information for publication of quantitative real-time PCR experiments* (MIQE) guidelines. Some companies quickly joined this fight. "Bio-Rad was certainly the pioneer," says Bustin. "It has put a huge amount of effort in publicizing and implementing the guidelines with road shows and so on." Some other large vendors—including Agilent Technologies, **Life Technologies**, and **Thermo Fisher Scientific**—also provide MIQE-compliant assays. Bustin also points out that qbase^{PLUS} from **Biogazelle** in Zwijnaarde, Belgium, is "the only qPCR analysis software that complies with MIQE."

According to Frank Bizouarn, field application specialist in the **Bio-Rad** gene expression division, "qPCR is the tool to analyze gene expression, but you need diligence in how you set up and run your experiments." For sample preparation, he notes "Bio-Rad's Experion automated electrophoresis station determines if the RNA quality is good or bad for gene-expression analysis." He adds that Bio-Rad sponsored the first MIQE qPCR iPad app, developed by Michael W. Pfaffl of the **Technical University of Munich** and Afif Abdel Nour of the **Polytechnic Institute of LaSalle Beauvais**. "With this," says Bizouarn, "you can check off MIQE guidelines as you go along."

EXAMINING SINGLE-CELL EXPRESSION

"The population average is a lie," says Marc Unger, chief scientific officer at **Fluidigm** in South San Francisco, California. "For gene expression, the population average of any given gene is not necessarily characteristic of any of the cells in the population." For example, the histogram of gene expression of a particular gene's intensity might show two peaks from two different

FEATURED PARTICIPANTS

Agilent Technologies
www.agilent.com

Affymetrix
www.affymetrix.com

Beckman Coulter
www.beckmancoulter.com

Biogazelle
www.biogazelle.com

Bio-Rad
www.bio-rad.com

Fluidigm
www.fluidigm.com

Illumina
www.illumina.com

Life Technologies
www.lifetechnologies.com

Nanostring Technologies
www.nanostring.com

Polytechnic Institute of LaSalle Beauvais
www.lasalle-beauvais.fr/spip.php?lang=en

Queen Mary, University of London
www.qmul.ac.uk

SAS
www.sas.com

Technical University of Munich
www.tum.de

Thermo Fisher Scientific
www.thermofisher.com

groups of cells. The average—in the middle—would not represent the expression in any of the cells.

Instead of assessing the expression at a population level, some answers can only be found in single cells. "In lots of biological processes, single cells may be driving or dominating the behavior," says Unger. As an example, he mentions cancer stem cells.

In looking at single cells, though, Unger encourages researchers to study multiple genes. "If you're looking at only a few genes at a time," he says, "then you will miss patterns in the expression." He points out that the Fluidigm BioMark HD System for real-time PCR plus its Dynamic Array chips can simultaneously track 96 genes in 96 cells.

Not just any technology provides enough resolution to work with single cells. "By their very nature, single cells are tiny, which means there's not much material to work with," says Ken Livak, senior scientific fellow at Fluidigm. The microfluidic technology underlying the BioMark HD can actually control cells one by one.

To see how different cells vary in expression, though, a researcher must analyze many cells. In just three hours, according to Livak, the BioMark HD can analyze 96 cells, which can produce as many as 9,216 data points (from 96 genes in 96 cells).

EXPRESSION BY COUNTING

"All of the methods in the past that looked at gene expression were relatively indirect—measuring expression levels by labeling transcripts and hybridizing them to an array," says Gary Schroth, distinguished scientist at **Illumina**, headquartered in San Diego, California. "The trend is very clear that the world is moving more and more towards sequencing or counting assays." He adds, "RNA sequencing is more about counting and directly characterizing the molecules in your biological sample."

In January 2007, according to Schroth, his group **continued** »



“There’s lots of interest in noncoding RNAs, microRNAs.”

was the first to do what researchers now call RNA-seq, which is sequencing full-length cDNA with short-read technology. In just a few years, this technology contributed to many publications. As Schroth says, “Now, there are more than 300 peer-reviewed publications using RNA-seq methods.” He also notes the increasing throughput, going from 40–60 million reads per run in 2007 to 3 billion with today’s Illumina HiSeq 2000 system. The length of the reads has also tripled in that time—increasing from 35 to 100 base pairs. In addition most of today’s studies use paired-end sequencing, which was not available in 2007.

Ongoing releases of kits also make RNA sequencing more appealing. At the end of 2010, for example, Illumina released its TruSeq RNA Sample Prep Kits. Currently, these kits allow 24 samples to be indexed per sequencing lane, which enables researchers to multiplex samples and capitalize on the increased throughput of the sequencing systems.

“These prep kits allow labs to process dozens or even hundreds of samples a week,” says Schroth, adding that labs with very high throughput needs can even automate the sample preparation using the kits.

Other technologies also use counting. For example, the nCounter platform from **NanoString Technologies** in Seattle, Washington, detects up to 800 target molecules in a single-tube reaction for analyzing gene expression, micro RNA (miRNA), and copy-number variation. “It’s a direct digital counting method,” says Chris Grimley, vice president of marketing at NanoString. Barcode-labeled probes hybridize directly to a target molecule in solution. The reporter probe carries the signal, and the capture probe allows the complex to be immobilized for data collection.

Researchers can select from a range of existing panels, including the recently launched leukemia panel, which simultaneously detects and quantifies 25 fusion-gene isoforms and 23 additional mRNAs related to leukemia. Alternatively scientists can design custom panels using their specific targets of interest.

The nCounter Prep Station processes the samples and the nCounter Digital Analyzer uses the fluorescent reporter on the

probe to count the target molecules. “The workflow is simple and results are so reproducible that researchers choose not to run replicates,” say Grimley. “This performance combined with the fact that the technology is compatible with a variety of sample input types, such as [formalin-fixed, paraffin-embedded], is resulting in strong adoption by the oncology research community.”

COUNTING NONCODING RNA

“There’s lots of interest in noncoding RNAs, microRNAs,” says Dennis Fantin, product management leader for qPCR at Life Technologies in Carlsbad, California. In analyzing gene expression, results show that miRNA plays a regulatory role. “They don’t work through proteins, but interact directly with mRNA to inhibit translation,” says Fantin. “You can’t explain everything about gene expression with transcription alone.”

To help researchers study the gene-expression impact of miRNA, Life Technologies developed its TaqMan OpenArray MicroRNA Panels. Fantin adds that the company offers “predefined TaqMan, noncoding-RNA assays, which work just like TaqMan gene-expression assays.” For researchers who find new miRNAs and long noncoding RNAs and need an assay, Life Technologies offers a custom service.

Scientists can also use the nCounter miRNA Expression Assay Kits. “The human miRNA product includes more than 700 miRNAs,” says Grimley. The company also makes panels for mouse and rat miRNA. The study of miRNAs is becoming so popular that the company just launched its new nCounter miRGE assays, which enable the simultaneous detection and quantification of mRNAs and miRNAs in a single tube.

DEALING WITH EXPRESSION DATA

Datasets generated in gene-expression experiments keep getting bigger. “This is challenging when trying to analyze a whole dataset,” says Shannon Conners, JMP life sciences product manager at **SAS** in Cary, North Carolina. Beyond dataset size, researchers face other analytical challenges. “People want to look at several factors at once, and they want to remove effects that might be caused by samples being taken at different sites or times,” says Conners. She adds that researchers often want to search for correlations in gene-expression data acquired on different platforms.

SAS’s JMP Genomics software provides a wide range of analytical tools. For example, Conners says, “It includes tools for the analysis of continuous intensities from array data that work with very big datasets. It also includes tools for count data from high throughput-sequencing studies.” She adds that SAS tailors these analytical tools for specific data types. “For example, there are simplified workflows for things like RNA-sequencing analysis,” she says. These tools and more come in the new JMP Genomics 5.1.

The analytical tools from SAS and the other technologies discussed here reveal the growing power that comes from combining approaches to analyze gene expression. Today’s technological combinations make it faster and easier to dig deeper into gene expression under a wider range of circumstances.

Mike May is a freelance writer and editor for science and technology.

DOI: 10.1126/science.opms.p1100059