blending of experimental and computational approaches, but its scale also hints at the magnitude of the work that remains. Surely, additional double dsRNA screens beyond the 12 performed will reveal more JNK regulators. Moreover, there is room to optimize experimental protocols for dsRNA knockdown (e.g., in sequence selection and delivery) and for measuring the degree of silencing of each gene and off-target effects. Finally, the studies of Bakal *et al.*[1] were all carried out with unstimulated *Drosophila* cells. Given JNK's known roles in maintaining cell, tissue and organism fidelity in the face of cellular stress, additional experiments will be needed to determine if and how the architecture of the JNK network is affected by stress.

JNK is known to exert its influence through numerous transcription factors, which in turn regulate a diverse set of target genes. How might the influences of the JNK regulators extend into the downstream transcriptional program? Capaldi *et al.*[2], working with budding yeast, suggest a way of tackling this question. The authors focused on building a quantitative model of the Hog1 MAPK-dependent pathway, which regulates the osmotic stress response. They performed gene expression profiling experiments on single-, double- and triple-knockout mutants of *hog1*, *msn2/4*, *sko1*, *sok2* and *hot1*—all known components of the Hog1 pathway.

Their key insight was to computationally tease apart the effects of knocking out single genes from what they call a "cooperative component," which quantifies whether two genes function independently, cooperatively (epistasis) or partially cooperatively (**Fig. 1b**). Perhaps most importantly, their method computes the cooperative component for each gene in the genome, providing a fine-grained view of how pairs of regulators interact functionally over the entire genome. The resulting regulatory map shows, for the first time, how the Hog1 MAPK signal propagates through different combinations of transcription factors to regulate distinct subsets of genes. By applying their analytical method to expression profiles of salt- versus glucose-induced osmotic stress, Capaldi *et al.*[2] also suggest how different branches of the regulatory hierarchy are used in a context-dependent manner to respond to different types of osmotic stress.

In contrast to the Bakal study, Capaldi *et al.*[2] use complete knockouts of components of a well-understood pathway. These precise deletions have the advantage of being invariant from cell to cell and assay to assay, but have the disadvantage that they cannot be considered essential genes. An important feature of both studies is that they rely on phenotypic readouts other than fitness to define pathway architecture. Comparing the networks derived from multiple phenotypic measures should greatly expand the 'dynamic range' of network biology.

As these two studies show, the analysis of multiple mutants and multiply perturbed cells provides crucial information for reconstructing the 'wiring diagram' of the cell, but the impact goes further. Drugs, for example, are simply cellular perturbations that can be conditionally applied. This concept was recently explored by chemically perturbing multiple mutants to enrich for genetic interactions and to order the components in a DNA repair pathway[4,5]. In a clinical context, a patient taking a therapeutic agent represents a unique genetic background combined with a chemical perturbation. Thus, it is conceivable that the approaches discussed here might eventually help guide the analysis of complex perturbation experiments in therapeutic settings.

1. Bakal, C. *et al. Science* **322**, 453–456 (2008).
2. Capaldi, A.P. *et al. Nat. Genet.* **40**, 1300–1306 (2008).
3. Hartwell, L. *Nature* **387**, 855–857 (1997).
4. St. Onge, R.P. *et al. Nat. Genet.* **39**, 199–206 (2007).
5. Lehar, J., Stockwell, B.R., Giaever, G. & Nislow, C. *Nat. Chem. Biol.* **4**, 674–681 (2008).

Sequencing in real time

Michael L Metzker

DNA synthesis by single polymerase molecules has been visualized at the speed of catalysis, heralding a new sequencing technology of unparalleled throughput.

DNA sequencing methods generally work by halting the process of copying the template strand in one way or another, using dideoxynucleotides (in Sanger sequencing), reversible terminators or natural nucleotides[1]. Now, a report by Eid *et al.*[2] in *Science* shows that sequence information can be obtained by continuous monitoring of DNA synthesis itself. This strikingly different approach, which records the incorporation of fluorescently labeled nucleotides into single primer strands in real time, promises to increase the speed and read-length of DNA sequencing and to open new avenues in basic research on DNA polymerases and nucleotide analogs.

Most of the next-generation sequencing systems, such as those from Roche/454 (ref. 3), Illumina/Solexa[4] and Life Technologies/Agencourt Personal Genomics[5], are not single-molecule methods as they rely on DNA amplification. A single-molecule technique was recently reported by Helicos Biosciences[6], but its dependence on reversible terminators limits it to the analysis of short DNA fragments. DNA polymerases perform optimally with nucleotide concentrations in the low micromolar range, a requirement that presents a challenge to single-molecule detection methods, which typically use fluorophores at pico- to nanomolar concentrations.

Eid *et al.*[2], of Pacific Biosciences, solved this problem with the company's zero-mode waveguide array[7], a nanostructured device that reduces the observation volume to the zeptoliter range—an improvement of more than three orders of magnitude over confocal fluorescence microscopy. At this superresolution volume, an estimated 0.01–1 molecule enters the detection layer by diffusion, providing a very low background signal and a signal-to-noise ratio of ~25:1.

To enable parallel sequencing, the authors used a chip with thousands of nanoscale wells containing an immobilized DNA polymerase bound to a primed DNA template to be sequenced (**Fig. 1**). To allow uninterrupted monitoring of nucleotide incorporation, they labeled nucleotides with four distinguishable fluorescent dyes on the terminal phosphate group rather than on the base, creating nucleotide analogs that apparently do not interfere with DNA synthesis by φ29 DNA polymerase, a highly processive, strand-displacing polymerase.

The residence time of the phospholinked nucleotides in the polymerase active site is governed by the rate of catalysis and is on the millisecond time scale. The bound nucleotide generates a recorded fluorescent pulse as no other fluorescent molecules are present in the detection volume of the zero-mode waveguide. Formation of a phosphodiester bond releases the fluorophore, which quickly diffuses away, reducing fluorescence to background levels and generating a natural, unmodified DNA product (**Fig. 1**). Translocation of the template marks an

Michael L. Metzker is at the Human Genome Sequencing Center and the Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza N1409, Houston, Texas 77030, USA.
e-mail: mmetzker@bcm.edu

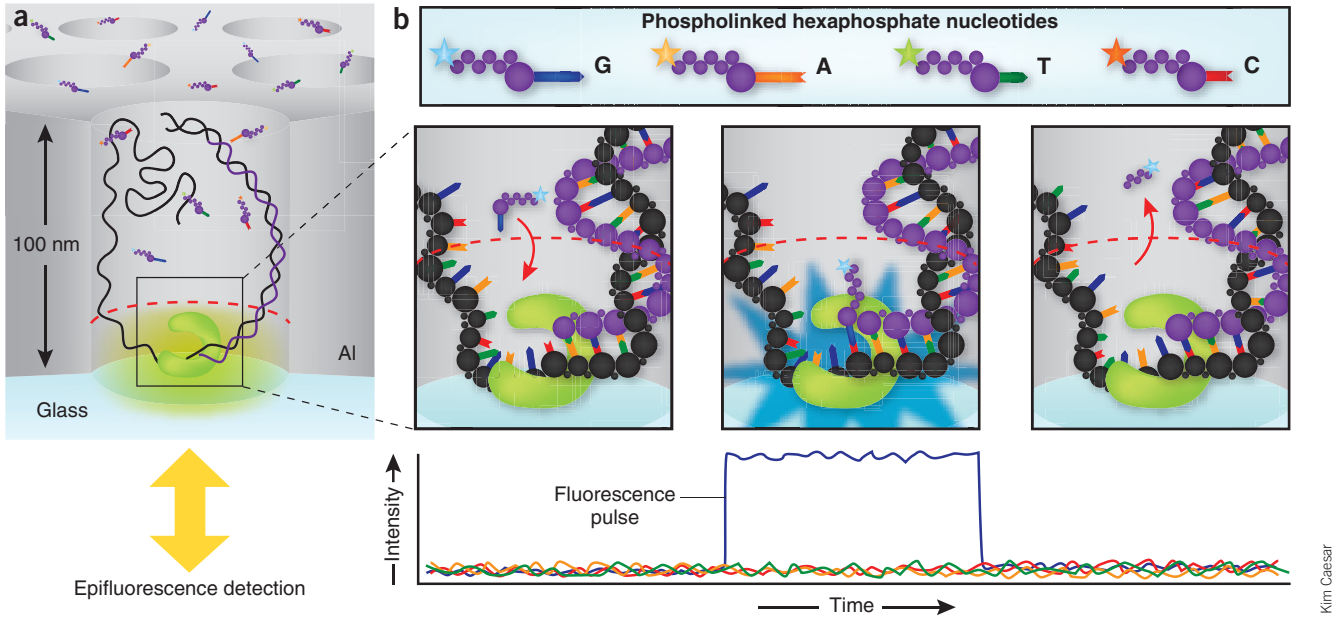

Figure 1 Real-time single-molecule sequencing. (**a**) A single φ29 DNA polymerase (green), immobilized at the bottom of a zero-mode wavelength nanostructure in an aluminium (Al) casing, adds fluorescently tagged nucleotides to a primed DNA template (black). (**b**) As light penetrates only the bottom 20–30 nm of the well, a fluorescent pulse is recorded only when the correct nucleotide is bound in the active site. Upon phosphodiester-bond formation and diffusion of the fluorescent polyphosphate byproduct out of the detection layer, a dark interphase period is observed. Translocation of the template prepares the DNA polymerase for the next incoming nucleotide. Fluorescent dyes are attached to the terminal phosphates of hexaphosphate nucleotides. Multiple lasers excite fluorophores only when they enter the region below the broken red line (zero-mode waveguide layer) and excitation and emission wavelengths pass through the same objective (epifluorescence detection).

interphase period before binding and incorporation of the next nucleotide. If processivity and fidelity are indeed unaffected by the phospholinked nucleotides, read-lengths of many thousands of bases should be possible.

Using a synthetic template with two dye-labeled phospholinked nucleotides, Eid *et al.*[2] demonstrated an impressive sequencing rate of approximately five bases per second, averaged from 740 single-molecule reads. By contrast, rates of at least 38 bases per second have been reported for ϕ29 DNA polymerase–mediated synthesis from primed M13 template[8]. The potential for long reads was also shown using a closed circular 72-bp template, from which polymerization was maintained for thousands of seconds, yielding several kilobases of synthesized DNA. The strand-displacing capability of the φ29 DNA polymerase enables resequencing of closed circular templates; multiple passes are likely to enhance sequence accuracy.

The team assessed accuracy by analyzing a known 150-bp linear template with a threshold detection method based on dye-weighted summation. When the read was aligned with a known reference, they identified 27 errors, including deletions, insertions and mismatches. This corresponds to a read accuracy of ~83% (131/158). Expanding the number of reads to 449 in a separate experiment revealed a lower accuracy of <80%. The authors attributed sequencing errors to very short interphase intervals (deletions), dissociation of the complementary nucleotide before phosphodiester-bond formation (insertions) and spectral misassignment of fluorescent dyes exhibiting significant emission overlap (mismatches). Because most errors are stochastic events, accuracy was improved to >99% by sequencing the same template molecule 15 times or more.

The Pacific Biosciences system can also measure enzyme kinetics of single polymerase molecules. The authors presented data revealing that φ29 DNA polymerase in this system occasionally pauses or exhibits distinct polymerization rates of approximately two bases per second and approximately four bases per second—modes that presumably interconvert. This provides fundamental insights into the kinetics of base incorporation (albeit in a surface-bound system) that would likely be overlooked in bulk solution experiments. It should also be possible to use this technology to study kinetic parameters affected by modified template bases and nucleotides, native and mutant polymerases, and polymerase activators and inhibitors.

Attaining the goal of the $1,000 genome will require substantial improvements in sequencing throughput, accuracy and cost. The array developed by Eid *et al.*[2] contains 93 × 33 wells, and the current method of populating them with DNA polymerases results in about a third of the wells being occupied with one polymerase molecule. Based on a daily throughput of roughly 400 kb, with 15-fold resampling of templates, the estimated throughput of the system is 27 kb of accurate consensus read data per well, or 30 mb per array. Although the authors speculate that increasing the density to 14,000 functioning zero-mode waveguides could produce a daily equivalent of 1× coverage of a human diploid genome, these calculations do not factor in the need for circular consensus sequencing.

In addition to increasing array density (to upwards of 100,000 wells per chip), other potential improvements include engineering more effective polymerases, creating brighter fluorescent dyes that reduce sequencing errors and developing methods that enable self-assembly of single polymerase molecules into each well and efficient circularization of large DNA fragments for consensus sequencing. Many of these challenges will likely be overcome in the near future, marking the entry of a formidable competitor in the next-generation sequencing market.

1. Metzker, M.L. *Genome Res.* **15**, 1767–1776 (2005).
2. Eid, J. *et al. Science* **323**, 133–138 (2009).
3. Margulies, M. *et al. Nature* **437**, 376–380 (2005).
4. Bentley, D.R. *et al. Nature* **456**, 53–59 (2008).
5. Valouev, A. *et al. Genome Res.* **18**, 1051–1063 (2008).
6. Harris, T.D. *et al. Science* **320**, 106–109 (2008).
7. Levene, M.J. *et al. Science* **299**, 682–686 (2003).
8. Soengas, M.S., Gutiérrez, C. & Salas, M. *J. Mol. Biol.* **253**, 517-529 (1995).