

Cracking the histone code

The protein spools around which the thread-like genomic DNA is packed, the histones, are very important in regulating gene expression. The N-terminal tails of histones contain many post-translational modifications, and it has long been suggested that the various combinations of modifications that appear at different times constitute a 'histone code'. This has not been an easy code to crack, however. Toward this goal, Kelleher and colleagues describe a new approach, in which intact N-terminal histone tails are first separated into simpler mixtures by hydrophilic interaction chromatography, then subjected to 'top-down' mass spectrometry, which aids in the preservation of sequence coverage and post-translational modifications. In total, they identified 150 distinct 'codes' for histone H3.2—the most comprehensive histone analysis to date.

Brief Communication p487, News & Views p480

achieve polytene chromosome banding patterns at unprecedented microscopic resolution. Moreover, these methods are compatible with fluorescence *in situ* hybridization and immunocytochemistry, meaning that genes and chromosome binding proteins can be localized with greater precision to a high-resolution cytological map.

Brief Communication p483

Getting a handle on sequence variants

Traditional mass spectrometry-based proteomics experiments require the use of databases containing peptide spectra, allowing users to map the peptides in the experimental sample to their known parent proteins. But sequence variants, such as coding single nucleotide polymorphisms (cSNPs), are not well-represented in most databases. Mann and colleagues have modified the European Bioinformatics Institute International Protein Sequence (IPI) database to maximize the potential to identify sequence variants. Without substantially increasing database size, they added information curated for cSNPs, sequence conflicts and N-terminal peptides, using as spacers letters that do not denote amino acids to allow search engines to handle the modified database. The new database for human and mouse proteins, MSIP1, is freely available.

Correspondence p465

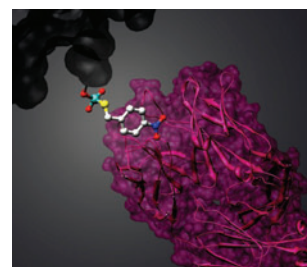
Testing times for metagenomics

The use of shotgun sequencing to extract genomic sequences from complex microbial communities has the potential to provide a host of biological insights. Unfortunately, assembly and analysis of this wealth of raw sequence data requires sophisticated algorithms, and the performance of these methods has been hard to evaluate. This lack of information has also impeded the development and testing of new methods. To remedy this, Mavromatis and colleagues at the US Department of Energy Joint Genome Institute created three simulated complex data sets as test standards for these algorithms and used them to evaluate existing analysis programs. Such work will be invaluable for advancing a field that holds so much potential.

Article p495, News & Views p479

Follow the thiophosphate

Those bringers of phosphorylation, the protein kinases, have a huge role in signaling in the cell. Explaining the specific roles of individual kinases, however, not to



mention developing effective inhibitors, remains quite a challenge as a result of the similarities in kinase activities. Shokat and colleagues report a method to identify the direct substrates of kinases engineered to accept a bio-orthogonal ATP γ S analog. These kinases, rendered 'analog specific', thiophosphorylate their substrates; alkylation yields thiophosphate esters. The researchers developed a new antibody to recognize the thiophosphate ester, and showed that they could purify and identify thiophosphorylated substrates of an analog-specific kinase expressed at endogenous levels.

Article p511

Producing polytene patterns with precision

The giant polytene chromosomes in larval tissue of *Drosophila melanogaster* have been used for several decades to physically map genes, breakpoints, deletions and binding proteins along the chromosome, as well as in the study of boundary elements and of structural changes associated with active transcription. But the manual 'squash' techniques presently in use for spreading these structures are inefficient and not very reproducible; typically, more than half of slides must be discarded. Using precision devices, Novikov and colleagues reproducibly