


[The Scientist](#) » [Magazine](#) » [Lab Tools](#)

Name That Metabolite!

A guided tour through the metabolome

By Jeffrey M. Perkel | July 1, 2013

0 Comments Like 0 +1 0 Link this Stumble Tweet this



© NODEROG/ISTOCKPHOTO.COM

The suffix '-omics' is synonymous with Big Data. It's simply a given that when one researcher publishes an omics data set, be it genomic, transcriptomic, or proteomic, other researchers will be able to take a crack at it, too.

Metabolomics data are no different. Researchers regularly report on dysregulated metabolites in disease and development. In one 2012 study, Scripps Research Institute metabolomics expert [Gary Siuzdak](#), with his then-postdoc Gary Patti, used mass spectrometry to identify dysregulated metabolites in a rat model of neuropathic pain. Of the tens of thousands of spectral peaks they examined, 733 were "significantly dysregulated" compared to control animals. The researchers eventually homed in on sphingomyelin-ceramide metabolism, one of the pathways these peaks represented (*Nat Chem Biol*, 8:232-34, 2012).

But what about the compounds they didn't pursue? Other pain researchers might want to see what other compounds were dysregulated in Patti's and Siuzdak's data sets. Or they might be interested in knowing whether their particular metabolite of interest also happens to exist in the neural tissues the team studied. Can a non-mass spectrometrists reanalyze metabolomics data sets and see what else they might say?

In a word: Yes. But it isn't easy. Mass-spec data are notoriously complex, and even experts can struggle with data interpretation. Caveat emptor. Still, if you want to take a shot, we can help. *The Scientist* spoke with Siuzdak, [Patti](#) (now at Washington University in St. Louis), and other metabolomics experts about their favorite data-analysis tools. This is what they said.

1. WHERE CAN I FIND METABOLOMICS DATA SETS?

Nucleic acids are archived at [GenBank](#) and protein-structure data at the [Protein Data Bank](#) (PDB) of the Research Collaboratory for Structural Bioinformatics (RCSB). Metabolomics data, though, have no standardized home. Authors are under no obligation to deposit their data in a specific repository or in a particular format. When the human serum metabolome was worked out in 2011, the authors of that study set up their own database ([www.serummetabolome.ca](#)) to house it. But that database doesn't include the actual raw spectral data—the only way to get that is basically to say "Please."

In the old days (that is, a little more than 2 years ago), researchers who wanted to share data essentially had to mail off a DVD, hard disk, or flash drive containing the raw data files. "These data sets could easily be several gigabytes large," says Patti. Even then, each mass spectrometer churned out data in its own proprietary format, meaning a dedicated [Bruker Daltonics](#) user might have trouble opening and interpreting data collected on, say, a [Thermo Fisher](#) mass spec, and vice versa.

Today, researchers can use XCMS Online ([xcmsonline.scripps.edu](#)). About 6 months after the site launched, Siuzdak and Patti wrote that they developed the metabolomics data-analysis suite "in response to the growing interest from the general scientific community for a user-friendly program to process untargeted metabolomics data" (*Anal Chem*, 84:5035-39, 2012). Anyone can open an account, and access is free. There are even a few sample data sets available to play with, including comparisons of the chemical composition of Coke and Pepsi and of two different kinds of beer. More than 2,350 users have registered to date, of which 2,000 or so use it consistently, Siuzdak says.

XCMS Online can accept data in some 11 different file formats, converting them into a standardized format for analysis and archiving purposes. Once uploaded, data sets can be shared specifically with other researchers, even those who lack the tools and expertise to handle the raw spectral data itself.

Data sharing and standardization, Siuzdak says, have proven to be surprisingly popular, and not only for

Follow The Scientist



Subscribe!



Print or Digital

• iPad
• Kindle
• Tablet



Stay Connected with The Scientist

- The Scientist Magazine
- The Scientist Careers
- Neuroscience Research Techniques
- Genetic Research Techniques
- Cell Culture Techniques
- Microbiology and Immunology
- Cancer Research and Technology

Popular Posts

1. [Bacterial DNA in Human Genomes](#)
2. [Silver Boosts Antibiotic Efficacy](#)
3. [Molecule Wards Off Mole-Rat Cancer](#)
4. [Image of the Day: Brain Fibers](#)
5. [Image of the Day: Palm Seed](#)

Current Issue



[View the July 2013 contents.](#)

newbies—more than half the system's users take advantage of data sharing. Even experienced mass spectrometrists can stumble a bit when examining data produced by other groups on unfamiliar equipment.

"Some of these [mass spectrometry] platforms are so convoluted that you need to take a multiday short course to get a sense of what's going on," Patti says. "That's one of the reasons why we developed this resource—it's very intuitive."

2. IS MY METABOLITE IN THE DATA SET?

Interrogating a metabolomics data set for a particular compound is simple in XCMS Online. Register for a free account, then under the tab "Public Shares," open the data set called Coke vs. Pepsi.

Clicking the button labeled "Browse Result Table" calls up a spreadsheet listing all the peaks identified in the data set, with column headings for fold change, p-value, mass-to-charge ratio (m/z), and more—some 2,553 features in all. You can search this list by name using the "Quick Compound Search" box; a search for "caffeine" pulls up two features, one of which is the sodium adduct of 8-methylcaffeine. On the right side of the page are plots showing the difference in abundance between the two samples, plus a table of putative chemical identifications.

Clicking a METLIN ID in the results table at the bottom right of the page (e.g., for 8-methylcaffeine, 84980) will take you to the corresponding METLIN record. METLIN is a database, also developed in Siuzdak's lab, which lists the structure, chemical properties, and purchasing information for some 76,000 metabolites. Among its niftiest features is a plot of spectral data found at the bottom of selected records, which Siuzdak's team has meticulously collected from more than 11,000 compounds to date. Mousing over the peaks on this graph pulls up the team's experimentally observed fragmentation patterns and predicted structures—information that can be useful if you wish to verify that a particular peak is actually what you think it is.

3. WHAT METABOLITES DIFFER MOST BETWEEN TWO DATA SETS?





There are two ways to answer this question in XCMS Online. Return to the data table for Coke vs. Pepsi (if necessary, click "Clear" next to the Quick Compound Search box). Click the magnifying glass icon at the top left of the table. Here you can build complex filters to show, say, only features that differ by more than 10-fold between the two data sets and that have a p-value less than or equal to 0.001. (Doing that for Coke vs. Pepsi reduces the list to just 32 features.)

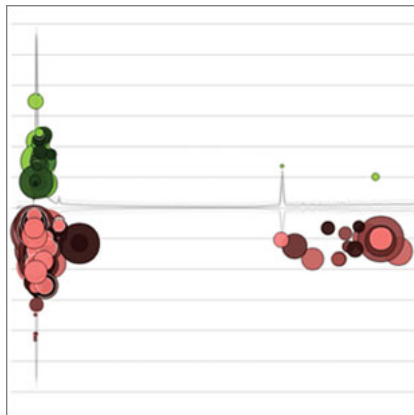
For a more visual approach, click "Return to Job Summary" at the bottom of the table. Next, click the job's "Interactive Cloud Plot" button at the top of the page.

A Cloud Plot provides a data-dense overview of two metabolomics data sets. Features are represented as colored bubbles plotted by m/z , ion intensity, and chromatographic retention time. (By default, green bubbles are upregulated and red downregulated, but these colors are configurable.) A bubble's diameter indicates the fold change between two samples, and color intensity corresponds to p-value. Thus, the most significantly different features between two data sets in a Cloud Plot will be those that are largest and darkest. A control panel at the top left of the page allows you to filter the display by p-value and fold change. (See plots at right.)

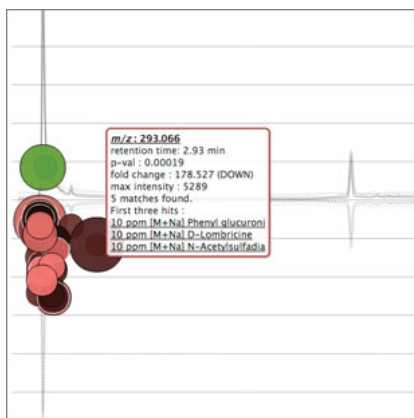
To see what a particular feature corresponds to, mouse over the bubble to view the top three matching METLIN identifications (if available; features that contain a METLIN ID are outlined in black). Selecting one of the identifiers pulls up the corresponding METLIN record. As with the table view, you can also see

Subscribe to RSS feed

-  All
-  The Nutshell
-  News & Opinion
-  Careers



[View Full Size JPG](#) | [PDF](#)
XCMSONLINE.SCRIPPS.EDU



CLOUD DATA: Interactive cloud plot representation of the Coke vs. Pepsi data set. Each spectral feature is displayed as a circle. A circle's diameter indicates how much that feature's abundance differs between the two data sets, while the color intensity indicates its statistical significance. Circles above the zero-line are upregulated (green), while downregulated features (red) are shown below the line. A black border around a circle indicates that a METLIN match for this compound has been identified. At bottom, the plot has been filtered to show only features that differ by 10-fold or more with p-values < 0.001.

[View full size JPG](#) | [PDF](#)
XCMSONLINE.SCRIPPS.EDU

at left the corresponding spectrum and box-and-whisker data plot.

4. WHAT IF I CAN'T IDENTIFY A METABOLITE?

Join the club. Few metabolites in metabolomics data sets correspond to the “canonical” biochemical pathways found in textbooks, Patti says, and many have never been seen before. This, combined with the fact that metabolites, unlike proteins, do not assemble from a small number of known and easily differentiated building blocks, means researchers can rarely name every molecule they see.

The Coke vs. Pepsi data set is loaded with compounds unknown in METLIN, but cells are no better. Leslie Silva, a postdoctoral fellow and metabolomics researcher at MD Anderson Cancer Center in Houston, who has been using XCMS Online since late 2011, says she typically identifies maybe 25 percent of the features she finds in her studies of ovarian cancer. “There are always a lot of features that can’t be identified,” she says. In those cases, Silva tries to work out the compound structure by fragmenting it in a tandem mass spectrometer and then figuring out how the pieces fit together. Still, she admits, “there are a lot of times where you throw your hands up in the air and have no idea what the compound is.”

Still, there are online resources that can help. MassBank (www.massbank.jp), a public repository of metabolite mass-spec data, currently contains more than 39,000 spectra on some 15,336 standard reagents, says Takaaki Nishioka of the Nara Institute of Science and Technology in Japan, who is MassBank’s PI. The service is akin to METLIN in that it can match *m/z* values with metabolites. But it also includes a “Metabolite Prediction” tool that can try to work out a compound’s structure by matching its fragmentation data against other known compounds. Similarly, METLIN’s “multiple fragment” search (or “similarity search”) tries to identify molecules with fragmentation patterns similar to a set of peaks input by the searcher.

5. CAN I DO A PATHWAY ANALYSIS?

Once you’ve identified a set of interesting metabolites, it’s helpful to find out if they are all related—say, if they are in the same or connected pathways—as Siuzdak and Patti did in their study.

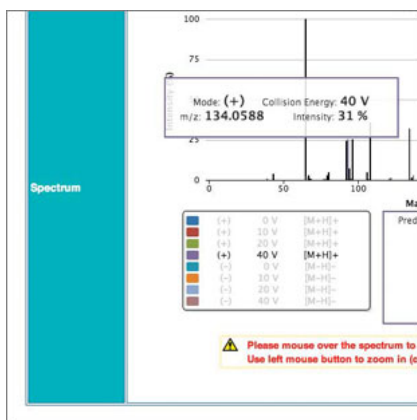
At the moment, such “pathway analyses” aren’t possible in XCMS Online. In part, Patti says, that’s because so many chemicals are unknowns. “It’s not really a limitation of the database in any way, it’s just a limitation of what we know about biochemistry.”

Commercial pathway analysis tools do exist—Silva uses a text-mining tool called Pathway Studio, from Elsevier—but for a quick-and-dirty (read: free) start, try PubMed. Or, view the metabolite in its pathway context on free data services like MetaCyc (metacyc.org), KEGG (www.genome.jp/kegg/), the Human Metabolome Database (www.hmdb.ca), Reactome (www.reactome.org), and the LIPID MAPS database (www.lipidmaps.org), any of which can provide structure, biochemical context, and relevant enzymes and genes.

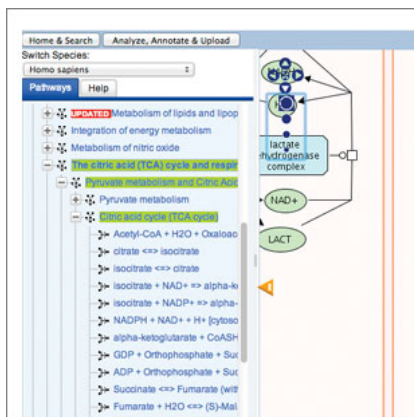
A MetaCyc search for caffeine, for instance, finds nine related pathways plus a variety of associated enzymes including caffeine dehydrogenase, caffeine demethylase, and caffeine synthase. Selecting “caffeine biosynthesis I” under pathways lays out the biosynthetic pathway for caffeine, as well as relevant genes, enzymes, and literature references.

Such information can help build new hypotheses. But before you actually start planning experiments, a word of warning: “Even experts have disagreements about how to work with [metabolomics] data,” says Akos Vertes, founder and codirector of the W. M. Keck Institute for Proteomics Technology and Applications at George Washington University.

Vertes (who uses the data-analysis software that came with his Waters mass spectrometer in his own work) has seen firsthand what happens when neophytes first start playing with metabolomics data—after



CHEMICAL ID: The METLIN record for N-acetylsulfadiazine gives the compound’s molecular formula, structure, and mass, and notes that the compound is used as an antibiotic for burn wounds. Predicted structures for the various mass-spec fragments appear as you mouse over them (arrow). View full size [JPG](#) | [PDF](#)
XCMSONLINE.SCRIPPS.EDU



ADDING CONTEXT: Pathway analysis in Reactome.org. Shown here is a portion of the citric acid cycle focusing on oxaloacetate. View full size [JPG](#) | [PDF](#)
WWW.REACTOME.ORG

all, he has students. "It's not point-and-click," he says. "It requires very judicious choices and decisions about what to take seriously and what to ignore. And we always go back to the original data and double-check."

So, he advises that newbies contact the researcher who generated the data in the first place, and think about collaboration. "There are too many factors that are not obvious for somebody who did not collect the data," he says. "I would say, pick up the phone and talk to the person and see where that takes you."

Tags

[XCMS Online](#), [techniques](#), [METLIN](#), [metabolomics](#) and [mass spectrometry](#)

0 Comments Like 0 +1 Link this Stumble Tweet this

Add a Comment



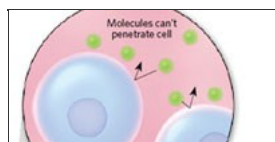
You

Sign In with your LabX Media Group Passport to leave a comment



Not a member? [Register Now!](#)

Related Articles



Narrow Straits

By Ruth Williams

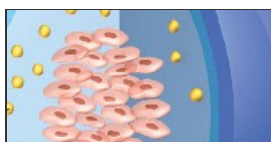
Transfecting molecules into cells is as easy as one, two, squeeze.



Transcriptomics for the Animal Kingdom

By Amy Maxmen

Using RNA-seq to study nonmodel organisms



Dead or Alive?

By Ruth Williams

Scientists create nontoxic pH nanosensors to assess viability of transplanted therapeutic cells.

TheScientist

[Home](#) [News & Opinion](#) [The Nutshell](#) [Multimedia](#) [Magazine](#) [Advertise](#)
[About & Contact](#) [Privacy Policy](#) [Job Listings](#) [Subscribe](#) [Archive](#)



© 1986-2013 The Scientist

Now Part of the LabX Media Group: [Lab Manager Magazine](#) | [LabX](#) | [LabWrench](#)