Magnetic sequencing

Sten Linnarsson

Single-molecule DNA sequencing takes an important step in a surprising new direction with a sequence-detection method based on magnetic tweezers.

© 2012 Nature America, Inc. All rights reserved.

Instead of directly observing the incorporation of fluorescently labeled nucleotides into DNA molecules, as has been done in many previous sequencing methods, Ding *et al.*[1] in this issue of *Nature Methods* measure the change in length of a DNA hairpin molecule tethered between a glass surface and a magnetic bead (**Fig. 1**).

As the beads are relatively large, they can be directly imaged using a simple microscope equipped with a video camera. When a magnetic field is applied, the hairpin is stretched and unfolds into a single DNA strand. When the magnet is removed, the hairpin refolds. As beads drift in and out of focus, diffraction rings are visible in the image, which allows the distance between bead and glass surface to be measured with great precision. Thus a very difficult optical problem (detecting light from single fluorescent molecules) is transformed into an easy one (detecting micrometer-sized beads under bright-field illumination). This remarkably simple optical setup can be used to detect length changes on the order of a few nucleotides, on DNA hairpins up to a kilobase long.

But how can the sequence of a DNA strand be inferred from changes in the length of a hairpin? The authors explore several methods[1]. The first is sequencing by hybridization: when they hybridize a probe to an open hairpin, complete refolding of the hairpin is prevented and the position of the hybridized probe can be inferred with great precision. Thus when a large set of probes is hybridized one by one, the sequence can be inferred from the overlapping sequences of probes. A smaller set of probes could be used to fingerprint a DNA sequence—for example, for pathogen detection.

Perhaps more promising is sequencing by ligation, a concept that is currently implemented commercially in Life Technology's SOLiD instruments. In this approach as implemented by Ding *et al.*[1], a primer is extended one step at a time by the ligation of a short degenerate oligonucleotide fragment. Extension is first attempted with a fragment starting with adenine, which can only be ligated if the next nucleotide on the opposite strand is a thymine. Then fragments starting with cytosine, guanine and thymine are attempted in turn, and the cycle is repeated. After each ligation, the magnetic field is released, and the length of the extended primer is measured. Upon ligation the primer is extended by seven bases, which is readily detectable as an increased distance between the surface and the magnetic bead. In preparation for the next cycle, the ligated fragment is cleaved at position 2 so that the next ligation is positioned just ahead of the previous one.

So-called 'next-generation' sequencers, which include instruments from Roche[2], Life Technologies[3] and Illumina[4], are all based on detection of clonally amplified DNA. For example, on the Illumina HiSeq, individual DNA fragments are amplified by PCR *in situ* to form micrometer-sized clusters containing thousands of copies of the initial fragment. This greatly simplifies detection as there is no need to detect single molecules. But the read length of such systems is limited by the phasing problem: in each chemistry cycle, some of the thousands of copies inevitably are not extended. Eventually the signal from such laggards overwhelms the true signal, and errors accumulate. The longest reads currently achievable are less than one kilobase and typically only around 100 bases.

In contrast, true single-molecule sequencing methods[5,6] are unaffected by phasing. Failed extensions can simply be ignored, and the error profile is independent of read length. Indeed, single-molecule fluorescence detection in the Pacific Biosciences RS instrument currently yields reads of up to several kilobases. But the drawback is a higher error rate, caused in part by the difficulty of capturing the fluorescence from a single molecule. The importance of the work by Ding *et al.*[1] is that to some extent they found a way to analyze single molecules without requiring any actual single-molecule optical detection. As a consequence, there is

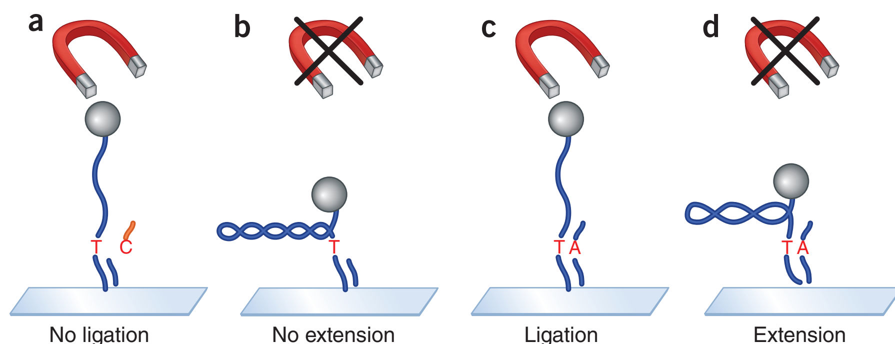

Figure 1 | Principle of sequencing by ligation. (**a**) A hairpin is stretched in a magnetic field, which allows the hybridization of a sequencing primer. A ligation fragment beginning with a cytosine (C) does not ligate. (**b**) The magnetic field is relaxed and the hairpin refolds. (**c**) The next ligation fragment, beginning with an adenosine (A), ligates to the primer. (**d**) When the magnetic field is relaxed, the hairpin refolds, and the distance between the bead and glass surface reveals a successful ligation.

Sten Linnarsson is at Karolinska Institute, Stockholm, Sweden.
e-mail: sten.linnarsson@ki.se

no need for expensive high-sensitivity optics used on other systems.

A conceptually different approach is nanopore sequencing, in which a single DNA strand is threaded through a tiny pore in a membrane, and the varying current through the pore is used to infer the sequence of bases. At the recent Advances in Genome Biology and Technology meeting, Clive Brown, chief technology officer at Oxford Nanopore, reported the sequencing of two viral genomes in the form of single multikilobase reads. Nonetheless, nanopore sequencing remains unproven until these results are published, and only the future will tell how far it can go.

What are the potential ultimate limits to magnetic sequencing? The number of beads that can be simultaneously monitored can probably approach densities equal to those of the best current amplicon-based sequencers (such as HiSeq 2000 at about 750,000 per square millimeter), being limited only by the size of the beads, the length of the tethered DNA template and the optical resolution limit.

The rate of imaging is limited by the mechanical movement of the bead, which is limited by the drag force. It would seem possible to achieve velocities of at least 10 micrometers per second (the speed of a swimming bacterium), so that ten open-close cycles can be measured in a second. The actual melting and annealing kinetics of the hairpin are unlikely to be limiting even at the millisecond scale. Thus an imaging rate of better than one image per second can potentially be achieved. It would seem possible to meet the specifications of current commercial 'next-generation' sequencers, with the potential for substantially longer reads. Nanopore sequencing, if successful, would be hard to match, however, as it promises reads up to 100 kilobases and throughputs greater than those of even the best current instruments.

However, although the work is a convincing proof of concept, it remains far from a practical, competitive sequencing method. Instead, its greatest virtue may be the many intriguing possibilities it immediately suggests. For example, could the positional accuracy be improved to detect single-nucleotide differences? This would open the possibility of sequencing by synthesis. Alternatively, would it be possible to monitor sequencing by ligation in real time, using fast enough modulation of the magnetic field? This would be a tantalizing prospect because it eliminates the need to repeatedly cycle reagents through a flowcell, thus simplifying instrument design. The four bases could be distinguished based on the length of the ligated probe.

Undoubtedly, many researchers will be stimulated to explore these and other developments of magnetic tweezers sequencing.

COMPETING FINANCIAL INTERESTS
The author declares no competing financial interests.

1. Ding, F. *et. al. Nat. Methods* **9**, 367–372 (2012).
2. Margulies, M. *et al. Nature* **437**, 376–380 (2005).
3. Rothberg, J.M. *et al. Nature* **475**, 348–352 (2011).
4. Bentley, D.R. *et al. Nature* **456**, 53–59 (2008).
5. Harris, T.D. *et al. Science* **320**, 106–109 (2008).
6. Eid, J. *et al. Science* **323**, 133–138 (2009).

© 2012 Nature America, Inc. All rights reserved. npg

Neighboring-gene effect: a genetic uncertainty principle

Anastasia Baryshnikova & Brenda Andrews

Deletion of a genomic locus may affect the function of neighboring loci, creating genetic uncertainty. Researchers now present a computational algorithm for identifying such neighboring-gene effects and improving the quality of functional annotations.

The budding yeast *Saccharomyces cerevisiae* remains a premier model eukaryote for post-genome-sequencing biology research, in large part because of the availability of an incredible array of different genome-wide reagents. In particular, the yeast deletion mutant collection has been available for a decade and has arguably revolutionized yeast genetics. The collection is a library of different yeast strains, each of which has a single open reading frame deleted and replaced with a dominant drug-resistance marker[1]. The deletion collection has been assessed for hundreds of phenotypes[2], which, in most cases, are directly associated with the deleted gene itself. In this issue of *Nature Methods*, Ben-Shitrit *et al.*[3] show that in some cases the deleted gene might not be directly responsible for the observed phenotype owing to an influence of the genetic perturbation on neighboring genes and describe a computational tool for addressing the problem.

Because genes occur in linear sequences on chromosomes, the change in gene sequence or alteration in DNA structure may influence not only a targeted locus of interest but may also perturb the function of neighboring sequences or genes. As a result, the phenotype observed in a deletion mutant strain may sometimes reflect the combined biological consequence of both inactivating the gene of interest and perturbing its neighbor(s). Thus, genetics based on precise gene deletions is associated with some uncertainty owing to a potential neighboring-gene effect (NGE).

Despite its possible impact on the accuracy of functional annotations, NGEs have not been systematically investigated so far. Ben-Shitrit *et al.*[3] describe an algorithm aimed at identifying erroneous functional annotations caused by an NGE and quantifying the extent of NGEs in budding yeast. Given a gene whose deletion causes a phenotype, NGE inference via a network-based approach (NIRVANA) can be used to explore adjacent genes to assess whether any of them are more likely to be responsible for the observed phenotype than the gene itself.

The NIRVANA algorithm uses protein-protein interactions to infer a functional relationship, independent of mutant phenotype. The idea is that the true causative gene should encode a protein that is connected by the shortest path on the protein-protein interaction network to a set of anchor proteins, which are known to contribute to the phenotype under investigation (**Fig. 1**). To select the most probable true effector, the algorithm also considers parameters such as network proximity to other selected genes and physical proximity to the deleted gene. Using this approach, the authors[3] estimated that ~12% of genes involved in telomere maintenance and ~10% of those associated with rapamycin sensitivity have been incorrectly associated with the observed phenotype, and this may indeed be

Anastasia Baryshnikova and Brenda Andrews are at The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada.
e-mail: brenda.andrews@utoronto.ca