SRF/CESB Metabolomics Services

August 2015

INTRODUCTION

The Center for Environmental and Systems Biochemistry (CESB) has been operating since Sept 2013, and has established collaborations with 6 groups from within the CCSG, some already leading to new NIH funding and several proposals are pending. The CESB also houses one of the six metabolomics centers supported by the NIH Commons Funds, called the Resource Center for Stable Isotope Resolved Metabolomics (RC-SIRM). The CESB/RC-SIRM in partnership with MCC, has provided 5 pilot and feasibility grants to CCSB members.

Workshop and Symposium

Under the aegis of the RC-SIRM, we host an annual Symposium and hands-on, wet-lab/dry-lab Workshop on metabolomics. In 2015 the Symposium was held at the Biopharm Building on July 26, immediately followed by the Workshop from July 27-August 7. The Symposium was open to all, but registration is required, whereas the Workshop has strict limits on numbers who can be accommodated. More details can be found at: http://symposium2015.cesb.uky.edu

A. What is Metabolomics?

Before this can be answered, we first have to define "metabolism". The working description used at CESB (http://bioinformatics.cesb.uky.edu/bin/view/RCSIRM/) is "Living cells are maintained under non-equilibrium conditions, which requires constant input of energy. The cells must also maintain their infrastructure, and perform tissue-specific tasks, all of which need energy and raw material. Metabolism is the set of processes that convert exogenous compounds to metabolic energy, which drives biochemical reactions within the cell, maintains homeostasis, provides the means to do work (e.g. contraction, movement, action potentials, secretion and so forth), for cellular repair, and to divide. Metabolism responds to exogenous signals as represented by diet and pollutants for example, and local environments (microenvironment) as represented by the conditions prevailing outside cells in tissues." *As such, metabolism is intrinsically a sensitive indicator of disease.*

Therefore, the ability to measure aspects of "metabolism" in detail is of fundamental importance in all aspects of biology. "Metabolomics" is defined as the technical means to carry out such analyses of metabolism, by identifying and quantifying a large fraction of all of the metabolites present in a cell, and how they change in response to perturbations of relevant metabolic networks. "Metabolomics" is NOT synonymous with "metabolism", because it is completely possible to carry out metabolomics analyses without gaining any insight, discovery, or understanding of metabolism. Terms such as "metabolome coverage" have been used to convey the central analytical need of metabolomics. Metabolomics therefore requires high (or appropriate) metabolome coverage, and this in turn requires very high-end analytical instrumentation, of which mass spectrometry and NMR together are the most appropriate technologies worldwide. Informatics is the third critical component of metabolomics, as high metabolome coverage is necessarily a "Big Data" activity in order to interpret the observations in a biological context. Careful sample preparation and processing is the often overlooked first critical component in metabolomics, which is necessary to prevent excessive perturbation and degradation of the observed metabolites. And before any of the above, appropriate experimental design and adherence to this design is the zeroth critical component of metabolomics.

From a cancer perspective, metabolism is often quantitatively reprogrammed compared with the untransformed state, and this may be reflected in changes in metabolite concentrations. Furthermore, gain of function mutations such as IDH1,2 in gliomas and AML result in the formation of novel metabolites (2-hydroxyglutarate in this instance) whereas loss of function mutations as found in for example hereditary renal carcinomas result in the build up of specific metabolites (e.g. fumarate or

succinate), with downstream effects on many metabolic and epigenetic events. However, more often than not, it is not the changes in the standing stock metabolite concentrations, but rather their synthesis and utilization that is metabolically reprogrammed in cancer. In order to detect this "metabolic reprogramming", a tracer such as ^{13}C and/or ^{15}N must be utilized in the experiment.

In summary, a fundamental cornerstone for this Metabolomics Component is stable-isotope tracing during the biological experiment. This means that the experimental design must be integrated with the analysis and informatics. In turn, this means that a simple fee-for-service model is unlikely to yield very useful results for cancer researchers, and this Component is structured for extensive advising and collaboration.

B. Step-by-Step Process and Key Questions

Investigators who seek information about metabolomics and the services offered through CESB should first visit our web site: http://rcsirm.cesb.uky.edu

This provides background information and a list of manuscripts using metabolomics to understand cancer function. Researchers have found that these papers can initially serve as a sort of "prix-fixe menu" (or "combo meals" if you want to be cynical) to get oriented with regards to experimental design, sample requirements, and expected types of information.

Other resources include:

Our Handbook of Metabolomics ("*The Handbook of Metabolomics* Methods in Pharmacology and Toxicology,
vol. 17" Humana Press 2012. DOI 10.1007/978-1-61779-618-0_4)

- U24 web site http://metabolomicsworkbench.org/nihmetabolomics/rcmrcs.php
- Metabolomics Society http://metabolomicssociety.org/

Once a CCSB member decides s/he needs metabolomics to understand aspects of their particular problem, they should first contact one of the 4 directors of CESB listed at the end of this newsletter.

The researcher should be prepared to discuss the following:

- What the problem is (e.g. I need to determine the change in molecule X when I do Y; I want to know whether the oxidative branch of the pentose phosphate pathway is contributing to the production of NADPH)
- What the experimental system is & why (e.g. cell culture, spheroids, animal models, human subject, PDX, etc.)
- Targeted or global metabolomics? Pathway analysis or flux analysis? For these questions, often the CCSB member will not have an answer, but should expect that these aspects will be discussed. It is highly recommended that the client be sure of the term flux, and refresh themselves on network dynamics and reaction diffusion systems under inhomogeneous conditions.
- What information is already available? For example, transcriptomics data that imply metabolic consequences, and preferably partially validated by protein analysis (cf. Westerns, MS, enzyme activity). Often the starting point for metabolic analysis is the Seahorse technology. The Seahorse XRF provides information about the rate of proton extrusion (ECAR) and the rate of oxygen consumption by adherent cells under normoxic conditions. By varying the extracellular nutrient source (e.g. glucose, glutamine, fatty acids) and with the use of specific inhibitors, it is possible to estimate net lactic fermentative flux, glycolytic reserve, respiration potential and the degree of coupling to oxidative phosphorylation. Such information is extremely valuable in designing metabolomics analyses or whether metabolism is a significant

correlate of the biological function of interest. If such data are not available, then it is generally recommended that they be acquired.

- Any Seahorse analyses? Are they appropriate to conduct first? This technique cannot presently be applied to whole organisms or organs.
- The consultation will then focus on the experimental design, which will also need to include biostatisticians for estimating sample sizes.
- It is also important to establish whether the user will do their own sample work up, what analytical techniques are needed, and what level of data reduction is required.
- The SRF has SOPs. Questions that cannot be answered using these SOPs will be considered in terms of development time to establish specific metabolic assays. This makes sense only in the context of a small number of metabolites (targeted metabolomics).
- Costs
- It is VERY common for data analysis to take 90% of the metabolomics labor, so the researcher should be prepared to discuss the time frame and cost structure with this consideration, versus obtaining data analysis training for their own laboratory. We have found, however, that the latter approach still incurs considerable labor for the CESB staff, as data analysis is highly nuanced and extensive consulting is still typically needed.

Once an experimental approach has been worked out, the user will be introduced to laboratory personnel who will carry out the analyses and oversee the project progress- this person will then be the primary contact.

The SRF is operating on a charge-back basis, which means that data quality is contingent mainly on experiment and sample quality. The center can perform feasibility or pilot analyses to assess this quality.

Data collection is not the rate limiting step. Global analyses often produce hundreds to thousands of identifiable and quantifiable compounds or "features", but this means the reduction of raw data is extremely time consuming. The researcher will work out with the SRF, the list of identified compounds and their amounts, with an estimate of reliability.

C. Examples/Case Studies

As stated above, our own publications are increasingly serving as a *de facto* "prix-fixe menu" of experimental design and expected types of information. Examples of the use of stable isotope resolved metabolomics can be found in the publications listed below and on our website:

Fan, T.WM., Lane, A.N., Higashi, R.M., Farag, M.A., Gao, H., Bousamra, M. & Miller, D.M. (2009) Altered Regulation of Metabolic Pathways in Human Lung Cancer Discerned by 13C Stable Isotope-Resolved Metabolomics (SIRM). *Molecular Cancer*. **8**:41

Fan, T. W-M., Lane, A.N., Higashi, R.M., Yan, J. (2011) Stable Isotope Resolved Metabolomics of Lung Cancer in a SCID Mouse Model. *Metabolomics* **7**, 257-269

Le, A., Lane, A.N., Hamaker, M., Bose, S., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., Zimmerman. L.J., Liebler, D.C., Slebos, R.J.C., Lorkiewicz, P.K., Higashi, R.M., Fan, T.W-M., and Dang, C.V. (2012) Myc induction of hypoxic glutamine metabolism and a glucose-independent TCA cycle in human B lymphocytes. *Cell Metabolism.* **15**, 110-121

Yang, Y., Lane, A.N., Fan, T.W-M., Rickets, C., Wu, M., Boros, L., Linehan, W.M. (2013) Understanding How Fumarate Hydratase (FH) Null Cells Use its Central Carbon for Energy and Malignant Development *PlosOne* **8**, e72179

Sellers, K., Fox, M.P., Bousamra, M., Slone, S., Higashi, R.M.,Miller, D.M., Wang, Y., Yan, J., Yuneva, M., Deshpande, R., Lane, A.N., Fan, T. W-M. (2015) Pyruvate carboxylase is upregulated in NSCLC. *J Clin Invest*. **125**(2): 687-698

The review by Fan, T.W-M., Lorkiewicz, P., Sellers, K., Moseley, H.N.B., Higashi, R.M., Lane, A.N. (2012). Stable isotope-resolved metabolomics and applications to drug development. *Pharmacology & Therapeutics*. **133**:366-391 is also a good background source of examples.

Untargeted metabolic profiling of biofluids is widely used, especially for biomarker discovery and validation. Technically this is quite straightforward, and its value is determined by the effect size, sample quality and robustness of the multivariate statistical analyses. Drs. H. Moseley, C. Wang, H. Weiss should be consulted for considerations of sample sizes and approaches. Metabolic profiling of retrospectively collected biofluids is not recommended for all but feasibility projects as it is rare that sufficient meta data are available, and collection and storage methods are usually suboptimal.

Follows an example study and pipeline.

What is the role of tumorin in tumor development and survival in TNBC?

The goal is to understand how central metabolism is impacted by the knockdown, including sources of NADPH needed for proliferation. The user will already have information from Seahorse analysis, and phenotypic effects of the knockdown. Cell cycle distribution analysis is also important for the overall interpretation.

(i) 3 cell lines ±shRNA against tumorin- triplicate experiments, 3 tracers ([U-^{13}C]-Glc, [$^{13}C_{1,2}$]-Glc, [U-^{13}C,^{15}N]-Gln = **54** experiments.

Polar + non polar metabolites = **108** analytical samples.

Protein may be used for additional experiments (e.g. expression), and/or normalization.

Polar metabolites analyzed by NMR and GC-MS or IC/FT-MS= **216** experiments,

+ **54** FT-MS of nonpolar fraction

total = **270** analyses.

Media samples at 5 time points for each dish =270 media samples, analyzed by NMR and MS

total= **540** analyses.

Total analyses = 810

(ii) Same cell lines as orthotopic xenograft in NSG mice (5 mice/group), two tracers = 60 mice.

Tumor tissue + nontumor tissue + 2 blood sample per mouse = 240 samples.

120 blood analyses by MS and NMR = **240** analytical samples.

Tissue polar and non polar = **480** analyses.

Total mouse sample analyses = 720

Grand total = 1530 analyses.

The number of analytes including isotopomers and isotopologues is > 200,000 quantified analytes for this one study.

At this point, the data are reduced to lists of compounds, their isotopomers and their amounts for biological interpretation.

Outline Procedure

Design experiment->execute biological experiment with tracers-> harvest sample->prepare sample for analysis->analytical data acquisition->data reduction->information retrieval and interpretation. Unfortunately, the pipeline is essentially linear, it isn't ethical to interpret data before they have been produced.

Detailed SOPs are available on request or via our web site.

Grow cells in culture in triplicate with each tracer ± shRNA. 18 experiments (x3 for each cell line).

Sample media at defined time points and store (e.g. 0, 3,6,9,24 h-the zero time point is critical) (90 samples)

Harvest and extract cells

Store metabolite fractions (polar, non-polar, protein)

Dried samples must be reconstituted in appropriate volume of buffers for different analytical platforms, and loaded into the correct labeled sample tubes.

Prepare for GC MS

Prepare for NMR – run first for quality control on sample and extraction integrity

Prepare for FT-MS

Record spectra on the different platforms

Reduce data to raw isotopologues distributions for each tracer

Repeat any bad experiments

Repeat for next cell line

Tumor bearing mice: treat with tracer, sample blood. Harvest tissue

Extract tissues and blood

Prepare for analytical spectroscopy

Record spectra on the different platforms

Reduce data to raw isotopologues distributions for each tracer

Repeat any bad experiments

Data acquisition and reduction.

GC-MS takes 1 h per sample to run. QC/standard samples must be run in interleaved mode.

NMR spectra take 1.2 h per cell or tissue sample to run, 0.5 h for plasma extracts

FT-MS for lipids takes 10-15 min per sample.

Data reduction for this density of data is 1-1.5 h/spectrum.

The results can be interpreted in terms of specific networks related to cell growth or survival, with limited flux information (exclusively in this design for inputs and outputs).

Data Processing and Analysis

The "data" comprise several components, as follows.

1. Metadata that describe in exact detail the entire workflow from sample receipt to final products. No useful results can be obtained without these data. An Excel template is available for these data.
2. Raw analytical data, i.e. the streams of bits coming from the instruments. For FT-MS and NMR these represent digitized electrical signals in the form of free induction decays comprising both real and imaginary parts. For other MS data, these are digital representations of ion counts.
3. Raw analytical data have to be transformed into a usable form, which for FT-MS and NMR is the discrete fourier transform and associated digital processing to suppress truncation artefacts, optimize signal to noise ratios etc. The resulting output is a spectrum of intensity versus frequency. For NMR, the frequency is usually transformed to chemical shift, in ppm, that is independent of magnetic field strength. For FT-MS, the frequencies are mapped onto an m/z range.
4. Intensities (ordinate values) must be internally normalized to obtain amounts of materials (i.e. numbers of moles of substances or of ions), and back calculated to the values associated with the original spectrum, on an agreed upon measure of the amount of that species (such as biomass weight, protein mass etc.). This absolutely requires accurate metadata. Actually, amounts are related to peak areas (or volumes) NOT peak heights; appropriate numerical or analytical integration procedures must be correctly applied, taking due account of baseline drift, phasing errors and peak overlap.
5. For isotopomer and isotopologue analyses, the intensities are usually expressed as mole fractions ("enrichments"). As these are ratios, normalized to cell amount is not needed. For MS, the natural abundance needs to be corrected.
6. Spectral features need to be mapped onto identifiable molecules ("assignment"), using the available spectra information, and by reference to our databases.
7. For "profiling" typically one is concerned with case-control comparisons, which require large numbers of specimens (each unique). Multivariate statistics are generally appropriate for initial analyses- are the groups different?; What is different about these groups? PCA and OPLSDA (SimcaP) may be used.
8. Normalization. To compare case and control, the quantity of each metabolite must be normalized to the appropriate amount of specimen. Cell number is generally not appropriate as cell volumes vary widely among types, and also in response to treatment. Total biomass or a surrogate is appropriate (e.g. dry weight, total protein).
9. Total DNA may not be appropriate in a case-control study because the amount f DNA per cell varies twofold during the cell cycle, and the control and treated samples do not necessary have the same cell cycle distribution. Comparison of different cell types is then further compromised where there are different numbers of chromosomes present (diploid G1 normal cell, triploid cancer cell, tetraploid cardiomyocyte arrested at G2/M).

10. With SIRM studies, a question is often what pathways were impacted, which requires pathway tracing (SIRM) and biochemical expertise.

Quantitative analyses may also be carried out (e.g. what is the rate of nutrient utilization and waste product excretion). Kinetic models (flux analysis) based on enzymology can also be applied where needed. These studies need careful consideration of the time dependence of the biomass as a function of time for accurate normalization of rates. Flux: the number of moles of nutrients (e.g. glucose, glutamine) consumed and the number of moles of product excreted (e.g. lactate, alanine, glutamate) is measure as a function of time, producing a time course of consumption and excretion. To determine rates, it is essential to normalize to the functional unit of metabolism which is the amount of enzymes present in the system. This is proportional to the concencentration of the enzymes times the relevant intracellular volume (unknown).

11. With tracers, the time course of the isotopomer distributions can be determined, as can the fraction of glucose (glutamine) consumed that is converted to excreted product (e.g. lactate, alanine, glutamate).

12. Further statistical analyses if needed can be carried out in the Biostats unit.

1-11 is usually carried out in the Analytical Core, where the expertise resides. #10,11 can be rather open ended, and is usually done as a collaboration rather than as a fee for service (which requires that the result be known in advance).

D. Link to RM SRF website

http://bioinformatics.cesb.uky.edu/bin/view/RCSIRM/ServicesAndRates

E. Contacts:

Rick Higashi (rick.higashi@uky.edu) Associate Director, Metabolomics Component of RM-SRF, Director of RC-SIRM
Teresa W-M. Fan (twmfan@gmail.com), CESB overall Director
Hunter N.B. Moseley (hunter.moseley@uky.edu), CESB Informatics Director
Andrew N. Lane (andrew.lane@uky.edu), CESB NMR Director
Alicia J. Colliver (ajduna0@uky.edu), CESB Administrative Operations Manager
Allan Butterfield, Ph.D., Director, Redox Metabolism SRF Phone for RM SRF Laboratory: 323-1106; Email: dabcns@uky.edu