Doing it <i>in situ</i> ...	402
...Or in solution	405
Box 1: SAGE resources	404
Box 2: The other side of SAGE: CAGE	405
Box 3: An atlas of human gene expression	407

Gene expression: one size does not fit all

Analysis of gene expression is critical to understanding the molecular underpinnings of normal and disease processes. Although in recent years the field has been flooded with microarray data, a multitude of non-array-based methods for studying the expression of genes are on the market. Laura Bonetta takes a look at them.

Think about analyzing the expression of several genes in your favorite cell or tissue, and microarrays probably come to mind. But this powerful and popular technology is not always the best suited to a given application or budget. Other approaches can provide more accurate measurements on the level of expression of a handful to a few dozen genes involved in a given signaling pathway, or give information about where and when these genes are expressed inside a cell. Like microarrays, several methods can be used to monitor the expression of thousands of transcripts at once but without any prior knowledge of the genes involved. This article surveys recent developments in non-array-based strategies for monitoring mRNA expression and the products available to researchers interested in such studies.

Mx3000P and Mx3005P thermocyclers. (Courtesy of Stratagene.)

Multiplex PCR

Microarray studies generate huge volumes of data, in many cases leading to the discovery of a subset of genes, numbering a few to a few hundred, that provides information about a particular biological system or pathway. PCR is one of the technologies poised to turn microarray discoveries into more in-depth, targeted studies. PCR-based procedures typically start with the reverse transcription of mRNA in a sample to produce cDNA. Using the cDNA as template, thermostable *Taq* DNA polymerase extends pairs of short single-stranded, gene-specific primers, during repeated cycles of heat denaturation, primer annealing and primer extension. With each cycle, the amount of template cDNA doubles until it reaches a plateau.

A variation of the technique, called real-time PCR, has gained popularity in recent years as it not only identifies specific transcripts, but also gives information about their

quantities¹. One of the leading providers of real-time PCR reagents, Applied Biosystems, offers over 400,000 TaqMan gene expression assays for human, mouse and rat genes. The assays include gene-specific primers and TaqMan probes that hybridize to an amplicon as it is produced, to release a fluorescent signal. The signal is detected by a thermocycler, an instrument that monitors the accumulation of PCR product in real time and automatically analyzes the data. Real-time PCR instruments vary in sample capacity, fluorescent dye flexibility, speed and, of course, price. For example, Stratagene's Mx3000P is a low-cost, personal thermocycler that can detect up to four different colors; the company's Mx3005P, a more advanced option, detects up to five colors. Other suppliers of real-time PCR instruments include Roche, Applied Biosystems, Bio-Rad and Cepheid.

The fact that these instruments can detect and resolve several different colors

means, in principle, that more than one PCR—using probes labeled with different-colored dyes—can be carried out in a single tube, reducing reagent costs and increasing throughput. But so-called multiplexed PCR remains a challenge. “If a reaction tube contains one template that is much more abundant than the other templates in that tube, it will amplify to a very high copy number very early in the reaction. This will deplete the reaction components and can potentially force all the reactions into plateau phase before the lower-abundance targets have amplified sufficiently to be detected,” says Rena McClory, senior director for product marketing and instrumentation at Stratagene. Another challenge has to do with primer design. “When you have many different primers and probes in the same tube, they may interact with each other,” she adds.

To address this problem, Qiagen has developed a multiplex PCR kit that contains

preoptimized concentrations of HotStarTaq DNA polymerase and an innovative PCR buffer that promotes 'molecular crowding', bringing primers in closer proximity to their target template. According to the company, in this way, each amplification reaction proceeds with specificity and efficiency, even when several different primer sets are present in a single reaction tube. "It is an open platform for multiplexing PCR. You can use any types of probes, such as TaqMan or molecular beacons, and detection methods," says Kenneth Dwyer, a marketing manager at Qiagen. According to Dwyer, the product allows a researcher to run up to 400 reactions in a single 96-well plate.

Invitrogen has come up with a different solution: the LUX (light upon eXtension) gene expression assay, which does not require labeled probes, thereby reducing the complexity of the reaction. The LUX system uses two gene-specific primers, each about 20–30 bases in length, one of which is labeled with a single fluorophore near the 3' end. When the labeled primer is incorporated into the double-stranded PCR product, the fluorophore gives off a fluorescent signal. "While with TaqMan assays the fluorescence is released as the probe detaches from the product, in LUX the fluorescence remains part of the product," says Peter Welch, director of research and development at Invitrogen. This property of the assay is designed to permit the analysis of more than one transcript in a single reaction tube. "We offer three different colors for labeling the primers, allowing direct detection of three different analytes, and by doing melt curve analyses you can look at five or more different analytes in the same sample," says Welch.

According to Promega, its Plexor technology allows the monitoring of up to three different genes in a single reaction tube. Like the LUX platform, Plexor reactions require only two primers for each target cDNA. Plexor, however, works by measuring a reduction, rather than increase, in fluorescence during amplification. One of the two primers contains both a fluorescent tag and a modified base. As amplification proceeds, a fluorescent quencher is inserted opposite the complementary modified base, resulting in a reduction in the fluorescent signal. Multiplex assay design is further simplified by the use of the web-based Plexor primer design program. Other providers of real-time PCR reagents include Stratagene, Clontech and Maxim Biotech.

The GenomeLab GeXP Genetic Analysis System. (Courtesy of Beckman Coulter.)

Upping the ante

Real-time PCR methods can monitor up to four or five genes per reaction, but Althea Technologies has developed an eXpress Profiling (XP) process for highly multiplexed, quantitative measurements of gene expression levels. It combines endpoint PCR and fluorescent detection with a method that uses two sets of primers to maintain gene ratios constant during the reaction. Two gene-specific primers, which carry on the 5' ends a consensus or universal sequence, are used to amplify the specific gene targets. The resulting PCR templates are tailed with the universal sequences. As the reaction proceeds, universal primers, present at high concentrations in the reaction mixture, take over the amplification process. "After the first few cycles you are only amplifying using the universal primers, which locks the gene ratios" says Joseph Monforte, vice president and chief scientific officer for Althea. The transition from using many different primers to just two effectively collapses the reaction complexity, as all the products are treated as the same chemical species and not differentially amplified.

The technology is available as a service from Althea and commercialized as Beckman Coulter's GenomeLab GeXP genetic analysis system. The system can run two 96-well plates in 24 hours to look at the expression of 20–35 genes in a single reaction per well. The system is particularly useful for pathway analysis or microarray

data validation. "Microarrays monitor the expression of thousands of genes. We see our technology play an important role after this discovery work, when you really want to look at the expression of 5 to 100 genes at a go," says Noreen Galvin, GenomeLab business manager. Beckman Coulter sells a master mix with the polymerase and universal primers, as well as gene-specific kits developed for different groups of genes, such as those involved in apoptosis or cell proliferation. "[The gene-specific kits are] something we will continue to add to and deliver. What drives the development of different kits is what our customers want," says Galvin. In addition to the PCR reagents, Beckman Coulter offers several software tools that assist in the design of experiments and management of data.

Doing it *in situ*...

Although different PCR platforms can be used to measure the quantities of different transcripts in a sample, *in situ* hybridization provides temporal and spatial information about their expression. The technique makes use of labeled nucleic acid probes that bind to specific targets in cells and can be detected with various fluorescence- or enzyme-based protocols. Until recently, it was difficult to detect more than one or two transcripts simultaneously in a single cell, but improvements in probe design, fluorescent dyes and confocal microscopes are making multiplex fluorescence *in situ* hybridization (FISH)

possible. Invitrogen's FISH Tag DNA and FISH Tag RNA kits are optimized for multiplex FISH applications. The kits include the tools needed for synthesizing, labeling and purifying a probe, and then preparing the labeled specimen for imaging.

David Kosman and colleagues at the University of California, San Diego used these probes to detect seven different transcripts in a single *Drosophila* embryo². "In our hands, we estimate that [FISH] is 5 to 10 times more sensitive than histochemistry," says Ethan Bier, one of the two senior investigators on the study. According to Bier, the technology could, in principle, be optimized to detect as many as 20–50 expressed genes at one time. "We see it as a niche application between detecting one or two genes by histochemistry and analyzing 20,000 transcripts by microarray," says Bier. "If a microarray points to 50 to 100 interesting genes, researchers might be able to do 2–3 multiplex *in situ* hybridizations to look for coordinate regulation or patterns of expression for these genes." The technology could also

BOX 1 SAGE RESOURCES

Once an experiment has generated a catalog of SAGE tags, the identity of the corresponding mRNA for each tag sequence must be determined. US National Institutes of Health have created a repository for SAGE sequence data, developed data processing algorithms and maintain the SAGEmap website (<http://www.ncbi.nlm.nih.gov/SAGE>). SAGEmap provides a convenient look-up table that links SAGE tags to mRNA transcripts. It is possible to enter a query sequence in the SAGEmap resource to determine what SAGE tags are in the sequence, then map it to associated SAGE-tag records and view the expression of those tags in different SAGE libraries. Alternatively a researcher can enter a SAGE tag sequence and find a gene match, if one exists, or view expression data for that tag.

be used to construct comprehensive gene expression atlases and lead to the development of diagnostic tools for examining gene expression in disease states.

Exiqon developed a line of locked nucleic acids (LNA) that can be integrated into different types of probes to provide superior hybridization characteristics and enhanced biostability. The probes can be applied to different applications, including FISH.

Using LNA probes, Wienholds *et al.* carried out whole-mount, *in situ* detection of the conserved vertebrate microRNAs (miRNAs; short RNA molecules known to regulate the expression of other genes and control development) in zebrafish embryos. The researchers were able to construct a catalog of miRNA expression patterns³. LNA probes are available from Proligo, Sigma-Aldrich, Isogen Life Sciences and others.

...Or in solution

Other products in the medium-throughput arena include those for solution hybridization-based methods. Luminex has developed the xMAP technology, which relies on beads carrying variable quantities of two different fluorescent dyes to produce up to 100 different shades of color. Each bead is coupled to a unique probe that recognizes a specific molecule. After the beads are mixed with a sample—for example, a cell lysate—and placed into the Luminex instrument, the unique color signature on each bead reveals the identity of the bound molecules. Several companies are licensed to sell Luminex instruments or reagents for different applications.

Panomics has combined the multiplex bead platform with branched DNA technology, which is a sandwich nucleic acid hybridization assay used to amplify a reporter signal. Unlike PCR, the Panomics product, called QuantiGene, provides a tool for analyzing gene expression directly from cell lysates. RNA transcripts are released from cells in the presence of a lysis mixture and hybridized to a set of probes specific to different genes. The RNA-probe complexes are captured by their respective beads. The signal on each bead is then amplified by hybridizing the branched DNA and a biotinylated labeled probe, which binds to a fluorescent dye. The captured beads are then analyzed with the Luminex instrument; the identity of each bead reveals the identity of the transcript and the amount of fluorescent signal reveals its quantity. “There is no RNA clean up and no amplification of the transcript. We created an amplification solution on the signal rather than the

target,” says Ian Ley, vice president for marketing at Panomics. Probe sets and capture beads for different genes are available from Panomics; if a panel required for research is not yet available, the company says it will manufacture it.

Another in-solution hybridization protocol for the multiplex, quantitative analysis of RNA transcription has been developed by the VTT Technical Research Center of Finland. The technique, called transcript analysis with the aid of affinity capture (TRAC), is based on the hybridization of mRNA with multiple fluorophore-labeled probes of distinct sizes. After hybridization the mRNA-probe complexes are captured, and the probes are eluted and analyzed by capillary electrophoresis. The probe signal intensities correspond to the amount of RNA in the sample, and the probe size indicates the transcript identity. “In many methods you have to turn RNA to cDNA first. This step can destroy the experiment or introduce errors,” says Hans Söderlund, research director at VTT. The whole procedure, starting from sample collection, can be carried out in 2 hours, making the assay suitable for high-throughput analysis of a limited set of mRNAs, such as gene expression monitoring in microorganisms⁴.

Sequencing-based analysis

Like microarrays, serial analysis of gene expression (SAGE) monitors the expression patterns of thousands of genes in one sample. But microarrays are limited to measuring the expression of previously identified genes assigned to the array. SAGE, however, is an open platform that can lead to the discovery of new genes. A SAGE protocol

BOX 2 THE OTHER SIDE OF SAGE: CAGE

A variation of the SAGE technique, called cap analysis of gene expression (CAGE)⁸, uses tags corresponding to the 5' ends of transcripts. To generate a CAGE library, linkers are attached to the 5' ends of full-length enriched cDNAs to introduce a recognition site for the restriction endonuclease *MmeI*. The enzyme cleaves cDNAs at a sequence 20 and 18 nucleotides away from its recognition site, creating a two-base overhang. After amplification the tags are concatenated for sequencing. The technology relies on the generation of full-length cDNAs by the cap-trapping technology, developed at the RIKEN institute. “The first step is more complicated,” says Piero Carninci at the Genome Science Laboratory at RIKEN. “But in the end the technique gives you more information than SAGE.” Indeed, CAGE not only monitors the expression levels of different genes but also yields information about the location of transcriptional start sites and helps to identify promoter sequences. “We wanted to be able to combine the analysis of transcriptional networks with transcriptional profiling,” says Carninci, who developed the technique. The company DNAFORM provides full-length cDNA libraries and is in the process of developing kits for CAGE library construction.

requires the conversion of mRNA from a cell or tissue to double stranded cDNA that carries a biotin label at the 3' end. The cDNA is digested and the products are captured using the biotin label. A *BsmFI* restriction site adapter is ligated to these fragments. The restriction enzyme then cleaves DNA at a position approximately 14 base pairs from its recognition site, generating

14-base-long SAGE tags corresponding to each transcript in the original sample. These tags are ligated together to form concatemers, which are cloned into a standard plasmid vector and sequenced. Each clone can contain tags for up to 60 mRNA transcripts. It is thus possible to sequence many thousands of such tags from a cell specimen to obtain an accurate quantitative analysis

of the relative levels of the genes expressed.

As the use of SAGE has increased in recent years, it has been possible to combine data from different experiments and sites to produce large-scale analyses of genes expressed in a particular species (**Box 1**). A new method⁵, dubbed LongSAGE, uses 21-base-long tags, which makes the discrimination between genes and the identification of new genes easier. The downside is that "you can end up with approximately 30% more sequencing," says Dave Peters of the Department of Pharmacology at the University of Liverpool. Another method uses tags corresponding to the 5' ends of transcripts (**Box 2**).

SAGE was invented by a group of researchers led by Kenneth Kinzler and Bert Vogelstein of The Johns Hopkins University⁶. The commercial application of the technology is exclusively licensed to Genzyme Oncology, which provides a service in SAGE library construction, sequencing and analysis. SAGE is also offered under sublicense through I-SAGE and I-SAGE Long kits available from Invitrogen, which use standard molecular biology techniques to construct SAGE libraries. One of the main bottlenecks in terms of time and cost for SAGE procedures is the sequencing step. Although advances in sequencing technologies may remove this hurdle in the next few years, for now several companies, such as Seqwright, Agencourt Bioscience and MWG Biotech, provide a SAGE sequencing as a service. Agencourt gives Invitrogen's I-SAGE library construction kit customers special pricing for SAGE sequencing services.

Massively parallel signature sequencing (MPSS)⁷ is another technique that relies on the production of short tags proximal to the 3' end of transcripts, but operates on a larger scale than SAGE. Whereas about 50,000 tags are generated within a single SAGE experiment, MPSS can yield over one million. In principle, this number is sufficient to

The newly launched Genome Analysis System will soon be commercially available. (Courtesy of Solexa.)

provide very deep coverage of the transcripts expressed in a human cell and thus, characterize even those expressed at very low levels (**Box 3**).

Solexa has developed an alternative, powerful platform through its Genome Analysis System, which will be commercially available to researchers by mid-2006. After constructing a library of 20-base-long cDNA tags corresponding to each expressed transcript, millions of individual DNA molecules are linked, using short oligonucleotide adapters, to the surface of a glass flow cell. Each molecule then undergoes a solid-phase amplification procedure that creates a cluster of DNA molecules each with identical DNA tags. The DNA molecules in each cluster are simultaneously sequenced by Solexa's Sequencing-by-Synthesis chemistry, which uses proprietary fluorescently labeled modified nucleotides. The sequence information is then used to identify the corresponding transcripts and genes, and determine their abundance. "You can use our system for both profiling known genes and

BOX 3 AN ATLAS OF HUMAN GENE EXPRESSION

A recent study⁹ used MPSS to analyze pooled RNA samples isolated from 32 human tissues to document the patterns of expression of 18,667 genes. Only about 34% of the sequence tags collected could be mapped to transcripts, suggesting that more than 65% of transcribed sequences are yet to be characterized. In addition, the data demonstrated that differences between cell and tissue types are largely determined by the expression of a limited number of tissue-specific genes. The data available on the web (<http://mpss.licr.org> and <http://sgb.lynxgen.com>) can be used to identify tissue-specific genes or global changes in gene expression induced by pathological conditions.

discovering unknown ones," says Gary P. Schroth, director for expression applications research and development. "Solexa's new technology will allow researchers to interrogate up to 5,000,000 tags per sample for a few hundred dollars—costs that are a fraction of those of MPSS and SAGE, and on par with those of hybridization arrays."

Laura Bonetta is a freelance writer based in the Washington, DC area (lbbonetta@nasw.org).

1. Bonetta, L. *Nat. Methods* **2**, 305–312 (2005).
2. Kosman, D. *et al. Science* **305**, 846 (2004).
3. Wienholds, E. *et al. Science* **309**, 310–311 (2005).
4. Rautio, J.J. *et al. J. Microbiol. Methods*; published online 27 September 2005.
5. Saha, S. *et al. Nat. Biotechnology* **19**, 508–512 (2002).
6. Velculescu, V.E. *et al. Science* **270**, 484–487 (1995).
7. Brenner, S. *et al. Nat. Biotechnology* **18**, 630–634 (2000).
8. Kodzius, R. *et al. Nat. Methods* **3**, 211–222 (2006).
9. Jongeneel, C.V. *et al. Genome Res.* **15**, 1007–1014 (2005).

SUPPLIERS GUIDE: COMPANIES OFFERING SYSTEMS AND REAGENTS FOR ANALYSIS OF GENE EXPRESSION

Company	Web address
Abgene	http://www.abgene.com/
Active Motif	http://www.activemotif.com/
Agilent Technologies	http://www.home.agilent.com/
AlleLogic Biosciences Corp.	http://www.allelogic.com/
Althea Technologies	http://altheatech.com
Ambion	http://www.ambion.com/
Apogent	http://www.apogent.com
Applied Biosystems	http://www.appliedbiosystems.com/
Artus Biotech	http://www.artus-biotech2.de/
Beckman Coulter	http://www.beckman.com
Biochain	http://www.biochain.com/biochain/
Biogene	http://www.biogene.com/welcome.cfm/
BioGenex	http://www.biogenex.com/
Biognostik	http://www.biognostik.de/
BioHelix Corp. (NEB)	http://www.biohelix.com
Bioneer	http://www.bioneer.com/
Bio-Rad Laboratories	http://www.bio-rad.com/
Biosearch Technologies	http://www.biosearchtech.com
Biosource	http://www.biosource.com/
Biotrove	http://www.biotrove.com/
Bio S&T	http://www.biost.com/
Brinkmann Instruments	http://www.brinkmann.com/
Chemicon International	http://www.chemicon.com/
Cepheid	http://www.cepheid.com/
Clontech (Takara Bio)	http://www.clontech.com
Corbett Research	http://www.corbettresearch.com/index2.html
DxS Limited	http://www.dxsgenotyping.com/
Enzo Life Sciences	https://www.enzolifesciences.com/
Epicentre	http://www.epicentre.com/
Epoch Biosciences	http://www.epochbio.com/
Eurogentec	http://www.eurogentec.be/
Exiqon	http://www.probelibrary.com/
Fermentas Inc.	http://www.fermentas.com/
Finnzymes	http://www.finnzymes.com/
Fluorescentric	http://www.fluorescentric.com/
Genetic Research Instruments	http://www.gri.co.uk/
Genetix	http://www.genetix.com/
Gene Express	http://www.geneexpressinc.com/
Gene Therapy Systems, Inc.	http://www.genetherapysystems.com/
GenScript	http://www.genscript.com/
GenoMechanix	http://www.geno-mechanix.com/
GeneWorks	http://www.geneworks.com.au/
Genzyme Corporation	http://www.genzyme.com/
Idaho Technology Inc.	http://www.idahotech.com/
Integrated DNA Technologies	http://www.idtdna.com/
Invitrogen	http://www.invitrogen.com/
DNAform	http://www.dnaform.jp/
Lark Technologies	http://www.lark.com/
Lucigen	http://www.lucigen.com/
Luminex	http://www.luminexcorp.com/
Maxim Biotech Inc.	http://www.maximbio.com/
MJ Research (Bio-Rad)	http://www.mjr.com/

SUPPLIERS GUIDE continued

Company	Web address
Midwest Scientific	http://www.midsci.com/
MobiTec Molecular Biotechnologies	http://www.mobitec.de/us/index.html
Molecular Probes (Invitrogen)	http://www.probes.com/
MWG Biotech	http://www.mwg-biotech.com/
New England BioLabs	http://www.neb.com/
Novagen (EMD Biosciences)	http://www.emdbiosciences.com/
Origene	http://www.origene.com/geneexpression/
Panomics	http://www.panomics.com/
PGC Scientifics	http://www.pgcspecifics.com/
PerkinElmer	http://las.perkinelmer.com/
Premier Biosoft International	http://www.premierbiosoft.com
Promega	http://www.promega.com/
Qiagen	http://www1.qiagen.com/
Quantum Dot (Invitrogen)	http://qdots.com
Roche Applied Science	http://www.roche-applied-science.com/
Seegene	http://www.seegene.com/
Serologicals Corporation	http://www.intergenico.com/
SeqWright	www.seqwright.com
Sigma-Aldrich	http://www.sigmaaldrich.com/
Sigma-Genosys	http://www.sigma-genosys.com/
Solexa	http://www.solexa.com/
Stratagene	http://www.stratagene.com/
SuperArray Bioscience Corp.	http://www.superarray.com/
Synthegen	http://synthegen.com
TagSorter	http://tagsorter.com
Takara Mirus Bio	http://www.takaramirusbio.com/
Thermo Electron Corporation	http://www.thermo.com/
Techne	http://www.techne.com/
USB Corp.	http://www.usbweb.com/
VTT	http://www.vtt.fi/