

Profiling a multiplayer game	232
From discovery come tools	233
A new toolkit	234

# Small RNAs: biology's brave new world

Nathan Blow

Small RNA discovery and profiling efforts are dramatically reshaping fundamental concepts of how genes are regulated and are leading to new tools for studying gene function.

The ability of double-stranded RNA to silence genes was first described in *Caenorhabditis elegans* in 1998—a discovery that would earn Andrew Fire and Craig Mello the 2006 Nobel Prize in Physiology or Medicine. Our knowledge of the cell's small RNA world and its ability to regulate genes through silencing has greatly expanded during the short period of time since then with the identification of many RNA classes such as small interfering RNAs (siRNA) and microRNAs (miRNAs) and a class of RNA that is associated with the Piwi family of Argonaute proteins (piRNA). Although in many cases their targets and mechanisms of action are still not clear, what is becoming certain is that many of these small RNAs play large roles in basic biological processes.

As scientists enter the second decade of small RNA research, the good news is that new technologies and tools are starting to provide answers to fundamental questions surrounding small RNA distribution in cells, patterns of expression and mechanisms of action. But as the answers emerge, so does a slew of new questions.

## Charting the unexplored country

In 2006, David Bartel, professor of biology at the Massachusetts Institute of Technology and an investigator at the Whitehead Institute in Cambridge, Massachusetts, USA along with his colleagues decided to use the high-throughput pyrophosphate sequencing approach developed by 454 Life Sciences (Branford, Connecticut, USA), now a Roche company, to peer inside the small RNA world of *C. elegans*. Their five sequencing runs profiled 394,926 small RNA sequences resulting in the identification of 18 new miRNAs and several endogenous siRNAs as well as an entirely new class of noncoding piRNA in *C. elegans* called 21U-RNA<sup>1</sup>.

While greatly expanding our knowledge of small RNA composition in the worm, the work also highlighted the potential of using sequencing as a discovery tool.

"That was just a first look in *C. elegans*, and we found all sorts of stuff because up to that point small RNAs had really been the unexplored country," says Chad Nusbaum, who is co-director of the Genome Sequencing and Analysis program at the Broad Institute of MIT and Harvard in Cambridge, Massachusetts, USA and a co-author on the paper<sup>1</sup>. According to Nusbaum, the discovery of small RNAs is a challenge best suited to a sequencing approach.

Sequencing works when it comes to small RNA discovery for two main reasons. "The fact is that no prior sequence knowledge is required, and the dynamic range is almost unlimited," says Roland Wicki, director, Applied Biosystems' SOLiD Application Development at Life Technologies in Foster City, California, USA. This nearly unlimited dynamic range is the direct result of the ability to increase the number of sequencing reads to find even very rare small RNAs in cells and tissue.

Many researchers contend that the latest next-generation sequencing platforms such as Applied Biosystems' SOLiD3 and the Genome Analyzer from Illumina in San Diego are particularly well suited to small RNA discovery applications because they can produce millions of short reads, each around 50–75 base pairs long, in a relatively rapid period of time.

Shawn Baker, a senior product manager of RWA analysis at Illumina, says that with the Genome Analyzer researchers will often generate two to five million reads when they are interested in exploring a population of small RNAs. But when it comes to more detailed inspections of RNA populations,



Chad Nusbaum suspects small RNA detection and profiling experiments will migrate toward sequencing in the future.

applications that require an increased dynamic range, researchers can ramp up some of the newer platforms to generate 40 million or more reads.

Exploring the RNA world by sequencing was on the mind of many developers from the beginning. "The small RNA discovery system was one of the first applications that we created for the Genome Analyzer," notes Baker. The initial step in all sequencing approaches for small RNA discovery is to isolate RNAs of specific sizes that can be used to construct a sequencing library by attaching specific adapters on the ends of the molecules. The library of small RNA is then used as the template for very deep sequencing. Although the library creation and sequencing steps are starting to become routine with the availability of improved kits and protocols, it is the final data analysis step that can often prove most tricky for scientists.

## TECHNOLOGY FEATURE

“The shorter the small RNA sequence, the harder it is to find a unique mappable location,” explains Wicki, adding that the diversity of small RNA types, from miRNA to piRNA to siRNA, along with the extensive sequence variations of many miRNA species makes identification and classification difficult. Both Life Technologies and Illumina are providing software to researchers to aid with the mapping of these short sequences back to the genome, and academic researchers like Nikolaus Rajewsky and his colleagues in Berlin who recently described a new algorithm called miRDeep for the identification of miRNAs from deep sequencing datasets<sup>2</sup>, are also adding to the small RNA analysis options.

### Profiling a multiplayer game

While scientists look to next-generation sequencing platforms when it comes to discovery, profiling the expression patterns of small RNAs is a multiplayer game with researchers using approaches that range from bead-based assays to those using microfluidic cards. But it is the traditional microarray that remains the option of choice over sequencing, at least for the moment.

The development of high-content miRNA microarrays has exploded during the past year as scientists continued to discover more miRNAs from humans and other species, and as developers improved the performance of these short strands of nucleic acid on arrays. “Profiling using microarrays is difficult because the conventional methods that are used for label-

ing and probe design do not work well with miRNA being so small,” says Hui Wang, a research scientist at Agilent Technologies in Santa Clara, California, USA. To overcome the challenge of profiling miRNA populations with arrays, Wang and her colleagues at Agilent developed a method for direct labeling miRNA using RNA ligase under specific conditions and designed the probe sequences on the arrays based on melting temperatures to maximize specificity and sensitivity. “The philosophy behind the direct labeling approach is that we have minimal perturbation of the sample, so you measure exactly what is there without further processing steps such as amplification or size fractionation that can potentially distort the miRNA content,” she explains.

Agilent Technologies now offers their direct labeling approach with different miRNA microarrays covering humans, mice, rats and other species. Other companies including both Life Technologies and Illumina along with Exiqon in Vedbaek, Denmark and Genosensor Corporation in Tempe, Arizona, USA are also offering various miRNA microarray options for profiling purposes.

Technology advances are making microarrays more productive and robust, but economics and speed are still their biggest advantages over other technologies. “If you know what you are looking for, then arrays remain the less expensive option,” says Nusbaum. The costs of producing microarrays and associated labeling kits have declined over the past year. In fact, Illumina’s Baker says for some of



Illumina’s Genome Analyzer has been used in a number of small RNA discovery efforts.

Illumina's whole-genome microarrays and labeling reagents the cost is currently under \$100 per sample, considerably less than the cost of a single lane in a flow cell on their Genome Analyzer.

Although the low consumable cost presents a substantial incentive to many researchers, time savings can be an additional benefit of arrays over sequencing. Joshua Levin, a research scientist at the Broad Institute, recently studied the time required for a typical next-generation sequencing project. Levin says that from RNA isolation to Illumina cDNA library construction takes around ten working days. From there, the actual sequencing can take anywhere from 3 to 4 days, in the case of unpaired sequencing reads, to a week or 8 days when obtaining paired reads. Processing of sequencing data requires another day, followed by analysis, which can often take at least a couple more days even if the researchers know the exact analyses they intend to perform. After considering all the steps, using sequencing to profile a sample can take up to a month in many cases. Arrays, in contrast, do not require a library creation step and have gone through years of standardization and bioinformatics development. In most cases this allows researchers to perform experiments from start to finish in a little over one week.

Microarrays have their downsides, however—a fact that is allowing sequencing to gain a small foothold within the community when it comes to profiling. As a result of their limited dynamic range, microarrays are substantially less quantitative than sequencing. “You are pretty much limited with microarrays to generally 3 to 4 logs of dynamic range,” says Illumina's Baker, “and it is very difficult to increase that range.” Therefore when it comes to very rare small RNAs, sequencing should provide a clearer picture of the expression patterns than microarrays.

When it comes to looking at multiple samples at the same time, even the high-throughput advantage of arrays is diminishing as researchers develop techniques to increase the number of samples that next-generation sequencers can analyze. “Bar-coding allows multiple samples to be interrogated in a single sequencing run,” says Wicki. Several developers, including Life



Life Technologies' SOLiD3 system, which uses a ligation-sequencing approach, has the capacity to profile tens of millions of small RNAs.

Technologies and Illumina, now offer bar-coding or indexing solutions that allow the profiling of several samples at the same time in the same flow cell lane, thereby reducing both the cost of the assay and increasing the number of samples that can be interrogated in a single run.

Even the economic argument might only be able to drive microarrays for so long. “I think the gap is certainly narrowing,” says Nusbaum. “There are things I am doing now with sequencing that I would have done last year by array.” Whereas this is partly driven by the sharply declining cost of sequencing that arrays cannot match, Nusbaum also notes that sequencing technology can be easily generalized, which he suspects will eventually drive all DNA readouts toward sequencing.

#### From discovery come tools

While discovery and profiling efforts advance our understanding of the small RNA world, there continues to be a strong interest among researchers in using these molecules to study gene function. And when it comes to small RNA tools, siRNA is the group's elder statesman.

Synthetic siRNAs have proved effective in gene silencing studies over the past 6 years. But figuring out the best way to design siRNAs for maximum silencing efficiency as well as to understand off-target effects is still a challenge for many researchers according to Devin Leake, director of research and development at Thermo Fisher Scientific

Dharmacon Products in Lafayette, Colorado, USA.

siRNAs consist of a sense and an antisense strand: the former strand acts as a passenger, and the latter is taken up by the RNA-induced silencing complex (RISC) to mediate gene silencing. Designing an antisense strand such that it is preferentially taken up by the RISC is at the heart of many siRNA design algorithms that optimize siRNA sequences so the two strands are thermodynamically different, with the antisense strand favored by the complex for uptake and effective silencing. Several online siRNA design programs exist for researchers to use in their designing efforts, including siDESIGN Center from Dharmacon Thermo Fisher Scientific, RNAi Designer from Clontech (Mountain View, California, USA), siRNA Target

Finder from Ambion (Austin, Texas, USA) and BIOPREDSi from Qiagen (Valencia, California, USA). Leake also notes that certain chemical modifications such as 2'-O-methyl modifications in specific positions on siRNA molecules and pooling approaches have also been shown to improve both specificity and potency.

But even with improved design tools, Chiang Li, an investigator at Beth Israel Deaconess Medical Center of Harvard Medical School in Boston and president of the Boston-based company Boston Biomedical, wondered for several years why it was so challenging to design good siRNAs to silence some of the genes his group was studying. This curiosity led Li and his colleagues to take another look at the basic structure of siRNA. At first they tried shortening the length of several siRNAs, but quickly found out that anything shorter than 19 base pairs lost activity very rapidly in their hands. They next tried asymmetric configurations, where one RNA strand was longer than the other, and suddenly everything changed according to Li. “As soon as you switch to the asymmetrical configuration, you come out of a tunnel and see this other world,” he says—a world where these much shorter asymmetrical RNAs not only result in increased silencing but also in diminished off-target effects as well as reduced immune response in tested cells.

The work, which demonstrates the use of 15-base-pair asymmetric RNAs (aiRNAs) to mediate silencing in mammalian cells, was recently published in *Nature Biotechnology*<sup>3</sup>.

“Many people have asked me ‘why does aiRNA have such silencing advantages when the molecule is just a few base pairs shorter than siRNA?’” says Li. He suspects that it all has to do with how the aiRNA molecule is loaded into the RISC: “Because of the asymmetrical structure, the sense strand is shorter, which makes it structurally unfit to be retained by RISC.”

#### A new toolkit

miRNA tools and inhibitors are also becoming more widely available for curious researchers. “Many researchers in the field find it difficult to link a specific miRNA with a phenotype,” says Leake. He strongly suspects this is the result of the subtle phenotypes that can occur when miRNAs are perturbed because unlike siRNA, which target a single gene for silencing, miRNAs can target hundreds of genes.

“An amazing new application of miRNA tools uses viral vectors,” says Leake. For

researchers interested in the effects of specific miRNAs, viral vectors offer an effective method for expressing a mature miRNA in many cell types. Several companies now offer viral vectors and miRNA libraries for researchers including Open Biosystems of Huntsville, Alabama, USA, which offers specialized retroviral vectors as well as a collection of human miRNAs cloned into lentiviral expression vectors, and System Biosciences of Mountain View, California, USA, which offers lentivirus-based miRNA expression vectors.

But the researchers in Leake’s group have been taking a different approach to study miRNA function, developing synthetic miRNA inhibitor molecules. “We evaluated the performance of traditional inhibitors and determined that chemical modifications alone would not be sufficient to improve inhibition,” says Leake. Instead, the group found that if they created a molecule with highly structured ends, the level of inhibition was enhanced<sup>4</sup>.

Small RNAs have given scientists a new way to look at gene regulation along with a toolkit to study gene function. With all that has been learned in the first decade and the new technologies available now, it is difficult to imagine the wonders the RNA world will reveal over the next ten years.

1. Ruby, J.G. *et al.* Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *Caenorhabditis elegans*. *Cell* **127**, 1193–1207 (2006).
2. Friedlander, M.R. *et al.* Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* **26**, 407–415 (2008).
3. Sun, X., Rogoff, H.A. & Li, C.J. Asymmetric RNA duplexes mediate RNA interference in mammalian cells. *Nat. Biotechnol.* **26**, 1379–1382 (2008).
4. Vermeulen, A. *et al.* Double-stranded regions are essential design components of potent inhibitors of RISC function. *RNA* **13**, 723–730 (2007).

---

Nathan Blow is the technology editor for *Nature* and *Nature Methods* (n.blow@boston.nature.com).

## SUPPLIERS GUIDE: COMPANIES OFFERING SMALL RNA ANALYSIS TOOLS AND SERVICES

Company	Web address
454 Life Sciences/Roche	<a href="http://www.454.com/">http://www.454.com/</a>
Agilent	<a href="http://www.agilent.com/">http://www.agilent.com/</a>
Amata biosystems	<a href="http://www.amata.com/">http://www.amata.com/</a>
Ambion	<a href="http://www.ambion.com/">http://www.ambion.com/</a>
Applied Biosystems	<a href="http://www.appliedbiosystems.com/">http://www.appliedbiosystems.com/</a>
Asuragen	<a href="http://www.asuragen.com/">http://www.asuragen.com/</a>
BD Biosciences	<a href="http://www.bdbiosciences.com/">http://www.bdbiosciences.com/</a>
BioCat	<a href="http://www.biocat.de/">http://www.biocat.de/</a>
BioChain	<a href="http://www.biochain.com/">http://www.biochain.com/</a>
Cellectricon	<a href="http://www.cellectricon.com/">http://www.cellectricon.com/</a>
Cenix Bioscience GmbH	<a href="http://www.cenix-bioscience.com/">http://www.cenix-bioscience.com/</a>
Clontech	<a href="http://www.clontech.com/">http://www.clontech.com/</a>
CombiMatrix	<a href="http://www.combimatrix.com/">http://www.combimatrix.com/</a>
CytRx	<a href="http://www.cytrx.com/">http://www.cytrx.com/</a>
Eurogentec	<a href="http://www.eurogentec.com/">http://www.eurogentec.com/</a>
Exiqon	<a href="http://www.exiqon.com/">http://www.exiqon.com/</a>
Geneservice	<a href="http://www.geneservice.co.uk/">http://www.geneservice.co.uk/</a>
Gene Tools	<a href="http://www.gene-tools.com/">http://www.gene-tools.com/</a>
GenoSensor Corporation	<a href="http://www.genosensorcorp.com/">http://www.genosensorcorp.com/</a>
genOway	<a href="http://www.genoway.com/">http://www.genoway.com/</a>
Helicos BioSciences	<a href="http://www.helicosbio.com/">http://www.helicosbio.com/</a>
Illumina	<a href="http://www.illumina.com/">http://www.illumina.com/</a>
Imagenex	<a href="http://www.imgenex.com/">http://www.imgenex.com/</a>
Integrated DNA Technologies	<a href="http://www.idtdna.com/">http://www.idtdna.com/</a>
Invitrogen	<a href="http://www.invitrogen.com/">http://www.invitrogen.com/</a>
In vivo Gen	<a href="http://www.invivogen.com/">http://www.invivogen.com/</a>
Life Technologies	<a href="http://www.lifetechnologies.com/">http://www.lifetechnologies.com/</a>
Mirus Bio	<a href="http://www.genetransfer.com/">http://www.genetransfer.com/</a>
Molecular Devices	<a href="http://www.moleculardevices.com/">http://www.moleculardevices.com/</a>
MWG Biotech	<a href="http://www.mwg-biotech.com/">http://www.mwg-biotech.com/</a>
New England Biolabs	<a href="http://www.neb.com/">http://www.neb.com/</a>
Novagen	<a href="http://www.emdbiosciences.com/">http://www.emdbiosciences.com/</a>
Oligoengine	<a href="http://www.oligoengine.com/">http://www.oligoengine.com/</a>
Open Biosystems	<a href="http://www.openbiosystems.com/">http://www.openbiosystems.com/</a>
Orbigen	<a href="http://www.orbigen.com/">http://www.orbigen.com/</a>
Perkin Elmer	<a href="http://www.perkinelmer.com/">http://www.perkinelmer.com/</a>
Polyplus Transfection	<a href="http://www.polyplus-transfection.com/">http://www.polyplus-transfection.com/</a>
Promega	<a href="http://www.promega.com/">http://www.promega.com/</a>
Qiagen	<a href="http://www.qiagen.com/">http://www.qiagen.com/</a>
Regulus Therapeutics	<a href="http://www.regulusrx.com/">http://www.regulusrx.com/</a>
Roche Applied Science	<a href="http://www.roche-applied-science.com/">http://www.roche-applied-science.com/</a>
Rosetta Genomics	<a href="http://www.rosettagenomics.com/">http://www.rosettagenomics.com/</a>
SABiosciences	<a href="http://www.sabiosciences.com/">http://www.sabiosciences.com/</a>
Sigma Aldrich	<a href="http://www.sigmaaldrich.com/">http://www.sigmaaldrich.com/</a>
Stratagene	<a href="http://www.stratagene.com/">http://www.stratagene.com/</a>
Sylentis	<a href="http://www.sylentis.com/">http://www.sylentis.com/</a>
Targeting Systems	<a href="http://www.targetingsystems.com/">http://www.targetingsystems.com/</a>
tebu-bio	<a href="http://www.tebu-bio.com/">http://www.tebu-bio.com/</a>
Thermo Fisher/Dharmacon	<a href="http://www.dharmacon.com/">http://www.dharmacon.com/</a>
Vectalys	<a href="http://www.vectalys.com/">http://www.vectalys.com/</a>