TOOLS IN BRIEF

BIOCHEMISTRY

Hyper-accurate Cas9

The rising popularity of the CRISPR–Cas9 system for genome editing and regulation goes hand in hand with calls for increased specificity and reduced off-target binding. Chen *et al.* use single-molecule Förster resonance energy transfer to investigate the mechanism by which two recently described highly specific Cas9 variants discriminate between on- and off-target sequences. They find that a noncatalytic domain in the Cas9-recognition domain recognizes complementarity between the single-guide RNA and the target DNA duplex and undergoes a conformational change that facilitates the activation of a Cas9 nuclease domain. In the high-fidelity Cas9 variants, this conformational switch is only triggered by a perfect match between target DNA and guide RNA. Based on this insight the researchers then design a hyper-accurate Cas9 (HypaCas9) with high specificity and no loss in on-target activity.

Chen, J.S. *et al. Nature* http://dx.doi.org/10.1038/nature24268 (2017).

MICROBIOLOGY

Large-scale microbial genome reconstruction

The growing piles of metagenomic data, generated by shotgun sequencing of environmentally sampled microbial mixtures, present a tantalizing target for data mining. Parks *et al.* take a deep dive into over 1,500 publicly available metagenomes and surface with a huge trove of newly characterized microbial genomes. The authors use the CLC *de novo* assembler to generate long contiguous sequences, which are then binned based on similarity and taxonomic compatibility and pruned for quality. The nearly 8,000 resulting draft-quality metagenome-assembled genomes (MAGs), which they call the Uncultivated Bacteria and Archaea (UBA) data set, are more than 50% complete, with nearly half of the MAGs over 90% complete. The selected metagenome data sets were mainly collected outside of the relatively well-characterized human host context, and thus the MAGs generated in this analysis expand known phylogenetic diversity by over 30%.

Parks, D.H. *et al. Nat. Microbiol.* http://dx.doi.org/10.1038/s41564-017-0012-7 (2017).

SENSORS AND PROBES

Improved split fluorescent proteins

Split fluorescent proteins are used for a variety of purposes, including protein labeling and bimolecular fluorescence complementation assays that probe whether proteins of interest are in close proximity. Split fluorescent proteins are typically composed of two separately expressed fragments of a circularly permuted version of a fluorescent protein that come together to form an intact fluorescent protein. Feng *et al.* have developed improved split fluorescent proteins based on mNeonGreen and mCherry. Compared with a commonly used split GFP, split-mNeonGreen2$_{1-10/11}$ has an improved ratio of complemented signal to background. Compared with split-sfCherry2, split-sfCherry2$_{1-10/11}$ shows ten-fold higher brightness. The researchers also developed a photoswitchable form of split-sfCherry2$_{1-10/11}$ that enables super-resolution imaging of split fluorescent proteins. They demonstrate the power of these tools to study the abundance of Sec61B in different parts of the endoplasmic reticulum.

Feng, S. *et al. Nat Commun.* **8**, 370 (2017).

GENOMICS

Alternative splice atlas

Much of the diversity in human gene expression is due to alternative splicing of pre-mRNAs and resulting RNA isoforms. To get a comprehensive picture of such splice events in vertebrate species, Tapial *et al.* quantified major splice types such as those involving intron retention and alternative 5′ and 3′ splice choices. The resulting Vertebrate Alternative Splicing and Transcription Database (VastDB; http://vastdb.crg.eu/) lists tens of thousands of alternative splice events in different tissues, cell types and developmental stages in human, mouse and chicken. Users can now easily determine which exons and introns in a gene are differentially spliced in a given cell type and draw conclusions about possible function. For example, genes with multiple major isoforms are enriched in functions involved with transcriptional control.

Tapial, J. *et al. Genome Res.* **27**, 1759–1768 (2017).