Proper prep	688
Shorter is harder	689
A common reference	689
Box 2: Sequencing for discovery	689
Box 3: Microarrays and more	691

MicroRNA profiling: separating signal from noise

Monya Baker

Various platforms for measuring microRNAs can provide different answers.

MicroRNAs may be small, but these noncoding RNAs that regulate gene expression are creating a big stir. Finding differences in the expression of microRNAs between, say, healthy and diseased cells could potentially be used to diagnose diseases or to assess treatment effects. If researchers can understand how they work, microRNAs could provide tools for manipulating genes, not to mention help to untangle how genes are regulated.

At first glance, studying microRNAs seems more manageable than studying the menagerie of other types of RNA. Typical expression profiling experiments for protein-coding genes examine thousands of molecules; those for microRNAs examine hundreds. But researchers are still figuring out the most reliable ways to measure these important molecules.

The most common techniques for profiling microRNAs are deep sequencing, microarrays and quantitative real-time PCR (qPCR). All are supported by several commercial offerings (see **Boxes 1–3**). Though specific products and techniques vary, researchers generally agree on the relative strengths and weaknesses of the platforms. The best choice depends on the application, says Muneesh Tewari, who studies microRNAs at the Fred Hutchinson Cancer Research Center. "It's a balance of cost, precision, accuracy and sample quantity," he says. "If the purpose is to screen a bunch of samples to find a few microRNAs that change and you can tolerate a false negative, then the microarray may be the best platform. If the purpose is to detect microRNAs where the sample amount is limiting, then qPCR has better sensitivity, and if you are trying to see different isoforms or very similar microRNAs, then sequencing is going to be the best approach."

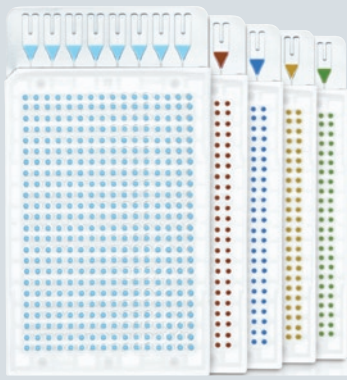
But not all researchers are aware of how the choice of product influences the data. "If you take the same sample and analyze microRNAs in different profiling technologies, the overlap can be surprisingly poor," says Robert Blelloch, a stem cell biologist at the University of California, San Francisco. "It's a really murky world," he says. "The community has to come together to come up with a strategy here."

Uncertain profile

Scientists at the Cancer Research UK Cambridge Research Institute and the

BOX 1 qPCR PROFILING

"The fastest growing field within microRNA is qPCR. There's no doubt," says Peter Mouritzen, director of life science product development at Exiqon. He should know; besides qPCR, his company, which offers modified nucleic acids designed to boost specificity and sensitivity for complementary oligonucleotides, offers microRNA products for *in situ* hybridization, microarrays, northern blotting and more. Several manufacturers, including Applied Biosystems, Exiqon, Fluidigm and SA Biosystems, offer qPCR kits that can assess hundreds of microRNAs in parallel, and some offer customizable assays.



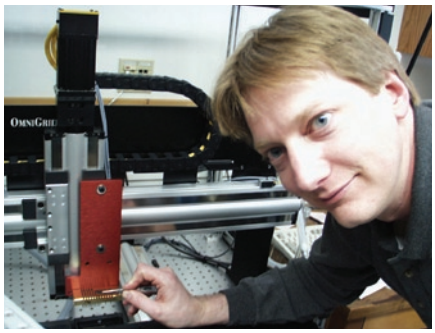
Life Technologies

A qPCR card used for profiling microRNA.

qPCR is a well established way to determine transcript levels. RNA is converted to cDNA, then run through amplifying cycles for each transcript being studied. Depending on the manufacturer, either double-stranded DNA or primers are labeled, and the number of PCR cycles necessary to produce a given signal is used to assess against a standard curve the number of transcripts originally present.

MicroRNAs pose two chief challenges for qPCR experiments: specificity and normalization. As microRNAs are single stranded, techniques that amplify these molecules usually use one unique and one universal primer, which provides less specificity than methods that use two unique primers. The uniform length of microRNAs also obscures unwanted side reactions because differences between amplified product lengths cannot be used as a quality check.

More problematic for qPCR are the techniques used to normalize microRNA expression. An ideal solution would be to identify reference microRNA genes, but none are available, write Vladimir Benes and Mirco Castoldi from the European Molecular Biology Laboratory in Heidelberg[4]. Instead, most normalization curves rely on genes for small RNAs that might not be transcribed by the same polymerases as microRNA precursors and are less representative of general microRNA regulation. An alternative normalization technique uses the mean microRNA present in each sample. The best option is still unknown, says Don Baldwin of the University of Pennsylvania Molecular Profiling Facility. "You just have to hope that what you've selected as a reference holds constant across your samples."



Don Baldwin of the University of Pennsylvania Molecular Profiling Facility believes a synthetic reference library of microRNAs can help to develop clinical assays and to make profiling more accurate.

European Bioinformatics Institute in Cambridge recently assessed how well deep sequencing, microarrays from six manufacturers, and two forms of qPCR identified differences in microRNA amounts among three biological samples[1]. One analysis examined transcripts that were upregulated in a breast progenitor cell line compared with normal breast tissue: 136 microRNAs were identified in total, but only 53 were found in common by five assays. (Results from qPCR and microarrays from two manufacturers were excluded because of their high rates of false calls.)

"I don't think any of the platforms demonstrate a significantly better view on the absolute truth," says Don Baldwin of the University of Pennsylvania Molecular Profiling Facility, who cochairs a research group with the Association of Biomolecular Resource Facilities that recently compared four microarrays and two sequencing platforms commonly used to profile microRNAs (http://www.abrf.org/ResearchGroups/Microarray/Activities/R7_Baldwin.pdf; **Table 1**). Baldwin advises researchers to find the technologies that they or their core facilities can use with the least technical variance, then stick to that and pour their energies into planning their study. "The microRNA fraction is less complex," he says, "but that doesn't mean you can get away with a poor experimental design or fewer replicates."

The obvious solution is to verify results using different techniques, but that kind of cross-validation does not always happen, says Carlo Croce, director of human cancer genetics at Ohio State University. "Most people use one method and that, I think, is wrong." He emphasizes that platforms are far from the only cause of variable results. Getting the right sample is crucial, and some samples and questions are more suited for different techniques. Still, validating findings takes significant effort, and researchers could be more efficient if they understood patterns of false positives and false negatives across different platforms.

Proper prep

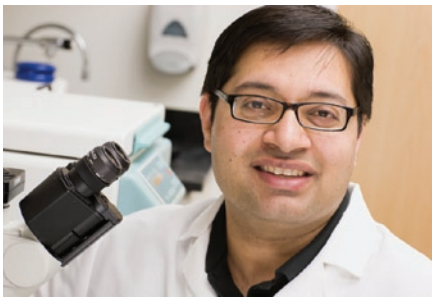
Before launching any kind of profiling study, researchers need to assess the quality of their RNA sample, says Kelli Bramlett, senior manager of research and development with Ambion, a division of Applied Biosystems. To test sequencing applications, she recommends spike-in controls generated from the External RNA Control Consortium plasmids, an effort coordinated by the National Institutes of Standards and Technology. Analysis on an RNA gel or Agilent Bioanalyzer can assess whether RNA is too degraded to be used for a particular experiment. Getting rid of unwanted RNA can also be useful. Illumina recently introduced an enzyme designed specifically to clear out ribosomal RNA, which can comprise more than 99% of total RNA in a sample.

Some technologies require an amplification step, although University of Pennsylvania's Baldwin says that it is difficult to represent all microRNAs in a sample while preserving their relative abundance. The amplification step is unavoidable in current deep sequencing protocols, but microarrays often skip amplification in favor of directly labeling microRNA.

This labeling step introduces much of the variability, says Anna Git of the Cambridge Research Institute, who co-led a comparison of profiling techniques. The detection components of various platforms often work well, she says. "The main problem is that the methods we use to label RNA are imperfect." Each company's preparation treats some RNA molecules differently and so creates different artifacts. Worse, such biases tend to be more serious in degraded samples.

L'Oreal

Anna Git of the Cambridge Research Institute believes differential labeling of microRNAs causes much experimental variation. The field has too few controls, she says.



Muneesh Tewari of the Fred Hutchinson Cancer Research Center thinks profiling technologies for microRNAs are more different from each other than are those for profiling mRNAs.

Many techniques add oligonucleotides to microRNA transcripts using RNA ligase, but that enzyme favors certain sequences over others. Enzymes that add other labels also have preferences. Deep sequencing is prey to similar biases; some microRNA sequences are preferentially ligated and amplified depending on the preparation technique[2].

Not everyone is convinced that these differences pose a big problem for expression profiling. It is true that each method has its own peculiarities that will affect quantification, and some microRNAs survive some preparation techniques better than others, says David Bartel of the Whitehead Institute, but that should not make a huge difference in the results. "Usually those biases are going to be the same in the different samples, so if you are looking at the same microRNA in different samples you can still see if it has changed," he says.

Shorter is harder

Technologies for studying microRNAs have been adapted from techniques for studying DNA and RNA molecules that are hundreds or thousands of nucleotides long. MicroRNAs are much smaller, typically about 22 nucleotides long. The short length gives researchers fewer options for designing complementary sequences: the entire microRNA sequence is often used for a single probe on a microarray.

MicroRNAs also exist in families in which members frequently vary by as little as a single nucleotide, and so are hard to distinguish. One solution is to boost the specificity of a probe or primer for its target with high temperatures, so that only the best matches bind. Genome-wide, however, microRNAs vary greatly in their GC content, or the percentage of their sequence that comprises guanines and cytosines. This means that the temperature at which microRNAs dissociate from complementary sequences varies greatly, perhaps by more than 20 °C, complicating efforts that depend on the separation and reannealing of complementary sequences.

Sequencing experiments have found variation even within microRNAs encoded by the same gene. Some single-nucleotide polymorphisms occur within microRNAs, and these variants are linked to differences in the expression of protein-coding genes. Another source of variation, post-transcriptional modification, can be identified through sequencing but complicates profiling by other techniques. "Not only does this bear significant implications on the function of the resulting microRNA," says Git, "but it also introduces a mismatch between probes [which are designed against the genomic sequence] and the real edited target, resulting in an inaccurate readout of expression."

To make matters worse, kits and algorithms designed from genome-scanning algorithms or even from miRBase, the common repository for microRNA sequences, could very well be designed to find molecules that are not transcribed or functional. "Papers have been written about imaginary microRNAs," says Bartel, who believes that more than a quarter of mouse microRNAs deposited in miRBase may not really be microRNAs.

Bartel and colleagues sequenced 60 million small RNA molecules from a wide variety of mouse tissues and found that some 150 miRBase microRNAs were either not represented or likely to be artifacts[3]. (The study also found 108 genes not represented in miRBase that did seem to represent microRNAs and showed that representative transcripts were recognized by microRNA-processing enzymes.) The bright side, says Bartel, is that these false microRNAs could be used as negative controls that indicate what researchers should expect if a microRNA is not present.

A common reference

Git advises researchers to use many more controls than they think they need. For

BOX 2 SEQUENCING FOR DISCOVERY

Illumina

A sequencer used for profiling microRNA.

Late last year, Illumina discontinued its microRNA microarrays, the better to focus on sequencing. Cost should no longer keep researchers away from deep sequencing, says Shawn Baker, market manager of expression and epigenetics at Illumina. Assuming two hundred samples per sequencing run, he says, Illumina's latest machine can provide profiles at under $200 per sample, comparable to that of microarrays. Besides Illumina, other sequencing companies including Roche 454, Applied Biosystems and Helicos offer tools for studying microRNAs or have announced plans to do so.

The appeal of sequencing is that researchers can look for all RNAs in a sample, not just the ones printed on a chip, says Baker. "You're not having to design probes for things, you're just grabbing all the RNA that's there and looking at it." The converse is that sequencing machines will also find degraded RNA and small RNAs that do not function as microRNAs, so profiling results are only as good as the algorithms that are used; many scientists feel the necessary bioinformatics are still too complicated.

Still, sequencing is probably the only reliable way to identify certain known variants, says Alex Ebhardt of the Swiss Federal Institute of Technology in Zurich, who has identified isomiRs, products of the same microRNA gene that differ in length by one or two nucleotides and may be involved in cancer. But, he worries about biases from the necessary preparative technologies. The adaptive regions and PCR-priming regions that are added to microRNAs to enable sequencing are longer than the microRNAs themselves, he says. "Single-molecule sequencing without RNA ligase and without reverse transcription PCR would really be nice."

Table 1 | Platform comparison for microRNA profiling

	qPCR	Microarray	Sequencing
Throughput time	~6 hours	~2 days	1–2 weeks
Total RNA required	500 ng	100–1,000 ng	500–5,000 ng
Estimated cost per sample, including reagents and supplies	$400 (754 human microRNAs queried per sample)	$250–$350 (at least 950 microRNAs queried per sample)	$1,000–$1,300 (theoretically, all microRNAs queried per sample)
Dynamic range detected	Six orders of magnitude	Four orders of magnitude	Five or more orders of magnitude
Infrastructure and technical requirements	Few	Moderate	Substantial

Results reported by the Association of Biomolecular Resource Facilities. Newer protocols and equipment may have different prices, throughput, output and requirements.

example, to validate microRNAs whose expression changes between different types of sample, researchers need to identify several other microRNAs that can be used for comparison within those same samples, including some that seem to change in opposite ways and some whose expression seems constant.

The Association of Biomolecular Resource Facilities (ABRF) hopes to help researchers discover precisely which microRNAs are favored by different techniques, says Baldwin. By the end of this year, ABRF plans to release a pool of synthetic RNA molecules that can be analyzed across microarray, qPCR and deep sequencing platforms. As the concentrations and identities in the pool will be known, says Baldwin, differences caused by the techniques themselves can be revealed.

Other researchers prefer to use biologically derived references, combining cell lines and tissue samples into a large pool that can be assessed alongside the samples of interest. Agilent, Ambion and other companies produce such universal reference pools for microRNA studies. Tewari, who used a small synthetic reference collection for assessing sequencing and qPCR, says that the technique works well, as long as scientists keep in mind how it might perform differently from a biological sample: a synthetic sample that contains a greater variety of microRNAs than typically seen in a biological sample may introduce high levels of cross-hybridization or competition, and it might also be tough to mimic the dynamic range in biological samples, particularly if some assays saturate. Most importantly, says Tewari, biological samples probably contain less than 0.01% microRNA. The rest is ribosomal RNA, tRNA and mRNA, which represents a background matrix that could compete with preparative enzymes and otherwise affect results.

But Baldwin thinks simplicity makes assessment possible. "All our tests so far have used messy biological samples, so we don't really know the absolute truth.

BOX 3 MICROARRAYS AND MORE

Though sequencing garners lots of attention, rumors of the impending death of microarrays have been greatly exaggerated, says Don Baldwin of the University of Pennsylvania Molecular Profiling Facility. "Microarray usage continues to grow, even though we have several sequencers on campus." Indeed, the need to validate sequencing data could raise demand for microarrays.

Microarrays are slides spotted with thousands of oligonucleotides used as probes, each designed to detect a particular target sequence. The use of redundant probes and slightly different probes that target the same sequence improves specificity and sensitivity, and some manufacturers, such as Agilent and Combimatrix, allow researchers to customize probes. Some researchers, such as Carlo Croce at Ohio State University, buy equipment to print their own arrays, updating them as they see fit and also including microRNA precursors that they particularly want to investigate.

Researchers purchasing microarrays should make their selection carefully, says Anna Git of the UK Cambridge Research Institute, who has compared several platforms[1]. Manufacturers often tout the number of probes on their microarrays as an indicator of quality, but there is more to it than that, she says. "Just because there are lots of features lighting up doesn't mean it's a good array. You have to be very careful for what the probes are and their specificity, not just the sheer number."

Git uses arrays from Agilent, which, according to an algorithm she developed, have one of the highest rates of true positive and true negative calls. Ambion arrays have slightly lower rates of false reads according to her analysis, but require about 100 times more RNA, she says. How and how quickly manufacturers update their microarrays and other unique features can also determine which product is the best choice for a given lab. However, Git does not recommend switching between manufacturers often. Whereas results within a platform are generally consistent, comparisons across platforms are difficult.

As well as expanding the scope of microRNAs that they evaluate, many researchers hope to expand the kinds of samples that can be analyzed readily. They are particularly interested in blood and in formalin-fixed paraffin-embedded samples, in which microRNAs are surprisingly stable. Companies working with clinical samples include Asuragen, febit and Rosetta Genomics, but rather than selling microarrays, these companies are developing diagnostics and providing pharmaceutical services. febit, which works primarily with blood, avoids the need to label microRNAs by instead labeling probes. MicroRNAs from the sample serve as primers for an extension reaction that adds labeled nucleotides to probes on an array. The technique boosts specificity because microRNAs must fully hybridize to the probe for the reaction to occur, says Andreas Keller, vice president of biomarker discovery at febit.

Researchers studying blood and paraffin samples in their own labs often use qPCR, but they should soon have more options. The next big push for microRNA profiling will not be adding new probes but moving out of cells, predicts Alicia Burt, director of microarray systems at Agilent. "Studying tissues and cell lines, that's very well covered," she says. "Being able to analyze serum and clinical samples, that's very exciting."

But all these methods still provide limited information about how microRNAs function *in situ*. "At some point you want to know not just what microRNAs are in a cubic millimeter. You want to know where they are. Are they in the blood vessels, or the collagen?" says James Manfield of Cri, which provides hardware for multiplex imaging. Studying microRNA *in situ* can get pretty complex, he explains. To make sure that the microRNA is having a biological effect, both the microRNA and the proteins believed to be regulated by it need to be labeled, and researchers need to account for autofluorescence of tissues. Exiqon has developed a suite of probes specifically for labeling microRNAs *in situ* and even *in vivo*; a variety of conventional labeling techniques are also being applied, and a host of other labeling techniques have been developed[5].

That's why a synthetic reference sample may be useful." Such samples could also be spiked with cell extracts to mimic biological samples, he says.

No matter what the ultimate approach, researchers agree that no reference sample can make up for using high-quality controls and constantly questioning the tools and protocols used for experiments. "Treat your assays and your kits with suspicion," says Git, "and we're going to end up with better science."

1. Git, A. *et al. RNA* **16**, 991–1006 (2010).
2. Linsen, S.E. *et al. Nat. Meth.* **6**, 474–476 (2009).
3. Chiang, H.R. *et al. Genes Dev.* **24**, 992–1009 (2010).
4. Benes, V. & Castoldi, M. *Methods* **50**, 244–249 (2010).
5. Hunt, E.A., Goulding, A.M. & Deo, S.K. *Anal. Biochem.* **387**, 1–12 (2009).

Monya Baker is technology editor for *Nature* and *Nature Methods* (m.baker@us.nature.com).

SUPPLIERS GUIDE: COMPANIES OFFERING MICRORNA PROFILING TECHNOLOGY	
Company	**Web address**
454 Life Sciences	http://www.454.com/
Advanced Array Technology (now Eppendorf Array Technologies)	http://www.biochipnet.com/
Affymetrix	http://www.affymetrix.com/
Agilent	http://www.agilent.com/
Ambion	http://www.ambion.com/
Amersham (GE Healthcare)	http://www.amershambiosciences.com/
Applied Biosystems (Life Technologies)	http://www.appliedbiosystems.com/
Arcturus (Molecular Devices)	http://www.moleculardevices.com/
Axygen Biosciences	http://www.axygenbio.com/
BioCat	http://www.biocat.com/
BioChain	http://www.biochain.com/
Cepheid	http://www.cepheid.com/
CombiMatrix	http://www.combimatrix.com/
Cri	http://www.cri-inc.com/
Eurofins MWG Operon	http://www.mwg-biotech.com/
Epicentre Biotechnologies	http://www.epibio.com/
Exiqon	http://www.exiqon.com/
febit	http://www.febit.com/
GeneCopoeia	http://www.genecopoeia.com/
Genisphere	http://www.genisphere.com/
GenoSensor Corporation	http://www.genosensorcorp.com/
GenScript	http://www.genscript.com/
High Throughput Genomics	http://www.htgenomics.com/
Illumina	http://www.illumina.com/
Integrated DNA Technologies	http://www.idtdna.com/
Lambda	http://www.lambda.at/
LC Sciences	http://www.lcsciences.com/
Life Technologies	http://www.lifetechnologies.com/
Luminex Corporation	http://www.luminexcorp.com/
Metrigenix (Xceed Molecular)	http://www.xceedmolecular.com/
Millipore	http://www.millipore.com
Nanogen	http://www.nanogen.com/
NanoString Technologies	http://www.nanostring.com/
National Center for Genome Resources	http://www.ncgr.org/
Ocean Ridge Biosciences	http://www.oceanridgebio.com/
Phadia Multiplexing Diagnostics	http://www.vbc-genomics.com/
Phalanx Biotech	http://www.phalanxbiotech.com/
Precision Biomarker	http://www.precisionbiomarker.com/
Qiagen (SA Biosciences)	http://www.qiagen.com/
SciGene	http://www.scigene.com/
SeqWright	http://www.seqwright.com/
System Biosciences	http://www.systembio.com/
Thermo Fisher Scientific	http://www.thermofisher.com
Vysis (Abbott Molecular)	http://www.vysis.com/
Xeotron (Invitrogen)	http://www.invitrogen.com/