for assembling splicing sites. The ABI SOLiD and the Illumina GAIIX have not only increased the sequencing length to 50 and 75 bases, respectively, but have also developed methods for sequencing from both ends of the cDNA fragments to help in connecting more distant exons.

Other challenges of RNA-seq are how to distinguish the various start and end sites of RNAs. It is becoming evident that there are often multiple overlapping RNAs encoded from the same genome region, and intron-derived RNAs are recycled to produce functional ncRNAs such as microRNAs. Another source of complexity comes from the secondary processing of mRNAs, which produces shorter, likely functional, RNAs. Thus, protein-coding genes are associated with a plethora of short ncRNAs, including short RNAs associated with promoters[13], transcripts arising around termination sites and even exons. A fraction of these RNAs are produced by a novel cleavage and recapping mechanisms, resulting in capped RNAs that start in the middle of coding exons or in untranslated regions. These naturally truncated RNAs are likely to be ncRNAs that overlap larger mRNAs[13]. Another complication arises from the broad nature of many promoters[14], which produce various capped RNAs from multiple transcription start sites. Technologies that identify the cap structure in such mixtures are needed to distinguish the RNA fragments obtained by RNA-seq. At present RNA-seq does not perform well at unambiguously identifying transcription start sites, and RNA-seq protocols need improvement to simultaneously decipher the long, short and capped RNAs so the RNAs' function can be assessed.

Some of the third-generation sequencers such as those from Pacific Biosciences and Oxford Nanopore—which will be able to read thousands of nucleotides[15] of single cDNAs—may ultimately meet these challenges: their long sequences will quantitatively represent complete RNAs, and the use of tags and linkers that mark cap sites and other modifications will allow an all-in-one determination of transcriptome structure, including start and termination sites and the mapping of regulatory elements such as promoters. The accurate sequence of coding sequences will also help directed cloning of open reading frames in experimental verification of alternative splice isoforms[16,17].

Although many challenges are ahead, the direction is becoming clearer, and I am beginning to wonder if the dark age of the transcriptome is giving way to rays of light.

1. Liang, F. *et al. Nat. Genet.* **25**, 239–240 (2000).
2. Carninci, P. *et al. Science* **309**, 1559–1563 (2005).
3. ENCODE Project Consortium *et al. Nature* **447**, 799–816 (2007).
4. Carninci, P. *et al. Genome Res.* **13**, 1273–1289 (2003).
5. Kapranov, P. *et al. Science* **316**, 1484–1488 (2007).
6. Harbers, M. & Carninci, P. *Nat. Methods* **2**, 495–502 (2005).
7. Sultan, M. *et al. Science* **321**, 956–960 (2008).
8. Pan, Q., Shai, O., Lee, L.J., Frey, B.J. & Blencowe, B.J. *Nat. Genet.* **40**, 1413–1415 (2008).
9. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. *Nat. Methods* **5**, 621–628 (2008).
10. Cloonan, N. *et al. Nat. Methods* **5**, 613–619 (2008).
11. Tang, F. *et al. Nat. Methods* **6**, 377–382 (2009).
12. Faulkner, G.J. *et al. Nat. Genet.* **41**, 563–571 (2009).
13. Fejes-Toth, K. *et al. Nature* **457**, 1028–1032 (2009).
14. Carninci, P. *et al. Nat. Genet.* **38**, 626–635 (2006).
15. Turner, D.J., Keane, T.M., Sudbery, I. & Adams, D.J. *Mamm. Genome* **20**, 327–338 (2009).
16. Djebali, S. *et al. Nat. Methods* **5**, 629–635 (2008).
17. Salehi-Ashtiani, K. *et al. Nat. Methods* **5**, 597–600 (2008).

Engineered fluorescent proteins: innovations and applications

Michael W Davidson & Robert E Campbell

Despite expansion of the fluorescent protein and optical highlighter palette into the orange to far-red range of the visible spectrum, achieving performance equivalent to that of EGFP has continued to elude protein engineers.

Evolving proteins, evolving tools

During the past decade and a half, intrinsically fluorescent proteins have been under intense evolutionary pressure for 'fitness', not in the wild, but rather for utility in live-cell imaging experiments. This unnatural course of evolution has occurred on the benches of protein engineers around the world who have helped to drive progress in the ever-expanding repertoire of fluorescence imaging technologies.

An underlying theme that has guided advancements in fluorescent protein engineering is that, all other factors being equal, redder is better. It is generally accepted that excitation with longer-wavelength light entails less phototoxicity for the cells or tissue being examined and decreased autofluorescence and scattering. These desirable factors mean that red-shifted fluorophores generally provide improved contrast (owing to decreased background fluorescence) and superior performance in whole-organism imaging (owing to higher tissue 'transparency'). Early efforts to engineer red-shifted *Aequeorea victoria* GFP (avGFP) variants led to the development of enhanced GFP (EGFP) and yellow fluorescent proteins with emission maxima at approximately 507 nm and 529 nm, respectively (versus 508 nm for wild type)[1].

For a time, however, it appeared that fluorescent protein engineering had hit a 'yellow' wall in efforts to red-shift fluorescence emission. Fortuitously, this barrier had already been surmounted by natural evolution, as was revealed in October 1999 with a report that the *Discosoma* sp. mushroom anemone harbored a fluorescent protein homolog, commonly known as DsRed, emitting in the orange-red region (583 nm)[2]. Counterbalancing this favorable shift to the red were several undesirable properties, including oligomerization, 'contamination' by a green component and sluggish chromophore development, which dampened some of the initial enthusiasm.

The discovery of DsRed (and other Anthozoa fluorescent proteins of various hues) had a twofold impact on the

Michael W. Davidson is at the National High Magnetic Field Laboratory and Department of Biological Science, Florida State University, Tallahassee, Florida, USA. Robert E. Campbell is at the University of Alberta, Department of Chemistry, Edmonton, Alberta, Canada.
e-mail: robert.e.campbell@ualberta.ca