Metabolic network models of disease for integration and statistical analysis of omics data

James Eddy, Nathan Price | Contact: james.eddy@systemsbiology.org
Institute for Systems Biology, Seattle WA

Institute for Systems Biology
Revolutionizing Science. Enhancing Life.

Enabling biomedical insights with metabolic models

A different motivation for model & tool development

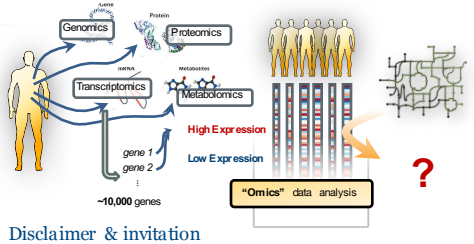
Hypothetical scenario: a non-COBRA researcher, bench or computational, is hoping to learn something about his or her dataset; how can models help?

Fostering independence: how can we allow researchers to use models and tools to generate hypotheses without relying on COBRA experts?

How can we develop useful pipelines to support biomedical researchers in the community?

- Accurate flux prediction is an ongoing challenge, even in microbes.
- Groups are continuing to make advances on this front with new tools and manual reconstruction efforts
- Lots of data available, should provide pipelines for data integration and analysis, without knowing the ins and outs of FBA et al.

What's needed?
Synergism within the COBRA field, *demonstrated utility* of analysis methods (especially those geared towards statistical learning), and *open-source, user-friendly* software and resources.



Disclaimer & invitation

The topics presented here are based on my own experiences and interactions with others; while I have some thoughts on future strategies, I certainly don't have all the answers worked out. **My goal is to encourage brainstorming and discussion.** To that end, I welcome any comments and suggestions via email. Depending on the volume of feedback, I would be happy to collate and organize ideas and share with the public.

Building better (metabolic) models of normal and disease systems

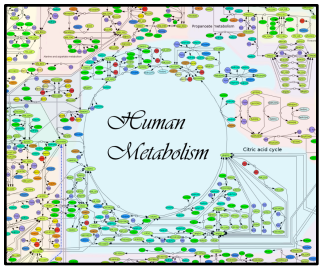
Keeping up with the latest and greatest human reconstruction

Recon 2 and beyond: community effort produced a more expansive and better curated picture of human metabolism, but was not free of lingering issues. Efforts are ongoing to update and improve the model:

- **Recon 2.03** (from I. Thiele) – corrected/expanded annotation (identifiers)
- **Recon 2.1** (from K. Smallbone) – elemental mass balancing
- Differences not always straightforward to reconcile

This presents a fair share of complications to the various owners and maintainers of Recon X – it's even more confusing to "the rest of us."

How can we avoid the "need" for more Jamborees?
Rather than wait for a massive event to unify the growing number of model branches, we can work to establish more centralized, open, and interactive environments for Recon X development. Moving away from monolithic reaction databases will ease coordination between groups and make changes more transparent to the community.



Developing a web-based system, not only for querying Recon information, but for editing or adding to the network will require careful consideration of *standards*, *version control*, *functionality*, and *visualization*. Finding funding and talent to expertise might be more challenging...

From global to tissue-specific metabolism (without reinventing the wheel)

How should we build tissue-specific models?
With the plethora of methods and tools available to generate tissue-specific models in humans, an *objective, comparative study* is needed. Do different methods have advantages/disadvantages, is there a global optimum? Current methods include GIMME (and newer versions), MBA, iMAT, mCADRE, INIT, FASTCORE, and more. Strategies to determine a consensus among method and model features could also be worth investigating.

How do we evaluate tissue-specific models?

- Comparing to **measured flux distributions** is theoretically ideal, but data is often unavailable for specific human systems (and presents its own complications, as measurements aren't *direct*)
- Same story goes for **gene essentiality** and **synthetic lethality**
- Other metrics like **coverage** of detected proteins and metabolites could be useful, but the negative case (exclusion of non-detected molecules) is harder to establish with confidence

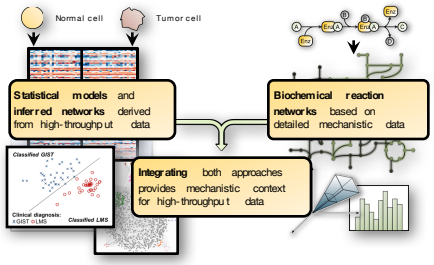
Acknowledgements & shout-outs

Cory Funk, Shuyi Ma, Andrew Magis, and of course **Nathan Price** from my group have been instrumental in helping with various aspects of data generation and processing to date, as well as general strategizing. **Neil Swainston, Kieran Smallbone,** and **Ines Thiele** (along with members of her group) have been patient and helpful in discussing and sorting out issues with Recon 2. Along with **Ben Heavner,** Neil and Kieran have also been great for thinking about ways to improve Recon 2 and the model-building process.

Markus Herrgard recently published a great comparison of methods for predicting flux from expression in microbes. **Nathan Lewis** and **Jonathan Monk** have also published nice review articles in the past year on modeling cancer metabolism and building better models in general, respectively. **Nikos Vlassis**' paper for FASTCORE is a mathematical tour de force, and doesn't seem to get enough attention. I'm sure there are plenty of other excellent people I'm missing who have either helped me directly or published papers well worth reading.

Exploring network-driven data analysis

From network refinement to biochemically informed inference



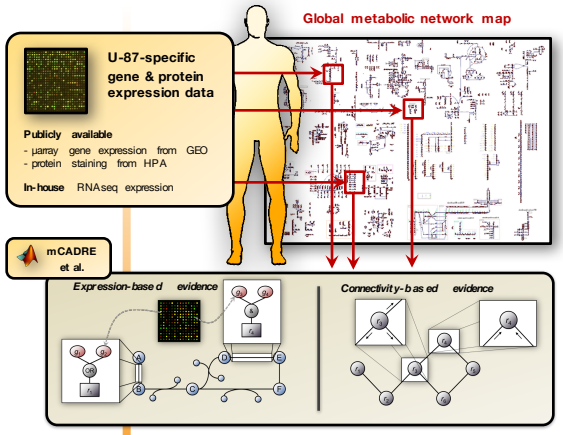
How much can a network tell us about the behavior of a biological system? Perhaps we don't need to limit evaluation to "predictive accuracy" of model simulations. In parallel to efforts aimed at building and improving metabolic network models based on experimental and omics data, we should also investigate how to *leverage network information to integrate and make sense of patterns in biological data.*

Statistical inference and machine learning: network structure and annotation can be an informative prior and enable easier functional interpretation of results – without relying as heavily on the uncertainty of constraints and stochastic nature of flux predictions.

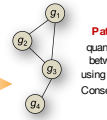
How can we get started with network-driven analysis?

- Differentially expressed metabolic pathways
- Mapping genomic variants to metabolic pathways to reduce noise
- Network structure provides attractive chassis for multi-omic integration

Evidence-driven reconstruction of a glioblastoma metabolic network



	Recon 2	U-87 model
Reactions	7,445	3,977
Metabolites	5,064	3,274
Transcripts	2,171	1,955
Unique Genes	1,766	1,583



Genomic mutations from TCGA mapped onto metabolic reactions to investigate pathway-level trends

Gene Name	Entrez ID	nMutated	%	Reaction Subsystems
PTEN	5728.1	65	24%	inositol phosphate metabolism
EGFR	1956.1	59	21%	Fatty Acid Metabolism
PIK3R1	5295.3	30	11%	inositol phosphate metabolism, Phosphatidylinositol phosphate metabolism
PIK3CA	5290.1	24	9%	inositol phosphate metabolism, Phosphatidylinositol phosphate metabolism
IDH1	3417.1	14	5%	TCA, urea acid cycle fatty acid metabolism

A case study in brain cancer metabolism

To investigate strategies for constructing and using tissue-specific metabolic networks with different types of omics data, we are focusing on the well-characterized U-87 MG cell line for glioblastoma multiforme. **The starting point is a slightly updated version of Recon 2.03:**

- Additional annotation from [Haraldsdottir '14]
- Definition of a U-87-specific biomass reaction
- Specification of a more constrained growth medium

Data available for U-87 include a fully sequenced genome, microarray and RNAseq gene expression, protein expression from HPA, secreted proteomics, targeted metabolomics, as well as metabolic and growth phenotype measurements.

Semi-automated model construction

We've incorporated **minor updates to mCADRE** to improve performance (thanks to FASTCC), handle multiple levels of evidence, and check for biomass production. However, we're also planning to **compare features and predictive strengths/weaknesses** of the mCADRE model with outputs of several other tissue-specific model building tools as well as the global Recon 2 model.

Model validation and evaluation

With data from collaborators, we can **compare qualitative predictions to measured effects** (increased/decreased growth, ATP production, oxygen consumption) of *inactivated reactions*, based on an experimental screen of enzyme inhibitors. This as well as coverage-based metrics provide a basis for cursory validation and comparison between different models.

Metabolic context for statistical analyses

Integrating genomics, transcriptomics, and other data for U-87 as well as GBM samples (e.g., from TCGA) allows us to identify *frequently mutated* subsystems or *variably expressed* pathways. The latter feature can present hypotheses for subnetworks perturbed in the cancer.