METHODS IN BRIEF

BIOINFORMATICS

A shortcut to high-dimensional data

Big data comes at a price, and it is a glutton for computer memory. Rather than using data compression to solve this problem, Cleary *et al*. have come up with ways to generate roughly equivalent genomic data from a smaller set of gene expression measurements. The authors borrow an approach called compressed sensing from the field of signal processing, in which random composite measurements are used to infer the activity of gene modules. The expression of unmeasured genes can then be estimated from the inferred module activities. Their blind compressed sensing with sparse module activity factorization (BCS-SMAF) software can successfully recover the transcriptome from 100 composite measurements without requiring training data. The approach may be useful for screens, single-cell profiling and other large-scale expression studies.

Cleary, B. *et al*. *Cell* **171**, 1424–1436.e18 (2017).

PROTEOMICS

Nascent proteome labeling *in vivo*

Methods for tagging newly synthesized proteins allow researchers to fish proteins out from cells for identification using techniques such as mass spectrometry. The mutant methionyl-tRNA synthetase (MetRS) system allows methionine tRNA to be charged with the unnatural amino acid analog azidonorleucine, which contains a bioorthogonal moiety that can be tagged via click chemistry with a purification handle. By combining this technique with cell-type-specific promoters, it is possible to identify nascent proteins in specific classes of cells. Alvarez-Castelao *et al.* developed a transgenic mouse line to perform *in vivo* cell-type-specific nascent protein labeling experiments using Cre recombinase to induce expression of the mutant MetRS in specific cell types. With this approach, the researchers analyzed protein expression in excitatory principle neurons and in Purkinje neurons.

Alvarez-Castelao, B. *et al. Nat. Biotechnol.* **35**, 1196–1201 (2017).

GENOMICS

Pooled screens expand to new phenotypes

Pooled screens have accelerated the rate of genetic discovery by allowing large libraries of diverse genetic variants to be screened together rather than individually. However, pooled screens typically require phenotypic enrichment or, if developed in single-cell format, use RNA sequencing as the only readout. Two methods now perform pooled screens for dynamic, image-based phenotypes. Single cells, each bearing a genetic variant or perturbation that is associated with an expressed sequence barcode, are imaged and then genotyped by reading out the barcode with multiplexed fluorescence *in situ* hybridization. Emanuel *et al*. use their approach to identify improved YFAST fluorescent proteins from among 60,000 variants screened in bacteria, and Lawson *et al*. develop dynamic u-fluidic microscopy-based phenotyping of a library before *in situ* genotyping (DuMPLING) to show that CRISPR-induced perturbations in multiple bacterial clones can be screened for gene-regulatory phenotypes.

Emanuel. G. *et al*. *Nat. Methods* **14**, 1159–1162 (2017); Lawson, M.J. *et al*. *Mol. Syst. Biol.* **13**, 947 (2017).

MASS SPECTROMETRY

False discovery rate estimation for metabolomics

In both untargeted mass-spectrometry-based proteomics and metabolomics studies, experimental tandem mass spectra are matched to reference library spectra for interpretation. Manual checking of these matches becomes impractical in large-scale studies; as a result, statistical methods for estimating the false discovery rate (FDR) of this matching process have become essential for proper reporting of proteomics results. However, equivalent approaches for assessing FDR in metabolomics studies have lagged far behind. Scheubert *et al.* now report FDR estimation methods for metabolomics based on empirical Bayes statistics and on the target–decoy approach (which is widely applied in proteomics for FDR estimation). They applied these approaches to 70 publicly available metabolomics data sets and found that scoring parameters must be adjusted for each data set. Statistical validation will help advance the reliability and robustness of metabolomics results.

Scheubert, K. *et al. Nat. Commun.* **8**, 1494 (2017).