Capturing the metabolome
Metabolites are tricky molecules to study. Whereas proteins and transcripts can be predicted to a certain extent from the genome sequence, metabolites are synthesized by enzymes and have no direct connection to the genome of an organism. In addition, metabolites are structurally more diverse than genes and proteins and can exhibit a wide range of properties. Enrichment methods exist for the isolation of metabolites from complex mixtures, but these target a small number of metabolite types. So, Benjamin Cravatt and Erin Carlson at the Scripps Research Institute developed a general tagging strategy to enrich and analyze many types of metabolites. With the method, called metabolite enrichment by tagging and proteolytic release (METPR), the researchers captured and studied metabolites with amines, thiols, ketones/aldehydes, and carboxylic acids.

A set of reactive groups targeting specific metabolite classes were synthesized and bound to resins with a small, arginine-containing peptide. In the METPR protocol, trypsin is added after metabolites are captured on the functionalized resins. Trypsin cleaves at the arginine residue; this event results in the release of the metabolite with a small tag. The rest of the peptide, including the arginine, remains on the resin. Finally, tagged metabolites are analyzed by LC/MS.

Cravatt and Carlson tested the METPR strategy by adding a simple mixture of small molecules to the capture probes. Each probe isolated the expected type of molecule. In addition, molecules present at low abundances were preferentially isolated with METPR. The researchers could detect some small molecules at a concentration of 100 pmol/10^7 cells, which is ~5–20-fold less than that detected with typical LC/MS workflows that do not include an enrichment step. The efficiencies were up to ~50% for the various probes that were tested.

In another set of experiments, METPR performance was demonstrated with breast cancer cell lines. Again, the expected classes of metabolites were captured. When the scientists performed MS/MS to determine the structures of the captured molecules, they discovered that the tag did not interfere with this process. The researchers also treated breast cancer cells with an antioxidant and compared the metabolomes of these cells with those of untreated cells. Differential profiles were observed for metabolites of all four classes with METPR. Cravatt and Carlson say that because different metabolites can be detected with METPR and conventional techniques, these methods are complementary. (Nat. Methods 2007, 4, 429–435)

Pinpointing phosphorylation sites in serine/threonine clusters
To identify phosphorylation sites, researchers typically use positive-ion MS/MS with collision-induced dissociation. In this process, however, modified peptides often undergo the neutral loss of phosphoric acid at the expense of backbone fragmentation. Therefore, little sequence information is obtained.

Although more gentle fragmentation techniques, such as electron capture dissociation and electron transfer dissociation, have been developed, they require expensive instrumentation or highly charged precursor ions. So, Wolf Lehmann and co-workers at the German Cancer Research Center investigated an alternative method called negative-ion MS/MS. They discovered that the method is well suited for determining phosphorylation sites within stretches, or clusters, of ≥2 serines and/or threonines.

Initial tests were conducted with two isomeric peptides that each contained two consecutive serines in the middle of the peptide. In one peptide, the first serine was phosphorylated, and in the other peptide, the second serine was phosphorylated. With the conventional method of positive-ion MS/MS, the spectra for both peptides were nearly identical, and the bond between the two serines was not broken. Therefore, the peptides could not be distinguished. With negative-ion MS/MS, however, z ions were formed. Two of the z ions likely represent fragments resulting from a cleavage between the adjacent serines. The relative abundances of these ions were different for each peptide; these differences allowed Lehmann and co-workers to correctly pinpoint the locations of the phosphorylations. Similar results were obtained for singly and doubly deprotonated peptide ions. The substitution of serines with threonines greatly reduced the extent of formation of the diagnostic ions, and substitution of serines with tyrosines completely suppressed z ion formation.

Finally, the researchers synthesized two additional peptides that included four consecutive serines with phosphorylations on two residues. One peptide had phosphorylations on the first and third serines, whereas the other had phosphorylations on the third and fourth serines. Again, the locations of the phosphorylations could be reliably determined with negative-ion MS/MS. (Anal. Chem. 2007, 79, 3476–3486)
p53 targets identified with proteomics methods
Because the p53 gene is mutated in about half of all human tumors, much attention has been focused recently on identifying its downstream effectors. To find these effectors, most researchers have only examined cellular mRNA levels or p53’s ability to bind to particular DNA sequences in the genome. Of course, genes and transcripts don’t tell the whole story—posttranslational modifications, protein stability, and cellular localization also could be important. So, Klas Wiman and colleagues at the Karolinska Institutet, Uppsala University (both in Sweden), the University of Schleswig–Holstein (Germany), and ETH Zurich conducted a proteomics investigation to discover additional protein targets of p53. In the study, the researchers observed that several expected and unexpected proteins were differentially regulated when p53 activity was induced in cancer cells.
HCT116 colon carcinoma cells (p53+/+)
were treated with mitomycin C (MMC), a DNA-damaging agent, to activate p53. Cells with a homozygous deletion of the p53 gene (p53+/−) also were exposed to MMC. By immunostaining and western blot analysis, Wiman and colleagues confirmed that p53 was induced in p53+/+ cells after 8 hours (h) of MMC treatment, but the protein was not detected in the p53−/− cells. Lysates and gel images, peak lists, search results, and summary pages. The data are available at www.proteomecommons.org.

Mass spectral library
Phillip Andrews and co-workers at the University of Michigan have compiled a standard reference set of spectra for spectral matching and to help researchers develop new algorithms. This publicly available library, called the Aurum Dataset, contains mass spectra of 246 purified proteins that were analyzed on a MALDI TOF/TOF instrument. Only those proteins that were deemed to be of high purity were included. In addition to spectra, scientists can access gel images, peak lists, search result files, decay database files, FASTA files, and summary pages. The data are available at www.proteomecommons.org.

Cleaning up oil with a bacterium
Nearly 20 years ago, the Exxon Valdez oil spill off the coast of Alaska brought the problem of oil pollution into the spotlight. Oil spills still are difficult and costly to clean up with conventional techniques, so researchers are investigating alternatives such as bioremediation, or the use of microorganisms to degrade pollutants. One bacterium that could help is Geobacillus thermodenitrificans, which was discovered in a deep oil reservoir in northern China. The bacterium can degrade long-chain alkanes, which are major components of crude oils. To better understand how G. thermodenitrificans works its magic, Lei Wang and co-workers at Nankai University (China) studied its genome and proteome. They discovered that the bacterium expresses several proteins that allow it to survive in oil-rich environments.

The G. thermodenitrificans genome, which is composed of a chromosome and a plasmid, was sequenced. In total, the bacterium has ~3500 predicted open reading frames. Genes involved in the uptake and usage of organic acids, aromatic compounds, and ammonia, which are present in oil reservoirs, were discovered. Denitrification and fermentation genes also were identified in the organism’s genome.

By real-time RT-PCR, Wang and co-workers observed that the transcription of a gene encoding a putative alkane-activating enzyme increased 120-fold when the bacteria were grown in medium containing crude oil rather than sucrose as the carbon source. The purified gene product activated long-chain alkanes in vitro, and a plasmid carrying the gene allowed a mutant bacterial strain to grow in media containing long-chain, but not short-chain, alkanes. The researchers called this gene long-chain alkane degradation A (LadA).

Finally, the proteomes of G. thermodenitrificans grown in media containing sucrose or long-chain alkanes were compared. Proteins were analyzed by 2DE and MALDI TOF/TOFMS. LadA and other enzymes implicated in alkane degradation were up-regulated in bacteria grown in the presence of alkanes. The researchers say that these results suggest that G. thermodenitrificans is a good candidate for the bioremediation of oil pollution. (Proc. Natl. Acad. Sci. U.S.A. 2007, 104, 5602–5607)
from both cell types were obtained at various time points (0, 8, 16, and 24 h) and analyzed by 2DE. A total of 115 spots were significantly changed between the samples. The researchers identified 55 differentially expressed proteins by MS/MS; 42 were up-regulated, and 13 were down-regulated. These proteins are involved in several processes, including mRNA processing, protein translation, and metastasis.

To validate these findings, the researchers performed western blots with antibodies against seven of the proteins. Although most of the results were similar to those obtained with 2DE, some differences were observed. For example, the increases in the protein levels of lamin A/C and heterogeneous nuclear ribonucleoprotein K in p53+/+ cells treated with MMC were not as large on the blots as they were on the 2DE gels. The scientists also treated two other cancer cell lines with MMC and obtained similar western blot results. Because of this finding, the investigators say that the results were not unique to HCT116 cells.

With an algorithm called p53MH, Wiman and colleagues discovered that 24 of the up-regulated proteins and 8 of the down-regulated proteins were transcribed from genes with at least one putative p53-binding site. These proteins are probably induced transcriptionally by p53. However, the rest lack these motifs, so they may be regulated by a post-translational mechanism. (Proc. Natl. Acad. Sci. U.S.A. 2007, 104, 5401–5406)

Targeting accessible proteins on tissues for therapies
Tumors are composed of rapidly dividing cells, which are targeted by many chemotherapeutic drugs. However, some populations of normal cells, such as hematopoietic progenitors and skin cells, also divide often. So, Philippe Kischel and colleagues at the University of Liège, University Hospital of Liège (both in Belgium), and ETH Zurich reasoned that a better way to target therapies is to bind drugs or other bioactive molecules to antibodies that are directed against proteins that are specific to the tumor microenvironment. Therefore, the researchers modified a recently reported protocol for identifying accessible proteins in perfusable organs and applied it to biopsy material. With the new method, they discovered a possible therapeutic target in human breast cancer tissue.

In previous work, the researchers injected a reactive biotin ester solution into the blood vessels of mice and vascularized human organs to identify easily accessed tumor proteins. In the current work, however, tissues such as biopsies that lack extensive vasculature are soaked in the reactive biotin solution. Biotin labels all accessible proteins, such as those in the extracellular space and membrane proteins. Then, the labeled proteins are purified with a streptavidin resin, digested, and analyzed by multidimensional protein identification technology (known as MudPIT).

With the modified method, Kischel and colleagues compared the profile of accessible proteins from biopsies of breast cancer tissue with that of noncancerous tissue that was adjacent to the tumor. Although most of the labeled proteins were extracellular or associated with the plasma membrane, some were intracellular. The investigators say that these inaccessible proteins may have been labeled inadvertently because they had leaked out of cells when the tissues were sliced, or these proteins may have bound strongly to membrane proteins. Alternatively, the biotin solution may have penetrated the cells.

Of the proteins that were identified, one was validated with additional tests. Versican, an extracellular matrix protein that is secreted by stromal cells, was identified only in the tumor samples. To verify this finding, the researchers stained tumor and nontumor tissue with an antiversican antibody and observed staining only in tumor tissues. To ensure that this protein was not expressed elsewhere in the body in healthy tissues, they also searched for versican expression with tissue microarrays. The protein was not present in most tissues, but it was moderately present in the placenta and the central nervous system, which is protected from proteins in the circulation by the blood–brain barrier. Therefore, versican is a possible target for antibody-based cancer therapies. (Proteomics 2007, 7, 1188–1196)

NOMIS
To remove platform-specific sources of variability in metabolomics data, Matej Orešič and colleagues at VTT Technical Research Centre of Finland and the Turku Centre for Biotechnology (Finland) developed NOMIS (Normalization using optimal selection of multiple internal standards) to help researchers choose the best standard compounds. (BMC Bioinformatics 2007, 8, 93)

UniRef
The presence of redundant protein sequences in a database can slow down searches and confuse the interpretation of results. To reduce the redundancy of protein sequences in the UniProt Knowledgebase (UniProtKB) and certain UniProt Archive records, Baris Suzek and co-workers at Georgetown University Medical Center have developed UniProt reference clusters (UniRef). Three versions are available. UniRef100, UniRef90, and UniRef50 contain clusters of sequences that are 100%, 90%, and 50% identical, respectively. UniRef records include general cluster information, such as the UniRef ID and common taxonomy, and cluster membership information, such as the source database and sequence length. UniRef databases are updated every 2 weeks in XML and can be downloaded at ftp.uniprot.org. The researchers caution that because the databases are actually regenerated every time an update is released, some clusters and sequences may change. (Bioinformatics 2007, doi 10.1093/bioinformatics/btm098)