Making sense of tumor mutations

Recent efforts of consortia such as The Cancer Genome Atlas have produced large-scale DNA sequence data from human tumors. Whereas somatic mutations are highly heterogeneous and may overlap only rarely between tumors, the networks that are disrupted in a given tumor type are more likely to be shared. Ideker and colleagues exploit this property to describe a network-based strategy for tumor stratification. They demonstrate that, by integrating somatic mutation data with gene networks, they can assign tumor samples to robust subtypes that are biologically or clinically meaningful. Also in this issue, Lopez-Bigas and colleagues report IntOGen-mutations, a Web platform that summarizes tumor mutation data from thousands of samples in multiple projects and identifies potential driver mutations. Finally, Mills, Liang and colleagues describe The Cancer Proteome Atlas, a portal for the visualization and analysis of functional proteomic data from thousands of tumors and cell lines.

Article p1108, Brief Communication p1081, Correspondence p1046, News and Views p1077

The hierarchical interaction score

Quantitative measurements of various aspects of biological systems—gene expression, for instance, or phenotypes due to systematic perturbation—require statistical analytical methods to fully extract the information present in the resulting large data sets. Pelkmans, Snidjer and colleagues present the hierarchical interaction score (HIS), a statistic that determines hierarchical relationships between variables in large data sets. They show that the HIS outperforms correlation coefficients in identifying functionally related genes from RNA interference screens. Interactions identified by the HIS show relatively low overlap with those inferred using existing methods, suggesting that it may allow the study of unexplored biology.

Brief Communication p1089

Image-based transcriptomics

Single-molecule fluorescence in situ hybridization (smFISH) provides information on gene expression that is complementary to single-cell RNA-seq data, but smFISH has been restricted to imaging a handful of transcripts at a time. Pelkmans and colleagues transform smFISH into a true expression-profiling methodology with modifications that allow hundreds of different transcripts to be imaged and counted in parallel in thousands of cells, at single-molecule resolution. They use brighter hybridization probes that enable robust imaging of cells in multiwell plates combined with automated robotics and software to localize transcripts relative to cellular features. The ability to analyze thousands of cells lends statistical power for highly accurate measures of low-level gene expression. The authors find that transcript localization can actually be more informative than transcript abundance in grouping genes by function to understand their biology.

Article p1127

Structured illumination microscopy in an instant

Methods to image below the diffraction barrier and attain resolutions better than 200 nanometers are typically limited to small fields of view or are very slow in comparison to conventional imaging methods, especially if illumination intensities need to be minimized for live-cell imaging. York and colleagues report an analog implementation of structured illumination microscopy that dispenses with the digital merging of multiple structured images to remove this speed restriction while allowing volumetric imaging at depth for large fields of view. They image structural details of moving mitochondria in two colors, cytoskeletal details in flowing blood cells 20 micrometers deep in zebrafish cranial vessels at 37 hertz, and the remodeling and growth of endoplasmic reticulum at 100 hertz.

Article p1122

Getting more out of single-cell RNA-seq

Sequencing the minute amount of RNA present in a single cell presents many challenges. Heisler, Marioni and colleagues address the high degree of technical noise in single-cell studies that can mask biological variation. They show that this noise can be modeled by quantifying spike-in RNA sequences, and they present a statistical method to identify the most variable transcripts that contribute to heterogeneity in populations of cells that are otherwise considered identical. Sandberg and colleagues provide a number of optimizations and improvements to their protocol for sequencing full-length transcripts based on template switching. Their new Smart-seq2 protocol improves sensitivity, coverage and accuracy, reduces bias and does not require a commercial kit.

Brief Communications p1093, p1096