

METABOLOMICS

Taming Metabolites

Metabolomics studies inch scientists ever closer to understanding phenotype. But to really make progress, pioneers are working on improving the technology and analytical tools of the field.

BY CIARA CURTIN

A 1966 *Biochemical Journal* article recently unearthed by Gary Siuzdak's lab at the Scripps Institute looks to be the first metabolomics experiment — though they certainly didn't call it that then. In it, researchers from Baylor College of Medicine describe using gas-liquid chromatography to separate metabolites from urine and tissue extracts. They also add that GC-coupled mass spectrometry gives a “diagnostic tool of great power” — something today's researcher already knows. But also in that article, C.E. Dalglish *et al.* grumble about overlapping peaks, resolving those peaks, and the lack of a database housing known metabolites. Sound familiar?

Today's metabolomics researchers have the same gripes, but are better poised to do something about them. The technology isn't very fast or robust as compared to other big players in systems biology, and identifying all the metabolites in a sample can be nearly impossible. At each step, from sample preparation



GARY SIUZDAK

to data analysis to metabolite identification, metabolomics as a field is still trying to resolve robustness, identification, and analysis issues that the other disciplines within systems biology have already overcome. Metabolomics, though, is working hard to live up to its potential and join the ranks of the high-throughput fields. Researchers are capitalizing on the track record of NMR and the sensitivity of mass spectrometry to increase the number of metabolites detected and then identified through new databases.

“When it comes to metabolomics, it's just a plethora of different tech-

niques and technologies and protocols, and some it is reproducible and some of it isn't. It's a bit of a Wild West, but there's a bit of consolidation and the trends are there,” says the University of Alberta's David Wishart, who heads up the Human Metabolome Project. “I think in a year or two it will be a more mature field with more consolidation and more consistency and more robustness.”

Detection

In metabolomics, the divide has been between nuclear magnetic resonance and mass spectrometry. Both tools can identify a sample's metabolite population to varying degrees of success. More and more, though, researchers are combining the approaches to take advantage of their strengths while minimizing their weaknesses. At the same time, people are using new tools and methods for both separation and detection to take a gander at their metabolome of choice.

NMR, which dates back to the 1940s, has long been used by chemists to identify molecules in a sample. As a tool, it gets kudos for being stable, reliable, and robust. “In NMR, you can analyze the same sample today and this time next year and get a very similar result,” says Warwick Dunn at the University of Manchester.

The tool has a variety of roles in the lab. For one, it can be used to get a first look at what the metabolome contains. “NMR is our main technique, which we would use first as sort of a survey technique,” says John Lindon, a professor at Imperial College London.

Or it can be used for metabolite profiling. “NMR is very good because it tells you exactly which position in a molecule contains a ¹³C or ¹⁵N, whereas mass spec only tells

Things to consider before purchasing a Next Gen System.

SCALABILITY

Throughput advances without costly upgrades

THROUGHPUT

More than 6 GB and 240M tags per run for cost-effective, large-scale resequencing and tag applications

FLEXIBILITY

Two independent flow cells to run multiple applications in a single or staggered run

ACCURACY

System accuracy of greater than 99.94% for the power to detect rare, causative SNPs without high false positive rates

MATE PAIRS

Libraries with inserts ranging from 600 to 10,000 bp for resolution of structural variation

DATA SETS

Publicly available human datasets for accelerating biological discovery and analysis tool development

SUPPORT

Trusted service and support team available when and where you need them

you how many positions are labeled, not which ones,” says Andrew Lane, a professor at the University of Louisville’s James Graham Brown Cancer Center.

And, of course, NMR has its downside. “The big knock about it is that it’s not very sensitive,” says Wishart. Indeed, according to Wishart and Siuzdak, an NMR-based characterization of a tissue sample or biofluid yields a little more than 50 molecules, but looking at that same sample with mass spec methods can yield hundreds or even thousands of molecules, depending on the chromatography technique coupled to the mass spec. Some scientists use NMR for surveying, and then apply mass spec for a more targeted analysis.

Manchester’s Dunn focuses on mass spec — particularly liquid-chromatography mass spec — and he works on developing and optimizing methods to use it in metabolomics. While mass spec may be able to see more metabolites than NMR, it has its own drawbacks, primarily reproducibility. In his lab, Dunn says, two separate sample sets might give 50 interesting metabolites, but only 10 of them overlap.

“In an ideal world, you’d use both technologies because, in any analytical technology, there is some bias in what it can detect, whether it be the type of metabolite it can detect or the sensitivity, for example,” Dunn says.

In particular, he works on increasing the reproducibility of mass spec by using an automated closed-loop strategy that has minimal human intervention. Over many iterations, the Robot Chromatographer, as his team calls it, initializes the instrument settings and then changes them as it cycles through looking for the optimal settings. When used on GC-TOF mass spec, Dunn and his colleagues increased the number of peaks seen by three-fold.

Newer technologies are also coming onto the scene to topple NMR, LC/MS, and GC/MS from the top spots in

metabolomic technologies. “The mass spec technology is wonderful now. The robustness is great, especially the new time-of-flight and quadrupole time-of-flight mass spectrometers. They have improved dramatically in the last couple of years,” says Siuzdak.

Not only is the detection step being improved, but advances are also coming along on the separation side. Ultra high-performance liquid chromatography came on the scene a few years ago, using higher pressures and smaller particle sizes to increase resolution and sensitivity, allowing scientists to detect even more metabolites. “The more things you can see, the greater overview you can get of the biology of the system,” Dunn says. He’s not the only one looking into UPLC: Lindon and his group have begun to couple it with time-of-flight mass spectrometry.

Another separation approach that is catching on is HILIC. Hydrophilic interaction chromatography allows researchers to detect more of the small, hydrophilic molecules that often are removed during a wash or that come out at the very beginning of the separation. Siuzdak is particularly intrigued by this method. “I’ve been recently surprised by some of the results that we’ve been getting that’s allowed us to certainly see new things,” he says. He is currently using HILIC to try to detect new molecules in knock-outs. “It’s just another window into these samples,” Siuzdak says.

Deconvolution

The data that comes out of the end of NMR or mass spec is a mess of peaks and spectra. Making sense of all that can require some serious analysis, though some old hands can recognize NMR peaks just by looking at them. Most scientists rely on software packages to resolve the curves and deconvolute the data into something resembling a list of metabolites. “If you use chromatography on very complex

samples like urine or plasma or something, there would be so many small molecules and metabolites eluting at pretty much similar chromatographic times that you get overlapped peaks, which makes them difficult to quantify and it makes it difficult to identify what it was,” says Henrik Antti, an associate professor at the University of Umeå in Sweden. Different research groups are developing new and better software to help deconvolute what’s in a metabolome.

At Imperial College, Lindon and his colleagues have developed and are using a statistical analysis method to identify NMR peaks. “It’s not like a gene chip where you have one spot equals one gene,” he says. “Here, a molecule, a metabolite will give many peaks on the NMR spectrum. We can use what we know about NMR to identify where those come from.”

Building on a previous tool, called TOCSY (for total correlation spectroscopy), Lindon’s team made a tool called STOCSY. This new method takes advantage of the correlation between peaks in NMR spectra — that multiple peaks can come from the same molecules and always occur in proportion. As an example, Lindon points to lactate, which has two NMR peaks — one from the methyl group and one from the CH group. Since NMR detects the hydrogen atoms of these groups, these two peaks will always be in a proportion of three to one. “We can use that statistical correlation to prove those two peaks are linked across hundreds or even thousands of samples,” Lindon says. This relationship can help researchers work out which peaks of an NMR spectra go with which and, Lindon adds, help them identify potential biomarkers.

At Scripps, Siuzdak and his colleagues developed their own tool to

analyze mass spectrometry data for metabolite profiling. Their XCMS is an open-source data analysis software package for LC/MS data that not only peak-picks, according to Siuzdak, but uses endogenous metabolites found in all the datasets as internal standards and aligns the peaks based on the retention time. Then, XCMS looks through its analysis and find the peaks that change between the dataset that are statistically relevant. “So now you have a set of molecules, typically, that look very interesting,” says Siuzdak. He and his colleagues also recently came out with XCMS² for MS/MS data.

For researchers blending NMR and mass spec data, Lindon and his colleagues have also been working on a tool that bridges the NMR-mass spectrometry divide. Their statistical heterospectroscopy, or SHY, works to put NMR and UPLC/MS data from the same samples together by analyzing signal intensities from the molecules as detected by the different methods. “You get a bit of information from the mass spec and a bit of information from the NMR, you can put the two together to identify molecules,” says Lindon.

Databases

With the molecules in hand, the identity of the metabolites can begin to be uncovered, though it isn’t always possible when they don’t correspond to a known metabolite. “There are still a lot of unknowns in terms of compounds that people see or identify. If you were to take a sample from a person or a plant and use our standard libraries of known endogenous

metabolites, you still won’t be able to identify all the compounds, or all the peaks,” Wishart says.

A few database projects — including efforts by Wishart and Siuzdak — are attempting to index and curate all the known metabolites. “Unlike in proteomics or

“We can use that statistical correlation to prove those two peaks are linked across hundreds or even thousands of samples.”

genomics where we can say we know all the amino acids and all the bases and therefore the library or the alphabet is known, the alphabet isn’t really fully known for all the things that we expose ourselves to,” Wishart adds.

Starting in January of 2005, Genome Canada funded the Human Metabolome Project; part of its mandate was to catalogue and consolidate all naturally occurring metabolites. It contains about 2,500 metabolites, culled from the literature and confirmed with NMR, LC/MS, or GC/MS, as well as from the group’s own experimental data.

Siuzdak and his colleagues are working on Metlin, a depository for mass spectral metabolite data. It currently contains about 23,000 molecules, and Siuzdak says they are adding more to it constantly. The 1966 paper, says Siuzdak, said the main problem with using GC/MS was that there are so many molecules that are unknown and there’s no comprehensive database. “What happens since then is now there’s a database that has well over 10,000 molecules in it,” Siuzdak says.

While these projects and others,

such as Riken's SpinAssign and the Madison Metabolomics Consortium Database, have made progress in cataloguing metabolites, estimates place the number of metabolites in the tens of thousands. The databases have a long way to go before they can be considered anything close to exhaustive. "[Metabolomic databases] still have a ways to go. They are not as robust as Blast or Mascot," Wishart says.

Not alone

Metabolomics isn't the be-all and end-all. Once the data is gathered and analyzed, with the metabolites identified, metabolomics often leads to new questions that can be followed up by using the other arms of systems biology. Because metabolomics may be more reflective of phenotype, as Siuzdak says, using it in combination with "the genetic information that we have, it gives us a really interesting story."

Andrew Lane agrees. "You can't do just one of the 'omics on its own," he says. "Once you've found something out from a metabolic pathway, you need to go back and verify that, OK, we're positing that this metabolic pathway has increased activity, that implies that there's either increased gene expression for those enzymes in that pathway or that some of the enzymes in that pathway have become more active by post-translational modification or by allosteric regulation. You have to look at gene expression, protein level, and protein post-translational modifications."

But that integration across the field is a challenge, not only for metabolomics, but for systems biology as a whole. "There's absolutely no point in just concentrating on one 'omics. We have to be able to integrate data across all the 'omiceses," Lindon says. "Making

sense of data that we collect at the different levels of the 'omics — genomics, transcriptomics, proteomics, and metabonomics — understanding all of that in the context of systems biology is very, very important. It's where we're going."

Tecan's new PowerScanner™

From high resolution to the art of perfection!



Can you afford to lose on content?

With Tecan's new PowerScanner you're ready for analyzing ultra high density microarrays. Integrated ozone filtering, auto-processing and internal calibration provide perfect quality conditions. Our premium microarray scanner gives you 2 µm pixel resolution at benchmarking sensitivity and accuracy. Compatibility of data formats and automated batch analysis is provided.

This system is available soon.



- 2 – 40 µm resolution
- < 0.04 fl/µm² detection limit
- < 2.5% CV uniformity
- 10⁵ dynamic range



More Reasons to Talk to Tecan



www.tecan.com/PowerScanner USA +1 919 361 5200 ext. 549 Europe +49 79 51 94 170 Asia +81 44 556 7311

Liquid Handling & Robotics | Detection | Sample Management | Components | Service & Consumables

© 2008 Tecan Trading AG, Switzerland, all rights reserved.