Big data, big picture: Metabolomics meets systems biology

Metabolomics—the study of the collection of an organism’s metabolites—provides a molecular measurement of phenotype, or the characteristics resulting from the genotype’s interaction with the environment. Using a range of analytical tools to scale the mountains of data collected, including molecular detection and bioinformatics, scientists use metabolomics to understand systems biology, which is the complete computational analysis and modeling of an organism and its well-being. By Mike May

A forgotten fourth category of molecular biology might be the one that teaches us the most about phenotypes. We’ve focused for decades on a molecular trio: DNA makes RNA makes proteins. And now, many scientists are mindful of an additional group: metabolites, which are made by proteins at work in an organism’s biochemical pathways. We hear so much about genes and genotypes, but—important as they are—we want to explore what happens when these genes and environmental elements interact. We want to find the molecular phenotypes that distinguish health from disease, and that’s what metabolomics provides.

Undoubtedly, systems biologists—those looking at interacting elements of life science—want to correlate the metabolome with the genome. “Metabolomics can’t really function on its own without the genome sequence,” says Jonas Korlach, chief scientific officer at Pacific Biosciences, in Menlo Park, California, a company that specializes in genomics analyses. “Once you find metabolites and want to analyze them for new antibiotics or new pathways, you need the genome information to identify the enzymes and clone the genes.”

Robert Trengove, director of the Separation Science and Metabolomics Laboratory at Murdoch University in Western Australia, agrees. Says Trengrove, the impact of metabolomics on systems biology is still in its infancy, but he sees positive advances underway, and the key is teamwork. “We’ve got people who are very good at doing informatics, processing ‘omics,” he says. “But few people have a solid understanding across all the ‘omics, including lipidomics and epigenetics.”

Trengrove’s group already teams up with others, collecting blood samples from an intensive care unit to create metabolic profiles of patients. “This way,” he says, “we can start to evaluate the use of various compounds as biomarkers and indicators of patient recovery.” To really know what’s going on, the biomarker panels must be correlated with proteomics and genomics, but that work is just getting started.

Nonetheless, systems biology has already turned biology on its head. Traditionally, biology has broken down an organism or cell into its subparts. Now, systems biology appears to embrace Aristotle’s view that “the whole is greater than the sum of its parts,” and is borrowing from various fields, including metabolomics. Tools to handle collections of data from these fields—collectively known as “big data”—are now emerging.

Thinking twice about data

Overall, metabolomics creates a numerical challenge. “Metabolomics is now often used to accompany large genomic cohort studies from biobanks, to correlate genotype and genomic variants with specific phenotypes, to complement nutritional studies monitoring food components or endogenous metabolites, or to support measurements in epidemiology studies,” says Andreas Huhmer, director of proteomics and metabolomics marketing at Thermo Fisher Scientific, headquartered in Waltham, Massachusetts. And that creates a lot of data. More than 7,500 metabolites have been detected in humans, and only about 2,500 of them come directly from the person; the others come from other sources, including foods and drugs that the person ingested.
Numerically, we’re probably just at the starting point. As Huhmer says, “More metabolites are expected to be identified in the future, particularly with increased efforts in understanding metabolites and microbiome-associated metabolism in the gut.” Bowel microflora play a significant role in metabolic profiles, providing a treasure trove of information on the impact of lifestyle and diet on chronic and acute diseases such as type 2 diabetes and obesity.

Two data repositories for metabolomics—the Metabolomics Workbench and MetaboLights—promise to give scientists more data to analyze as teams, enabling them to share data across the globe.

Leveraging lipidomics

Lipidomics is another integral piece of the metabolomics puzzle, but it has been a neglected area of ‘omics, because few researchers explore these molecules—at least compared to their investigation of DNA, RNA, and proteins. However, lipids provide a useful tool for systems biology, because “they can be quantified, which tells us about the state of a cell or tissue,” explains Kai Simons, CEO of Lipotype in Dresden, Germany.

The company’s Lipotype Shotgun Lipidomics Technology provides that quantification. This technique performs mass spectrometry on a whole extract, rather than separating it first by liquid chromatography. With only one microliter of blood, this technique provides what Simons calls “absolute quantification—we can identify up to 2,300 lipids.”

Scientists can send a sample to Lipotype and have a total lipid analysis completed in just two weeks. If desired, the company’s software is available to researchers who want to look through the data. Still, biologists need to collect even more data on various metabolites and determine what they do. In fact, learning the function of more metabolites could make the biggest impact on understanding complete biological systems.

Figuring out function

Even if we could detect all the proteins and metabolites in any biological matrix, says Jose Castro-Perez, director of health sciences marketing at Waters in Milford, Massachusetts, “we’d know what only a small percentage of them do.” Scientists need ways to both detect metabolites and understand their biological functions. That requires analytical and bioinformatics tools to conduct disease or therapeutic association and pathway analysis that combines various forms of ‘omics data.

For that, Waters developed its SONAR software, a data acquisition mode that works with the company’s Xevo G2-XS QToF, which provides quadrupole time-of-flight (QToF) mass spectrometry. SONAR can catalog a complete sample with precursor and fragment ion spectra from a data independent analysis (DIA) experiment in a single sample injection, giving researchers quantitative and qualitative information about the proteins or metabolites. “This new DIA acquisition mode is more advanced than other DIA approaches, because it provides faster and more selective data acquisition for complex samples,” Castro-Perez says. “Furthermore, this new approach allows for improved reliability of database library searches and quantitation accuracy.” The Xevo G2-XS QToF can be integrated with Waters’ chromatographic tools, such as ultra-performance liquid chromatography, for high throughput.

“Generating high-quality data is important, but ultimately you need to be able to go from raw data to meaningful and actionable biological information,” says Castro-Perez. To synthesize the information and simplify the data-handling and processing workflow, Waters developed its Symphony software, a client/server application that allows the automation of one, or several, data-handling or processing functions in a sequence. This tool can even initiate data processing immediately following an instrument run and complete it without user intervention—features that are very important in large-scale studies.

Combining complementary methods

Despite all of the advances in storing and analyzing data, scientists are still confronting significant obstacles in studying the metabolome. “One bottleneck in nontargeted workflows is the identification of unknown compounds,” says Aiko Barsch, market manager for metabolomics at Bruker Daltonics, based in Bremen, Germany. “This is where MS and NMR [mass spectrometry and nuclear magnetic resonance] both have advantages.”

For example, high-resolution, accurate-mass (HRAM) MS can reveal the elemental composition of unknown compounds. “MS has come a long way, and systems that provide ‘extreme resolution’ enable researchers to read out elemental compositions from the so-called ‘isotopic fine structure,’” says Barsch, “but if a real unknown—something not in a database—appears in a sample, then you need de novo structural elucidation capabilities, and that’s a key job of NMR.” So, MS and NMR can be used together, as complementary techniques in metabolomics.

To dig even deeper into complex samples, researchers often combine MS with liquid or gas chromatography (LC/HRAM-MS or GC/HRAM-MS, respectively). “This helps to pinpoint characteristic metabolites,” Barsch explains, “because the separation combined with high-resolution detection zeros in on specific components of a sample.”

Advances in NMR also help. Today’s platforms include standard operating procedures that allow scientists to transport a...
As scientists and manufacturers develop more tools to analyze metabolomics in more detail, we learn more about biological systems—and how they function in healthy and disease states, as well as how they change and by how much.

Analyzing those changes requires specialized software. In 2016, Agilent released VistaFlux, which performs stable-label isotope tracking. Since various biological routes can produce a metabolite, “the only way to know what route produced a given metabolite is by tracking the consumption of its tracer through intermediate metabolites to its ultimate fate,” Fischer explains. “VistaFlux shortens data analysis from weeks of manual data processing to hours using this software, while increasing the number of metabolites that can be examined.”

For a completely dedicated metabolomics solution, scientists can combine Agilent’s 1290 UHPLC and 6470 Triple Quadrupole LC/MS system with the Agilent Metabolomics dMRM Database and Method—this platform can measure 217 metabolites.

Other companies provide scientists with additional options in analytics. For example, Thermo Fisher Scientific’s HRAM Orbitrap-based MS systems can detect as many as 1,000 metabolites from several microliters of human plasma in less than an hour. “For high-throughput, untargeted analyses,” says Huhmer, “the Thermo Scientific Q Exactive mass spectrometers, combined with the breadth of chromatographic separations—including LC, [ion chromatography], and GC systems—detect and resolve chemically diverse compounds in the metabolome.” This technology can be combined with Thermo Scientific Compound Discoverer software, which can confidently turn data into meaningful results, says Huhmer.

Protocol from one platform to another. The growing level of teamwork in metabolomics makes these features crucial to scientists collaborating internationally, because they seek consistency in NMR data coming out of different labs.

Although metabolomics researchers use both MS and NMR, Steve Fischer, marketing director for metabolomics and proteomics at Agilent Technologies in Santa Clara, California, says, “The trend strongly favors mass-spectrometry solutions, because of instrument cost, breadth of measurement, and sensitivity.”

Given the wide range of metabolites that can be detected and the overlap in their masses, combining chromatography with MS provides separation and deeper analysis of a sample. “Broadly speaking, an LC/MS system can measure more things than a GC/MS system,” says Fischer. Some samples cannot be volatilized easily, which is required for GC. Yet, “both systems provide mass-spectral information that can be used to track and eventually identify what metabolite has been detected and the abundance of that metabolite,” he says. By comparing samples, this information reveals which metabolites are changing and by how much.

Analyzing those changes requires specialized software. In 2016, Agilent released VistaFlux, which performs stable-label isotope tracking. Since various biological routes can produce a metabolite, “the only way to know what route produced a given metabolite is by tracking the consumption of its tracer through intermediate metabolites to its ultimate fate,” Fischer explains. “VistaFlux shortens data analysis from weeks of manual data processing to hours using this software, while increasing the number of metabolites that can be examined.”

For a completely dedicated metabolomics solution, scientists can combine Agilent’s 1290 UHPLC and 6470 Triple Quadrupole LC/MS system with the Agilent Metabolomics dMRM Database and Method—this platform can measure 217 metabolites.

Other companies provide scientists with additional options in analytics. For example, Thermo Fisher Scientific’s HRAM Orbitrap-based MS systems can detect as many as 1,000 metabolites from several microliters of human plasma in less than an hour. “For high-throughput, untargeted analyses,” says Huhmer, “the Thermo Scientific Q Exactive mass spectrometers, combined with the breadth of chromatographic separations—including LC, [ion chromatography], and GC systems—detect and resolve chemically diverse compounds in the metabolome.” This technology can be combined with Thermo Scientific Compound Discoverer software, which can confidently turn data into meaningful results, says Huhmer.

Sythesizing on a systems level
To understand the systems biology of metabolites—or how they work together—scientists must connect them to pathways, which is precisely the purpose of the cloud-based XCMS Online platform. “You can take data from an LC/MS run, and—in one click—pull out the pathways that it predicts are being dysregulated,” explains XCMS Online’s creator Gary Siuzdak, professor and director of the Scripps Center for Metabolomics at the Scripps Research Institute in La Jolla, California. “It can also integrate proteomic and genomic data in the analysis.” This capability provides several levels of validation.

Most important, XCMS Online makes it easy to explore the results. For example, it creates a Pathway Cloud Plot, an interactive graph of the metabolites grouped by pathways. Clicking on a pathway bubble provides the name of the pathway, the metabolites that are associated with it and those that are not, related statistics, and more.

Over 14,000 scientists already use XCMS Online, which allows them to run analyses and share the results, because it’s cloud-based technology. “We’ve seen people from every continent—even Antarctica—using this,” Siuzdak says.

As analytical devices continue to improve and bioinformatics systems get more powerful, yet easier to use, systems biology approaches will stretch across more areas. “There is a strong movement in systems biology to become more translational,” says Fabian Theis, director of the Institute of Computational Biology at the Helmholtz Zentrum München, in Germany, “and medical research is generating lots of ‘omics measurements.”

Metabolites fluctuate over the course of a day, and blood makes a good sample for tracking that variation. Looking at blood samples from large patient cohorts proves especially interesting. “From this,” says Theis, “we can combine various ‘omics, build networks, and then correlate them with population cohorts or clinical trials.”

The Helmholtz Zentrum München specializes in such big cohorts. They measured metabolites in a few thousand patients with MS, then matched the concentrations of metabolites with single nucleotide polymorphisms (SNPs). “You usually find dozens or hundreds of metabolites associated with SNPs, and if you correlate a SNP with the ratio of two metabolites—those next to each other in a biochemical pathway, or reaction—you can see a SNP’s effect on this reaction,” Theis explains. “For example, we can choose healthy disease groups or simpler phenotypes, such as males versus females, and ask if we find a metabolic footprint of this variable.”

The data do not provide easily interpretable results, however. Many of the associations come from indirect effects, because one reaction in a pathway can drive long-reaching impacts on other reactions. Using computational and statistical tools, Theis and his colleagues link two metabolites after correcting for the influence of all others. “Then, the correlations expose clear pathways,” Theis says, “and you can then compare across diseases.”

As scientists and manufacturers develop more tools to analyze metabolomics in more detail, we learn more about biological systems—how they function in healthy and disease states, as well as how they change over time and in different environments. The key to learning even more revolves around collecting larger datasets—and sharing them for analysis with scientists around the world.

Mike May is a freelance writer and editor based in Florida.

DOI: 10.1126/science.opms.p1700115