

Phase change	540
Virtual 'multiplex' PCR	541
Key metrics	542
Capturing a moment in time?	542

Genomics: catch me if you can

Nathan Blow

Next-generation sequencing has made decoding entire genomes cheaper and faster. But what about those researchers who only want to sequence a small section of a genome or focus on a couple thousand specific exons? A wave of new technologies has recently emerged that should help these scientists target their sequencing efforts to sequences of interest.

In May 2009 Applied Biosystems, a division of Life Technologies, announced the sequencing of an entire human genome at 17-fold coverage, 50 billion mappable bases, all from a single run on the company's SOLiD 3 next-generation sequencing platform. Such an increase in next-generation sequencing output, alongside declining costs, have expanded the scope of what is possible in whole-genome sequencing studies. But researchers interested in using the technology to investigate specific regions of the genome have found it tougher going.

"We saw this desire within the research community to sequence not the whole genome but portions of a genome," says Xinmin Zhang, a product manager for emerging genomic applications at Roche NimbleGen. "And in those cases what the researchers needed was a way to target their region of interest." Traditionally these scientists would rely on PCR to amplify target sequences from the genome before sequencing. But next-generation approaches, with their capacity to generate gigabases of sequence data per run, require more input than what a traditional PCR capture strategy can deliver, a fact that has led to the development of new capture and enrichment technologies designed specifically for next-generation sequencing applications.

Starting line

In November 2007, three articles describing new capture and enrichment protocols for multi-megabase genomic regions appeared in the pages of *Nature Methods*. Two of the approaches relied on array-based hybridization capture methods^{1,2}, and the third took



The RainDance RDT 1000 can be used in PCR-based sequence capture applications.

advantage of molecular inversion probes to isolate specific genomic regions³.

"Our first thought was to use the array like an affinity column to pull down regions of a genome," explains Zhang, discussing the strategy behind array-based capture. The initial proof-of-concept work used fragmented and amplified genomic DNA that was hybridized to an array designed to capture 6,726 genomic regions as well as a series of chromosomal regions ranging in size from 200 kilobases to 5 megabases before sequencing. The results were promising with 65–75% of the captured sequence reads mapping back to the targeted region and an overall average enrichment of 432-fold, prompting Roche NimbleGen to commercialize the approach in 2008.

Roche NimbleGen has since expanded and advanced the method, now offering a variety of standard and custom capture arrays, including a new human exome array built on their high-density array platform with 2.1 million oligonucleotide probes, which enables the capture of 180,000 human coding exons as well as nearly 550 microRNA exons.

As Roche NimbleGen and others advance the use of microarrays for sequence capture, a company called febit is taking the approach of miniaturizing their capture arrays. "We use a microfluidic biochip that is partitioned to contain eight separate arrays capable of synthesizing up to 60,000 features each," explains Peer Staehler, chief scientific officer at febit. The technology, which they call HybSelect, was released in March 2009, and Staehler says

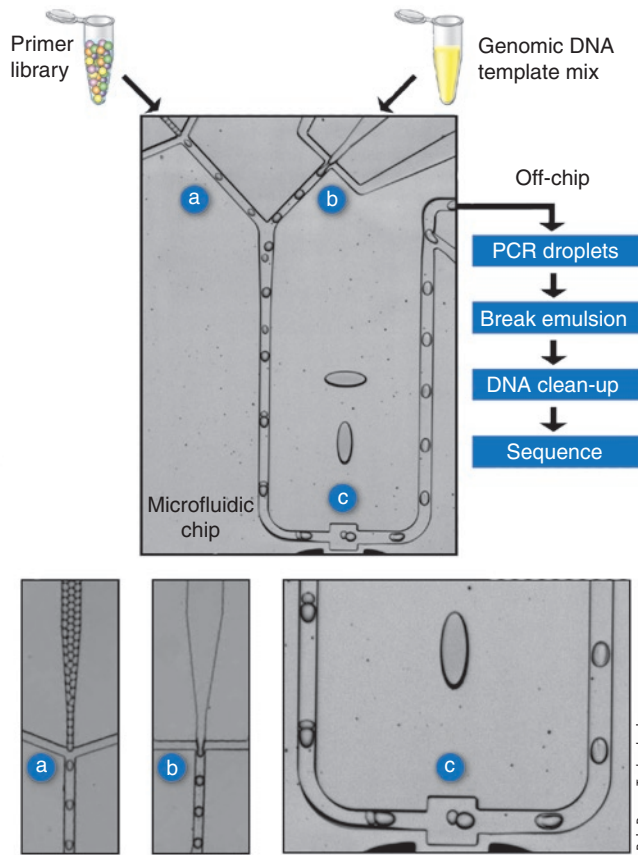


Diagram illustrating the RainDance approach using PCR for sequence capture.

several institutes and laboratories around the world have started using the approach for their sequence-capture applications.

The main differences between Roche NimbleGen's microarray approach and febit's biochip are in the amount of sequence capture possible and the intended applications. At the moment the NimbleGen Sequence Capture Human Exome array is capable of capturing over 30 megabases of sequence according to the Zhang, whereas febit's approach is limited to 9.5 megabases per custom array. But Staehler notes that unlike the other array-based hybridization approaches, the HybSelect technology permits a high degree of multiplexing owing to the eight separate arrays on a single chip. "We are convinced that for a lot of users it is highly relevant to have small target sizes and process a lot of samples in a single study," he says, especially as more and more researchers start to use resequencing approaches to validate and advance results obtained from genome-wide association studies.

Phase change

Arrays are not the only hybridization approach available to researchers interested in sequence capture. In fact, for some

developers the solution to the challenge of capture might be in a solution.

"Molecular inversion probes are extremely specific: you get what you want," says Jay Shendure, an assistant professor of genome sciences at the University of Washington. Molecular inversion probes, or padlock probes, have been previously used in large-scale single-nucleotide polymorphism genotyping applications⁴, but it was Shendure and

his colleagues who applied these probes to next-generation sequencing sample preparation in 2007.

Molecular inversion probes used for large-scale sequence capture are usually single-stranded and around 70 nucleotides in length, composed of a universal core of 30 nucleotides, which is flanked by specific 20-nucleotide targeting sequences on each side. The targeting sequences are designed to hybridize to specific genomic regions upstream and downstream of a sequence of interest. After hybridization in solution, polymerase is used to copy the targeted region, and both the probe and copied region are circularized through the addition of ligase. The initial work in 2007 demonstrated capture of nearly 10,000 human exons in a single multiplex reaction.

Although Shendure and others continue to advance the use of molecular inversion probes for sequence-capture applications, including recent work in which molecular inversion probes were used to target and capture specific methylated regions of a genome^{5,6}, the number of groups using these probes is still relatively small according to Shendure. "The key thing holding back the broad scale adoption of this approach is the lack of commercialization," he suspects.

But one solution-based approach that is seeing broader adoption at the moment was developed by researchers at the Broad Institute⁷ and is now being commercialized through Agilent Technologies. In this approach, which Agilent calls the SureSelect Target Enrichment System, biotin-labeled RNA 'bait' probes are used to 'fish' specific DNA sequences out of a 'pond' of DNA fragments. When the RNA is hybridized to a fragmented genomic library, DNA and RNA



The commercialization of sequence-capture technologies and approaches has occurred in a very short period of time.

© 2009 Nature America, Inc. All rights reserved.



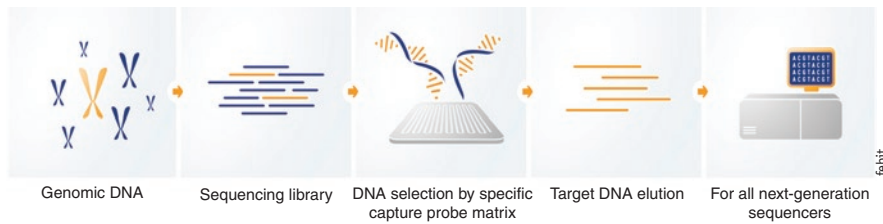


Illustration of the steps involved in sequence capture using febit's hybridization-based biochip approach.

hybrid duplexes are formed, which can be collected and captured using streptavidin beads, whereas the uncaptured, nontargeted portion is thrown away during washing. In an elution step, DNA is then released from the beads and RNA is digested.

Emily LeProust, director of genomics applications and chemistry research and development at Agilent, says that performing this whole operation in solution is one of the keys to the success of their capture method. "It is the combination of the fact that we use an excess of RNA bait in very small volume for the hybridization that drives the equilibrium forward," she explains. Probe length is also important, and LeProust notes

that the Agilent SureSelect method takes advantage of 120-base-pair probes, which she says helps avoid bias when capturing large numbers of different sequences. Still, hybridization approaches have difficulty when it comes to distinguishing between closely related gene sequences or capturing highly repetitive regions of the genome.

Virtual 'multiplex' PCR

"There are fewer limitations for PCR in these sections of the genome," says Jeremy Lambert, product manager for genomics at RainDance Technologies. Although it is potentially a less biased approach to sequence capture, traditional PCR has

always proved difficult to multiplex in large numbers. "Multiplex PCR is difficult after ten primer pairs, and even there you would need optimization and still have the issue of uniformity," says Shendure. But RainDance Technologies is betting that their new 'multiplex' PCR approach could provide the necessary throughput needed for next generation sequencers while also giving the benefits of a PCR-based approach.

RainDance was founded four years ago in an effort to advance the use of aqueous oil droplets as reaction vessels. "We like to think of each droplet as the functional equivalent of a test tube," says Darren Link, who is vice president of research and development at RainDance, as well as one of the founders of the company. The droplets, 30 micrometers in size and capable of holding 26 picoliters of liquid, are generated at a rate of 3,000 per second, nearly 10 million an hour. Link says taking advantage of these droplets as individual test tubes to perform PCRs, specifically for targeted resequencing, became a focus for the company a little more than a year ago.



The febit approach to sequence capture relies on the use of microfluidics and multiplexed arrays.

The theory behind the RainDance approach to PCR is simple: one droplet will contain a single forward and reverse primer pair and a second droplet will contain the PCR mix and DNA template of interest. The challenge for engineers at RainDance was to create a reliable microfluidic platform that could move and subsequently merge these droplets together in a high-throughput manner. They accomplished this by infusing the primer droplets in one microfluidic channel and the template droplets in another. Application of an electric field induces the two droplets to come close together and eventually merge into a single droplet, resulting in the creation of millions of individual PCRs in individual droplets over the course of an hour. At the moment, Lambert says the RainDance approach is capable of amplifying up to 4,000 PCR products per sample and processing eight samples per day on a single instrument.

“The cool thing about the RainDance approach is that in principle it should lead to good uniformity,” says Shendure. He notes that by doing independent PCR amplifications within each droplet, all the reactions reach saturation independently, and so they are not competing with one another, which should result in less bias in terms of which fragments are captured. But with the application only released in March, Shendure and others are waiting to see how users perform with the instrument to get a better handle on the exact overall uniformity of target amplification, which is just one of the key metrics for any capture strategy.

Key metrics

Specificity is the other key metric for sequence capture and enrichment applications. And when discussing various capture methods, uniformity and specificity are bound to come up during the conversation.

“We are going to capture everything in the prepped libraries, and when we couple this to the long oligonucleotides we are using, there is very little bias with the captured sequences,” explains LeProust when describing Agilent’s SureSelect solution-based hybridization approach. But Zhang argues that there are more enzymatic amplification steps involved in the production of probes for solution-based methods when compared to array-based approaches, creating a potential source of bias in the probe representation itself. “You might start with probe A and probe B at similar amounts, but after all the amplification you might not have the same amounts, which will affect the uniformity.”

A nearly 99% specificity is the advantage of using molecular inversion probes for sequence capture. But although the specificity is extremely high, the uniformity of the approach has been low, an issue Shendure’s group tackled in a May 2009 *Correspondence to Nature Methods*. “We just increased the hybridization time of the probes to genomic DNA from what was a relatively short time to quite a bit longer. There were a few other changes as well, but it was primarily an issue of hybridization kinetics,” he says. Shendure’s simple modification to the protocol increased the uniformity from 18% to 91%, putting the approach on par with the other capture techniques. “I would say that whereas before we were lagging way behind [in uniformity], we are now in the mix.” His group is now taking advantage of the new protocol in experiments aimed at capturing the whole human exome in a single tube.

Capturing a moment in time?

The big question that remains for all new sequence capture approaches is what will happen if developers finally reach the point

where an entire genome can be sequenced for \$1,000 or less? “This is something that I think about everyday, something everyone working in this area has to think about,” says NimbleGen’s Zhang.

Most developers suspect the future of sequence capture and enrichment will come down to a simple matter of economics. “I think in one or two years from now you will be able to sequence one megabase for less than \$5,” says Staehler, which he suspects will enable targeted resequencing studies examining tens of thousands of samples. And these studies with their tremendous sample numbers will generate different sets of data than sequencing only a small number of complete genomes for the same price tag.

“As sequencing becomes much higher-throughput and lower-cost, capture has to catch up and become cheaper as well to enable low-cost capture. If the cost of capture is 10–100 times less, then there will always be a niche,” suspects Zhang. Although he does add that just how big that niche will be remains to be seen.

Others are more doubtful about the future of capture and enrichment for next-generation sequencing. “This will not last forever; it is a transient moment, but what is hard to say is if it will go on for one year or five years,” says Shendure. Although lingering questions about the longevity of sequence-capture approaches remain, there is no questioning the value of the technologies to many researchers at the moment.

1. Albert, T.J. *et al.* Direct selection of human genomic loci by microarray hybridization. *Nat. Methods* **4**, 903–905 (2007).
2. Okou, D.T. *et al.* Microarray-based genomic selection for high-throughput resequencing. *Nat. Methods* **4**, 907–909 (2007).
3. Porreca, G.J. *et al.* Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
4. Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* **21**, 673–678 (2003).
5. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009).
6. Ball, M.P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* **27**, 361–368 (2009).
7. Gnirke, A. *et al.* Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat. Biotechnol.* **27**, 182–189 (2009).

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

TECHNOLOGY FEATURE

SUPPLIERS GUIDE: COMPANIES OFFERING GENOMICS INSTRUMENTATION AND REAGENTS

Company	Web address
454 Life Sciences, a Roche company	http://www.454.com/
Affymetrix	http://www.affymetrix.com/
Agencourt Bioscience Corporation	http://www.agencourt.com/
Agilent Technologies	http://www.agilent.com/
Applied Biosystems	http://www.appliedbiosystems.com/
Arrayit Corporation	http://www.arrayit.com/
Beckman Coulter	http://www.beckmancoulter.com/
Bruker Daltonics	http://www.bdal.com/
Clontech	http://www.clontech.com/
Cogenics	http://www.cogenics.com/
deCODE Genetics	http://www.decode.com/
DNADirect	http://www.dnadirect.com/
DNASTAR	http://www.dnastar.com/
Epigentek	http://www.epigentek.com/
Eurofins MWG Operon	http://www.operon.com/
Expression Analysis	http://www.expressionanalysis.com/
febit	http://www.febit.com/
GE Healthcare	http://www.gehealthcare.com/
GenScript	http://www.genscript.com/
Geospiza	http://www.geospiza.com/
Helicos Biosciences	http://www.helicosbio.com/
Illumina	http://www.illumina.com/
Intelligent Bio-Systems	http://www.intelligentbiosystems.com/
Invitrogen	http://www.invitrogen.com/
LI-COR Biosciences	http://www.licor.com/
LingVitae	http://www.lingvitae.com/
Maxim Biotech	http://www.maximbio.com/
NABsys	http://www.nabsys.com/
Nanogen	http://www.nanogen.com/
New England Biolabs	http://www.neb.com/
Open Biosystems	http://www.openbiosystems.com/
Pacific Biosciences	http://www.pacificbiosciences.com/
Perkin Elmer	http://www.perkinelmer.com/
Promega	http://www.promega.com/
Qiagen	http://www.qiagen.com/
RainDance Technologies	http://www.raindancetechnologies.com/
Reveo	http://www.reveo.com/
Roche Applied Science	http://www.roche-applied-science.com/
Roche NimbleGen	http://www.nimblegen.com/
Sequenom	http://www.sequenom.com/
SeqWright	http://www.seqwright.com/
Soft Genetics	http://www.softgenetics.com/
Stratagene	http://www.stratagene.com/
Taconic	http://taconic.transnetyx.com/
Transgenomic	http://www.transgenomic.com/
VisiGen Biotechnologies	http://www.visigenbio.com/