

More genes for less	459
Box 1: Building oligos	458

Microarrays, megasynthesis

Monya Baker

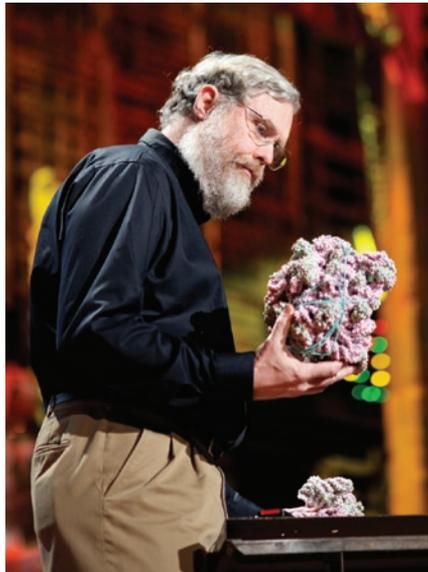
Microarrays are not just for analyzing gene expression.

Jonathan Weissman at the University of California, San Francisco wants to know how cells fold proteins into the right shapes, and he knocks down genes to find out. For that he uses massive RNA interference libraries, each containing about 50,000 oligonucleotide sequences, 20 or 30 for each gene he wants to interrogate¹. “It would be completely impractical to synthesize so many by traditional methods,” he says. Instead, Weissman relies on microarray technologies that make diverse oligonucleotides in parallel. In fact, Weissman does not even see the microarrays on which his oligonucleotides are printed; they are shipped in bar-coded vials straight from the supplier.

Carefully designed oligonucleotide probes made from microarrays also enabled Kun Zhang, at the University of California, San Diego, to compare genome-wide surveys of DNA methylation across two cell types². “The thing that changed the landscape is that we started a collaboration with Agilent, and they looked at the idea of using a microarray as a DNA synthesizer,” recalls Zhang. Though he had initially explored manufacturing a few hundred probes using column-based synthesis, costs were prohibitive. And Agilent increased the numbers of probes produced on its microarrays from an initial 10,000 to over 50,000.

There are many ways to use pools of diverse oligonucleotides, says Jay Shendure at the University of Washington, Seattle, who has developed target-capture and several other uses of oligonucleotides. “People are making complex libraries for many reasons,” he says. “Anything where you can synthesize a lot of oligos and you don’t care if they are in one big pool is a potential application.”

Huge, cheap libraries of oligonucleotides are changing the scale and scope of many experiments. In the last five years, oligonucleotides synthesized on microarrays have been used



Microarrays can drive down the cost of DNA synthesis and create a wealth of new applications, says George Church at Harvard Medical School.

for profiling regulatory elements of genes and creating cDNA and RNA libraries to examine how different sequences affect function. Another application is genome enrichment, in which oligonucleotides are used to select only certain portions of genetic material for subsequent amplification and analysis. Several vendors provide libraries of thousands to millions of oligonucleotide probes, each designed to amplify desired parts of the genome for analysis by next-generation sequencing. In fact, this ‘targeted-capture’ approach has led to relatively quick identification of the genetic culprits behind several Mendelian diseases, a search that had long eluded other approaches³. Another emerging approach is to use oligonucleotides produced on microarrays to reliably assemble genes.

Just as huge price drops in sequencing DNA opened up experiments in individual variation, gene expression and more, a fall

in the price of synthesizing DNA will also open new frontiers in synthetic biology and systems biology, says George Church, who is at Harvard Medical School and an early pioneer in using microarrays for synthesis. (Many scientists interviewed for this feature are collaborators with or former members of Church’s laboratory.) “It’s been hard to do synthetic experiments,” says Church. “The cost-curve for synthesis has lagged behind sequencing, but it’s catching up now. The time is ripe for turning to something that is high-throughput.”

Longer, better oligos

Standard DNA synthesis uses microliters of reagents in tiny glass columns to make individual sequences (**Box 1**). Microarrays miniaturize and parallelize the synthesis, producing thousands of sequences side by side; small volumes of reagents normally used for single reactions can wash over an entire slide of tens of thousands of nucleotides at once.

Compared with traditional DNA synthesis techniques, microarrays offer a far less expensive source of oligonucleotides; long oligos (about 100 to 200 nucleotides) generally cost about \$0.10 per nucleotide from commercial vendors, but microarray-based methods can be used to produce oligonucleotides for considerably less, about a million 60-mers for \$600 in some cases, though prices can vary for many reasons.

Several aspects of microarray synthesis make these oligonucleotides harder to use, however. Each oligonucleotide is produced in very small amounts. Hundreds of thousands of oligonucleotides may be made on certain types of microarrays, but there may be only a million or so molecules of each one. And when these are cleaved from the array, they mix together, forming a large heterogeneous pool.

BOX 1 BUILDING OLIGOS

David Lockhart, who led the team that invented the first commercial microarrays at Affymetrix, says that the idea of using a microarray as a DNA synthesizer appealed from the beginning, but the quality of the DNA produced on the array in the 1980s simply was not high enough. Chemistries have improved considerably since then, but microarrays still have a common problem of accumulating errors, he says. “You don’t get purification away at each step, so you end up with truncation products and places where things went wrong.” Thus, as oligonucleotides get longer, the yields go down because molecules either stop growing or incorporate the wrong nucleotides.

Several techniques can be used to add desired nucleotides to the growing oligonucleotides on a microarray¹⁰. Agilent uses a form of inkjet printing. Tiny nozzles for each monomer, plus others for additional monomers and chemical activators, precisely place picoliter drops of reagents in desired spots on an array. MYcroarray uses light-sensitive reactions combined with programmable mirrors that can activate molecules at specific locations. It synthesizes batches of 20,000 oligonucleotides per batch with lengths between 50 to 150 nucleotides.

Technology developed by CombiMatrix, now sold by CustomArray, uses tens of thousands of microelectrodes to direct reaction-promoting acidic conditions at desired spots. They have two platforms that can synthesize 12,000 or 94,000 oligonucleotides. Although the arrays can produce quality oligonucleotides of 200 base pairs, most customers order 120- and 140-mers, says senior scientist Marcelo Caraballo. In addition, CustomArray sells the synthesizer itself. “Our synthesizer does not utilize lasers or mirrors or precision-sensitive mechanical devices, which require extensive calibration and maintenance,” says Caraballo. “The array itself is the device that actively controls synthesis, while the machine simply moves chemicals around and sends signals to the array.”



“Where the microarrays become efficient is where you make a hundred oligos all at once,” says Emily Leproust at Agilent Technologies.

LC Sciences has a microfluidic system with picoliter reaction chambers and a digital optical device that uses light to generate acid, which drives desired reactions in each chamber. An advantage of the light-activated system is that it allows oligonucleotides to be built from the same chemical building blocks as those used in standard column synthesis but at a smaller scale, says Chris Hebel, vice president of business development at LC Sciences. “We see the same error rates as on an oligo column with no reduction in coupling efficiency as typically seen with modified building blocks.” Though the company routinely synthesizes oligonucleotides as long as 120 base pairs, says Hebel, most researchers want molecules that are in the 60–70–base-pair range; the most common application are carefully designed primers for targeted capture. The company recently introduced an array capable of making 30,000 distinct oligonucleotides at once, and for customers who want more diversity, arrays can be programmed to produce ‘degenerate molecules,’ in which any one of the four nucleotides is randomly incorporated into desired places in the molecules.

By improving reagents and reaction conditions, Agilent could consistently make high-quality pools of oligonucleotides up to 100 bases, explains Emily Leproust, who directs chemistry applications for genomics research and development at Agilent Technologies, and has collaborated with Church, Shendure, Weissman, Zhang and other academic scientists. Beyond 100 base pairs, however, depurinations build up in certain nucleotides, ultimately cleaving the nucleotide. Leproust’s group, however, recently published a technique that reduces this reaction and allows the synthesis of longer oligonucleotides that are still high quality¹¹. Though libraries are only available to a limited number of collaborators, hundreds of thousands can be made on the same microarray. “Now we are routinely making and sending oligos up to 200 base pairs long,” she says.

No matter how researchers intend to use libraries of oligonucleotides, they usually want more oligonucleotides, longer oligonucleotides and lower error rates. “Chemistry is chemistry. There is no 100% complete reaction, so there will be errors,” says Jingdong Tian at Duke University, who has used anti-sense oligo arrays to remove impurities⁴. If oligonucleotides are designed to be the same length, high-performance liquid chromatography and polyacrylamide gel electrophoresis can be used to purify oligonucleotides, but neither technique is perfect, and both are expensive and time-consuming.

“For most people, the economics are still in favor of just one-at-a-time small-scale synthesis,” says Tom Ellis, a synthetic biologist at Imperial College London. Because

few people want more than a dozen genes, he says, ordering genes from a vendor will cost less than making a microarray, sequencing products for accuracy and assembling products into genes.

When large quantities of highly accurate oligonucleotides are needed for, say,

Diverse oligonucleotides made by microarrays are letting researchers explore many genetic elements and do more efficient sequencing, says Jay Shendure at the University of Washington, Seattle.



standard PCR primers or gene synthesis, researchers typically turn to traditional synthesis from commercial suppliers. “I don’t think preparative microarrays will displace column-based syntheses,” says Shendure. “We’re adding to the application space rather than competing.”

Extending the length of oligomers extends their uses, says Shendure. For example, Shendure hopes to stitch together strategically varied nucleotide sequences for modestly sized proteins and so explore the functional landscape of known proteins. Although often-used technologies such as phage display use random mutation and selection pressure to hunt out new protein functions, microarray synthesis could enable a more systematic hunt of how new protein sequences might

"Microarrays traditionally have been used for analysis of gene expression, but now they are starting to be used for gene synthesis," says Jingdong Tian at Duke University.



yield new functions. Synthesized constructs could be placed into vectors and mixed with cells, and their products could be assessed. And that is just one of many possibilities, says Shendure. "If [the achievable length] was 300 base pairs or even a kilobase, there are a lot of things one could do that one can't do now."

More genes for less

Whereas genes can be produced with cloning, this involves several steps: synthesizing PCR primers, PCR, restriction digests and DNA isolation. Moreover, cloning allows relatively few changes to be made to an existing genetic sequence. In contrast, gene synthesis lets researchers dictate the desired sequence; once ordered from a vendor, genes are generally shipped in two weeks or less. For researchers who want to explore many genes, microarrays could make gene synthesis cheaper and faster.

Genes are sensitive to errors. A single missing nucleotide—the most common error in oligonucleotide synthesis—can mean that no functioning protein is made. Even when error-free oligonucleotides are produced, the job is not complete. Most genes are around 1,000 kilobases long, so separate oligonucleotides must be collected and then assembled, a task that is more difficult with the complex mixtures of oligonucleotides produced by microarrays.

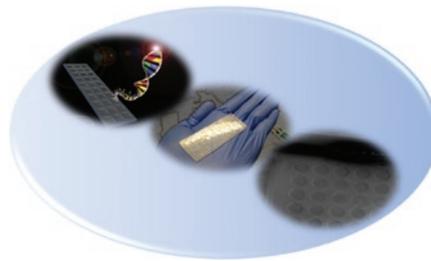
Right now, synthesized genes must also be sequenced to make sure that there are no errors. It is a Catch-22, says Church. "If you do high-throughput synthesis on chips, you have a bottleneck in sequencing. If you use high-throughput sequencing [to verify low-throughput synthesis], you have a bottleneck in making the oligos." Thus, he explains, "if you order a gene from a company, it's almost always first-generation synthesis and first-generation sequencing, meaning synthesis on a lot of columns and sequencing by Sanger sequencing." That is because microarray production of oligonucleotides and next-generation sequencing are hard to integrate.

In December 2010, the Church group described a technique to do so⁵. Next-generation sequencing generally involves

making many copies of a given DNA molecule. Church, along with scientists from microarray company febit and Stanford University, fed synthetically produced oligonucleotides into a next-generation sequencing machine, and then worked out a way to find and collect the error-free, amplified sequences. In this case, the researchers used the Roche 454 sequencing machine, which produces copies of sequenced DNA fragments clustered onto micrometer-sized beads.

Compared to the oligonucleotide pool originally generated by the microarray, error rates fell by around a factor of 500, explains Church: for 200-mers, one oligonucleotide in 250 contained an error; for 130-mers, the error rate was one in 1,300; for 130-mers subjected to enzymatic error correction, the rate was one in 6,000. Still, several steps are necessary before the approach can be used practically, he says. Methods to collect many beads have yet to be worked out, and sometimes the wrong bead is collected; although most next-generation sequencing machines could be used to generate highly accurate oligonucleotides, says Church, much optimization will be necessary for this use to become practical for routine gene assembly. The technique, he says, is "more of a work in progress; it's to stimulate people to think about integrating next-generation sequencing and synthesis."

Also in December last year, Church, in collaboration with scientists from the Wyss Institute at Harvard and Agilent Technologies, described another technique to use microarrays for gene synthesis⁶. This approach addresses the problem of assembling mixtures of oligonucleotides produced by microarrays. Although the complex mixtures are acceptable for applications such as capture probes or libraries of short regulatory elements, preparing microarray-produced oligonucleotides for gene assembly



Ali Saeem, Tian lab, Duke University

Several new techniques may soon allow oligonucleotides produced on microarrays to be used regularly for inexpensive gene synthesis.

is labor-intensive. Tens of thousands of diverse nucleotides are cleaved from the microarray all at once, which must then be segregated into smaller pools.

To overcome the need for such separation, the researchers designed a quarter-million short PCR probes so that the probes would interact only with desired sequences and then used PCR to amplify only selected 130-mer or 200-mer oligonucleotides for subsequent assembly. As proof of concept, they assembled 40 of 42 genes for single-chain antibodies, sequences which are particularly difficult to assemble because of frequent repetitive sequences and high proportions of guanine and cytosine.

Larger assemblies are also possible from short oligonucleotides. A report from the Venter Institute last year demonstrated that the 16-kilobase mouse mitochondrial genome could be swiftly assembled from 60-mers spanning the genome, in this case manufactured by column synthesis⁷.

Church and his collaborators plan to offer their technology as a service, starting this fall, through a program called SynBioSIS. Researchers who want thousands of assembled oligonucleotides or who have very robust selection systems could benefit from having many assemblies available at low cost, he says. A 500-base-pair construct would cost around \$10. Church says that the work is going to be used for applications in which the one in 1,000 error rates will not be a hindrance. For many applications, researchers will want to use synthetic genes that have been sequence-verified, he says. Still, he hopes the service offering will launch some interesting questions.

In fact, scientists are beginning to have several techniques to choose from. An approach described by Kathryn Sykes, Alex Borovkov and colleagues at Arizona State University combines hybridization with selection, preventing error-containing oligonucleotides from being incorporated into genes. Using their protocol, they assembled genes using oligonucleotides produced by Agilent, CombiMatrix and LC Sciences. Although less than one in 20 oligonucleotides directly eluted off of a microarray was perfect, assembled genes had only 2.7 errors per kilobase⁸.

Rather than using selective amplification or hybridization to collect desired oligonucleotides, Tian and colleagues at

Duke University built a microfluidics system to sequester them⁹. The microarray that his team built is divided into as many as 30 physically isolated subarrays, each of which contains the pieces necessary to assemble DNA molecules up to a kilobase long. After synthesis, oligonucleotides are amplified and assembled, all within the subarray. In an initial publication, raw error rates of assembled genes were one in 526 base pairs; adding an endonuclease that recognizes and cleaves mismatched double-stranded DNA reduced the error rate to one in 5,392 base pairs.

Tian used the chip to create a library of synthetic genes. One library consisted of over a thousand genes, all encoding the protein LacZa but using synonymous codons; about one-third of these variants had higher expression levels than the original gene.

Similarly constructed libraries were also used to optimize expression of the *Drosophila melanogaster* transcription factors. Tian estimates that the cost of the final synthesized sequences including error correction is about half a cent per base pair, a drop in costs that could make gene synthesis more widely available. At that rate, researchers could use large-scale gene synthesis not just to construct new genetic pathways but to optimize gene expression, by optimizing protein-coding sequences and regulatory elements, Tian says. "You can just synthesize whatever genes you dream of. It could make a big difference not just in synthetic biology, but in general biomedical research."

1. Bassik, M.C. et al. *Nat. Methods* **6**, 443–445 (2009).
2. Deng, J. et al. *Nat. Biotechnol.* **27**, 353–360 (2009).
3. Ku, C.S., Naidoo, N. & Pawitan, Y. *Hum Genet.* **29**, 351–370 (2011).
4. Tian, J. et al. *Nature* **432**, 1050–1054 (2004).
5. Matzas, M. et al. *Nat. Biotechnol.* **28**, 1291–1294 (2010).
6. Kosuri, S. et al. *Nat. Biotechnol.* **28**, 1295–1299 (2010).
7. Gibson, D.G., Smith, H.O., Hutchison, C.A. III, Venter, J.C. & Merryman, C. *Nat. Methods* **7**, 901–903 (2010).
8. Borovkov, A. et al. *Nucleic Acids Res.* **38**, e180 (2010).
9. Quan, J. et al. *Nat. Biotechnol.* advance online publication 24 April 2011 (doi:10.1038/nbt.1847).
10. Tian, J., Ma, K. & Saaem, I. *Mol. Biosyst.* **5**, 714–722 (2009).
11. Leproust, E.M. et al. *Nucleic Acids Res.* **38**, 2522–2540 (2010).

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