How to analyze RNA-seq data

High-throughput sequencing of transcripts (RNA-seq) has a lot of potential for identifying and quantifying genes and their isoforms. But realizing this potential is made difficult by a confusing array of computational tools for data analysis. In a Review in this issue of *Nature Methods*, Manuel Garber and colleagues discuss the concepts behind the three main categories in RNA-seq data analysis: read mapping, transcript reconstruction and quantification of expression. They introduce representative tools for each category and describe their main applications with pros and cons. Although the current crop of tools will not solve every RNA-seq data analysis problem, this Review will help researchers make more informed choices as to which program is best suited for them.

**Review p469**

Finding structural RNA modules from sequence alone

Understanding RNA structures can yield insight into RNA functions, but RNA structure-function relationships are just beginning to be sorted out. Three-dimensional (3D) structural RNA modules, defined as ordered non-Watson-Crick pairs, are found throughout the phylogeny. Almeida Cruz and Westhof describe a computational tool called RNA 3D modules detection, or RMDetect, which helps identify these 3D structural RNA modules from sequence analysis alone. Currently, RMDetect is trained to identify four common modules, including the G-bulge loop, the kink-turn, the C-loop and the tandem GA-AG loop. Almeida Cruz and Westhof demonstrated the performance of RMDetect for identifying 141 known modules and 21 new instances in a set of 1,441 alignments. A tool for building additional models corresponding to new modules, called RNA 3D modules builder or RMBuild, is also presented.

**Article p513**

STORM picks up speed

Localization-based methods of super-resolution imaging such as photoactivated localization microscopy (PALM) and stochastic optical reconstruction microscopy (STORM) are generally very slow. Two-dimensional (2D) imaging speeds of 30–300 seconds per reconstructed image frame are typical. Three-dimensional (3D) imaging is even slower, and resolution typically suffers. Through judicious choice of switchable probes and imaging conditions, Zhuang and colleagues report 2D and 3D STORM imaging at 2 hertz and ~1 hertz, respectively. Resolution varied between 25 and 50 nanometers depending on the imaging axis, probe and frame rate. Dual-color 3D live-cell STORM imaging is also reported. Unlike scanning-based methods, the field of view can be as large as the entire cell, but the number of snapshots is limited, so the frame rate must be decreased if long imaging durations are needed.

**Article p499**

Managing the data cascade

As automation becomes increasingly common in biological research, the amount of data generated even in individual laboratories is rapidly expanding. What is more, experimental design often changes frequently, driven by previous results and by the biological question at hand. Systems for computer-readable and flexible data management—storing, accessing and sharing both the data and the metadata describing the experiment—have become limiting. In this issue, Sorger and colleagues describe a data structure, semantically typed data hypercubes (SDCubes), for flexible data management. SDCubes combine hierarchical data format 5 (HDF5) and extensible markup language (XML) file formats and are applied, in this paper, to high-content imaging data in a study of the cellular response to varying doses of ligands and drugs.

**Article p487, News and Views p463**

Mutagenic protein traps in zebrafish

Annotation of the vertebrate genome needs tools that permit the study of gene expression and function *in vivo*. Ekker and colleagues report a Tol2 transposon–based mutagenesis system, termed RP2, in the zebrafish. Insertion of RP2 into a zebrafish gene results in abnormal splicing and robust abrogation of endogenous transcript expression, permitting the study of gene function. Simultaneously, a fusion to a fluorescent reporter protein is generated, elucidating the expression pattern of the trapped gene *in vivo*. The mutant alleles are reversible with Cre recombinase or with morpholinos that block key splice sites. The authors use the RP2 system to generate a collection of 350 freely available mutant fish lines.

**Article p506, News and Views p467**