

FRETing about sequencing	268
Sequencing by serendipity	269
The non-cutting edge	271
The proof is in the pore	271
Box 1: Of designs and nanopores	272

DNA sequencing: generation next-next

Emboldened by the success of next-generation sequencing, scientists are pursuing the holy grail of genomics—the ‘\$1,000 genome’—with single-molecule approaches. Nathan Blow reports.

Next-generation DNA sequencing has started a revolution in genomics and created the opportunity for large-scale sequencing projects, such as the recently announced 1,000 Genomes Project—an international effort to sequence the genomes of at least 1,000 people from around the world, which a few years ago would not have been feasible. But in the future, many researchers think that ‘next-next’ generation sequencing using single molecules might be able to take the genomics community even further.

“The reason that single-molecule methods have such appeal on paper is that they are the least manipulative and most direct way to get an answer from the sample,” says Tim Harris, senior director, research at Helicos Biosciences. Because these methods avoid the time, error and cost associated with sample

preparation and amplification, Harris thinks single-molecule methods will allow researchers to get at genomic data faster and at lower cost than presently available sequencing methods.

Steven Block of Stanford University in Palo Alto, CA, USA concurs: “When single-molecule sequencers finally come along, they will be terribly important and something to behold.” The development of single-molecule sequencing technology is just beginning with a growing number of groups taking very different approaches into how to sequence single molecules.

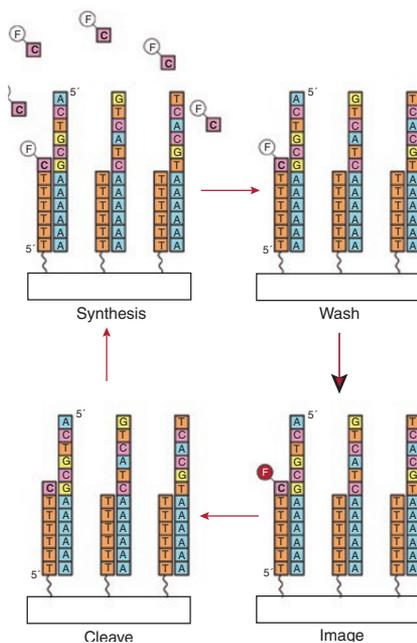
Snapshots of single molecules

Helicos Biosciences, located in Cambridge, MA, USA, may be the first company to offer a ‘next-next’ generation DNA sequencing system for single molecules. Steve Lombardi, president and chief operating officer at Helicos, says that Helicos’ single-molecule sequencing system, which they are targeting for a 2008 launch, can analyze DNA or RNA, examine sequence variation and potentially look at epigenetics. “Our goal was to build a production-level genetic analyzer that can enable big experiments.”

Helicos’ technology, which they call true single-molecule sequencing or tSMS, is a sequencing-by-synthesis approach for single molecules that is implemented in an instrument called the HeliScope single-molecule sequencer. To prepare a sample, DNA is fragmented into 100–200-base-pair pieces, and then adaptors of known sequence, often poly(A) tails, are attached to the ends of DNA fragments. “We can capture that poly(A) tail on a surface that has a poly(T) primer covalently attached,” says Harris. For imaging during tSMS, the spacing of the captured fragments is very important. “The resolution of optical microscopy is a few hundred nanometers, so captured molecules have to be at least that far apart,” notes Harris.



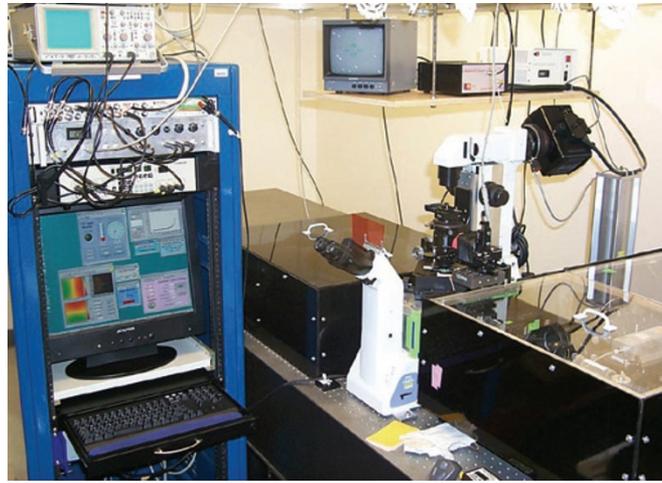
Helicos Biosciences’ HeliScope is expected to be the first commercially available ‘next-next’ generation DNA sequencer. (Courtesy of Helicos Biosciences.)



A schematic of the sequencing-by-synthesis approach used by Helicos Biosciences. (Courtesy of Helicos Biosciences.)

45 bases across billions of strands in a single run.

Imaging during each cycle was another challenge for the developers at Helicos. “We could not image the entire surface at once, so we take a picture every few hundred micrometers over the whole surface for each cycle,” says Harris. And the acquisition of this many high-resolution images during the sequencing process results in a tremendous amount of data points. Each run on the Heliscope needs 14 terabytes of computer storage space, but Harris says that those 14 terabytes can hold an enormous amount of sequence data; the imaging system has the potential to deliver 1 gigabase of sequence per hour. Although the current version is not even near this mark now, both Harris and Lombardi are quick to point out that this gives the system enormous headroom for future improvements in the sequencing chemistry.



Steven Block is using optical traps to hold strands of DNA in place for direct sequencing with RNA polymerase. (Courtesy of S. Block.)

FRETing about sequencing

“It has always been in my mind that sequencing is something that has to be improved,” says Susan Hardin, chief executive officer of VisiGen Biotechnologies located in Houston, TX, USA. She and four

other faculty members started VisiGen Biotechnologies in 2000 as a spin-off from the University of Houston.

“We are engineering the polymerase and nucleotides to act as direct molecular sensors of base identity in real time,” says Hardin about VisiGen’s fluorescence resonance energy transfer (FRET)-based approach to single-molecule sequencing. The engineered polymerase contains a donor fluorophore, and one of four differently colored acceptor fluorophores is attached to the gamma phosphate of each of the nucleotides. When a nucleotide is incorporated, the proximity causes a FRET signal.

The DNA molecule lights up, and the color indicates the base identity because the fluorophores on the nucleotides are color-coded. “You can think of it as watching the polymerase during the DNA synthesis process,” says Hardin.

Each time a nucleotide is incorporated, the pyrophosphate containing the fluorophore is released so that the nascent strand synthesized is natural DNA, and no additional processing is needed before the next nucleotide can be incorporated. “I think one of the unique aspects of our technology is that there is one reagent-injection step, and then it is almost like watching a movie.” And she is quick to note that this single injection sets VisiGen’s FRET-based approach apart from other sequencing-by-synthesis approaches, such as tSMS, that require cyclical additions of reagents.

Hardin hopes that using their new sequencing approach on a massively parallel array will provide tremendous amounts of data very quickly. “Our target sequencing rate is a million bases per second with the goal of a genome in less than a day.”

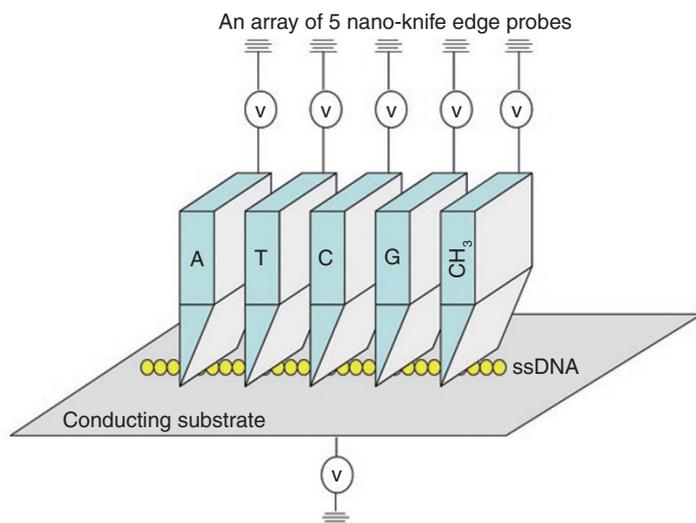
Sequencing by serendipity

Over the past 10 years Steven Block has built optical traps with the highest resolution on earth—capable of looking at the motion of a single protein. “We can now resolve the motion of a protein in real time down to a one angstrom level, the diameter of a hydrogen atom,” says Block. And using this system Block and his colleagues have been exploring the fundamental mechanics and properties of RNA polymerase, which can speed up, slow down, correct errors and even pause on a DNA template. Since the distance between the rungs of the DNA ladder is 3.4 angstroms, Block’s group can really watch all the steps taken by RNA polymerase. “We

were able to watch this thing move 1 base pair at a time.”

It did not take long for Block and his coworkers to realize something else about his new system. “If you can watch an enzyme move one base pair at a time, that immediately suggests a new way to sequence DNA not based on its chemistry but on its nanomechanics.” The method developed by Block and William Greenleaf¹ is in many aspects quite similar to traditional Sanger sequencing. In their system, RNA polymerase is placed in buffer containing normal concentrations of 3 of the 4 nucleotides, but a low concentration of the fourth. When the RNA polymerase moves along a DNA template stretched between two optical traps, the enzyme will move at a normal rate until reaching the first point on the template where the polymerase needs to put in a limiting nucleotide—and there it will pause briefly. Block says the pauses will occur everywhere that the limiting nucleotide is required, generating a record of movement with brief plateaus corresponding to all the locations of that nucleotide. “Now perform this four times with each nucleotide limiting in turn and you will be able to generate a ladder of all the bases; this is very similar to the way in which gel or capillary sequencing is done,” notes Block. The results of initial experiments were amazing and a bit surprising. “The first time we tried it as a quick experiment, we were able to get read lengths of 30 base pairs or so.”

Block thinks his system has headroom for longer read lengths in future versions. But he does not see this as the ‘next-next’ generation



Reveo is relying on voltage changes through ‘tuned’ nano-knife edge probes to sequence single molecules of DNA. (Courtesy of Reveo.)

system to deliver the coveted \$1,000 genome. “I do not think our method is something that you will be able to put in a box for \$50,000 in the next five years,” concedes Block. He notes that although it could work for niche applications, the method will be difficult to parallelize, and it takes time to prepare the sample and get things right with such a sensitive instrument. “Our lab is in the basement, and is sound-proofed and vibration isolated,” which he says is necessary as at one point they had to put parts of the instrument in a helium filled chamber because small fluctuations of air pressure in the room were bending beams of light minute amounts, influencing measurements.

The non-cutting edge

Elmsford, NY, USA-based Reveo was founded in 1991 with the goal of inventing solutions to difficult to solve problems. Now the company has turned its attention to DNA sequencing. “Current approaches to DNA sequencing have errors, are expensive and are not native,” says Sadeg Faris, chief executive officer of Reveo. So when Faris started to think about new ways to sequence DNA, he wanted a method that would use native DNA nondestructively and did not require extensive, costly imaging. This has led Reveo to develop a sequencing method using nano-knife edge probes.

In the system, multiple nano-knife edge probes pass over a stretched and immobilized strand of DNA in a channel that is 10 micrometers wide. Because the knife-edge nanoprobe is also 10 micrometers wide, the probe will always pass over the DNA. The key to this type of sequencing is that each nano-knife edge probe is ‘tuned’ to recognize one nucleotide and only that nucleotide. A unique voltage is applied to each nano-knife edge probe, and when the probe touches the corresponding nucleotide, electrons tunnel into the molecule, losing energy—which can be directly measured. In contrast, if a nanoprobe touches the wrong base virtually no tunneling current is detected. It is this selective excitation of molecular vibrations by electron tunneling that is the basis for the fast recognition of nucleotides using Reveo’s method.

And to avoid errors, each nucleotide is measured with 64 sets of the four nanoprobes. A fifth probe can be tuned to detect methyl groups, which Faris hopes will permit faster and cheaper exploration of the epigenome. “It is so cheap since the knife-edge nanoprobes will be disposable; it is error-free

and very fast.” Faris is so confident about this approach to single-molecule sequencing that in May 2007 Reveo officially announced its intention to compete for the Archon X-prize in genomics. And although he is still looking for additional financial support to build prototypes and advance the method, Faris is simply not worried about anyone else getting the X prize. He says: “We will win.”

The proof is in the pore

“When it comes to sequencing, nanopores hold great promise,” says Daniel Branton of Harvard University in Cambridge, MA, USA, “and have held great promise for such a long time that I think a lot of people are getting impatient.” Indeed, the basic idea for nanopore-based sequencing was first described more than five years ago, but as Branton

BOX 1 OF DESIGNS AND NANOPORES

One of the big issues for nanopore-based sequencing is slowing down the DNA as it transits through the pore. “If you put a moderate voltage across a nanopore, the DNA whips through so fast that you cannot resolve single bases,” says Steven Block. Dan Branton agrees that this is an issue that needs to be resolved for nanopores. “We reckon that the molecule has to be slowed down from the speed it normally travels through the pore.” And as Branton’s group explores carbon nanotubes to accomplish this slowing down, another group of researchers is trying to solve the problem by changing what it means to be a single nucleotide.

“The founder of the company realized that one of the challenges with respect to rapid sequencing was the size of the nucleotide,” says Robert Frostad, chief executive officer of LingVitae AS. Reasoning that a long molecule would take longer to go through a pore, researchers at LingVitae developed a method to convert DNA into a ‘magnified’ form called ‘design polymers’ using standard biochemical methods. “We magnified the DNA in terms of the length,” says Finn Terje Hegge, senior scientist at LingVitae AS; “each single

nucleotide can be represented by a block of nucleotides.” Synthesis of design polymers allows conversion of nucleotides into blocks of sequence of 4–40 nucleotides long in a bar-coded, binary format.

This magnification trick could allow sequence determination in nanopores where single-nucleotide detection is not possible. And Terje Hegge points out that some of the newly developed nanopores have been designed to do just this—determine blocks of bases passing through the pore rather than a single nucleotide in the pore.

“Our goal is to establish design polymers as the standard for the emerging generation of sequencing technologies,” says Frostad. And Terje Hegge thinks these polymers might work for more than just nanopore-based technologies. “I think design polymers can facilitate all the single-molecule sequencing technologies available now,” says Terje Hegge. Although the technologies would have to be altered to read design polymers rather than single nucleotides, he thinks this would be easy and the advantage of these magnified nucleotides could be huge.

notes, no sequencing has been accomplished directly using nanopores.

Although the idea behind nanopore sequencing is straightforward, the execution is turning out to be anything but that. Initially a small pore less than 2 nanometers in diameter is created in a very thin membrane. DNA is placed on one side of the membrane and current is applied across the nanopore. The early idea was to apply a current that would result in DNA entering the channel, and depending on the nucleotide passing through the pore there would be a change in the measured voltage across the channel. This voltage change could be determined as each individual base passes through the pore. “We were hoping this would demonstrate differences at the single-nucleotide level,” says Branton. But this did not turn out to be the case, so Branton and his colleagues have been working on new ways of measuring the current associated with a single nucleotide in the pore.

“Rather than use the current through the pore, our interest now is in measuring transverse currents across the pore diameter, namely tunneling currents, from probes on the pore perimeter,” says Branton. He says that the two major stumbling blocks to this new approach—achieving the same orientation for the each nucleotide in the pore and slowing down the translocation

rate of DNA through the pore—seem to have been overcome using carbon nanotubes as the electrical probes. Branton and his colleagues have found that each nucleotide interacts with the nanotube in a specific orientation and there is a strong interaction between the DNA and the carbon nanotube. His group hopes to take advantage of these properties of carbon nanotubes to finally achieve proper nucleotide orientation in the pore at a speed that can allow the measurement of the transverse current for an individual nucleotide.

Other groups, such as the companies Nabsys and LingVitae AS, are also exploring the potential of nanopore sequencing. Nabsys, located in Providence RI, USA is working closely with scientists at Brown University to develop what they call hybridization-assisted nanopore sequencing (HANS) for whole-genome sequencing. HANS relies on initially hybridizing a library of 6-mer probes to a fragmented, single-stranded genome. After hybridization, the DNA is driven through a nanopore and the resulting current is measured. Where the DNA is double stranded (a six-mer probe is attached) the flow of current will change. In this way a full-length probe map for the entire genome can be created and the sequence determined. LingVitae AS in Oslo is approaching the nanopore

problem from another angle—converting DNA into ‘design polymers’ to make sequencing easier (see **Box 1**).

The theoretical potential of nanopore sequencing to achieve the rapid, cost-effective sequencing is undeniable. “We are hoping for a parallelism of 100 nanopores as being reasonable and think this number would give a mammalian genome in 24 hours with the main cost being the chip itself, probably around \$1,000,” says Branton. And although it is now possible to manufacture this number of pores on a chip, the goal for Branton’s group in the coming years is proof-of-principle with a single nanopore.

But many wonder whether nanopore sequencing can ever reach its potential. “How they are going to get this to work with real DNA, in real time, remains to be shown,” says Block. But for Branton and others there is still great optimism. “My feeling is it has to work; there has been nothing in any of the investigations, in our lab or other labs around the world, that appears as a major stumbling block to further development.”

1. Greenleaf, W.J. & Block, S.M. *Science* **313**, 801 (2006).

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

SUPPLIERS GUIDE: COMPANIES OFFERING DNA SEQUENCING PRODUCTS

Company	Web address
23andMe	http://www.23andme.com
454 Life Science	http://www.454.com
ACGT	http://www.acgtinc.com
Affymetrix	http://www.affymetrix.com
Agencourt Biosciences Corporation	http://www.agencourt.com
Agilent Technologies	http://www.agilent.com
Applied Biosystems	http://www.appliedbiosystems.com
Beckman Coulter	http://www.beckmancoulter.com
Biotage	http://www.biotagebio.com
Brukner 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
Expression Analysis	http://www.espressionanalysis.com
GE Healthcare	http://www.gehealthcare.com
Genomic Solutions	http://www.genomicsolutions.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
Licor Biosciences	http://www.licor.com
LingVitae AS	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
Nimblegen Systems	http://www.nimblegen.com
Nugen	http://www.nugenic.com
Open Biosystems	http://www.openbiosystems.com
Operon Biotechnologies	http://www.operon.com
Pacific Biosciences	http://www.pacificbiosciences.com
Perkin Elmer Life Sciences	http://www.perkinelmer.com
Promega	http://www.promega.com
Qiagen	http://www.qiagen.com
Reveo	http://www.reveo.com
Roche Applied Sciences	http://www.roche.com
Saturn Biotech	http://www.saturnbiotech.com.au
Sequenom	http://www.sequenom.com
SeqWright	http://www.seqwright.com
Stratagene	http://www.stratagene.com
Taconic	http://taconic.transnetyx.com
TeleChem International	http://www.arrayit.com
Transgenomic	http://www.transgenomic.com
Visigen Biotechnologies	http://www.visigenbio.com