

Detection of novel metabolites and enzyme functions through *in silico* expansion of metabolic models

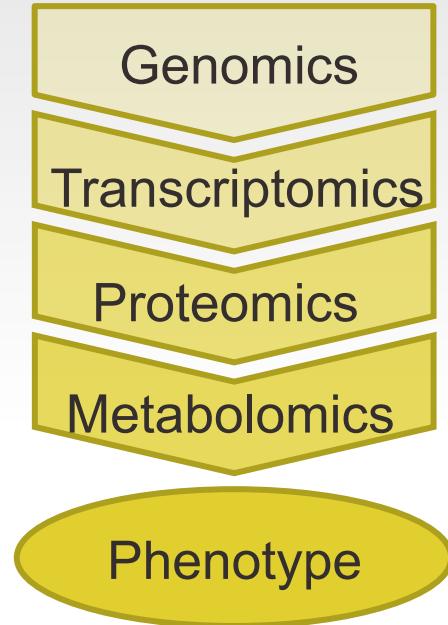
James Jeffryes - March 14, 2016
Advised by Chris Henry & Keith Tyo



Northwestern | McCORMICK SCHOOL OF
ENGINEERING

Measuring the metabolome is of medical and mechanistic interest

- Metabolomics seeks to measure all intracellular molecules < 1500 Da
- Directly measures many phenotypes on a rapid time scale^{1,2}
- Critical for gene and protein annotation and modeling^{3,4}



1. G. Patti et al. 2012

2. A. Scalbert et al. 2009

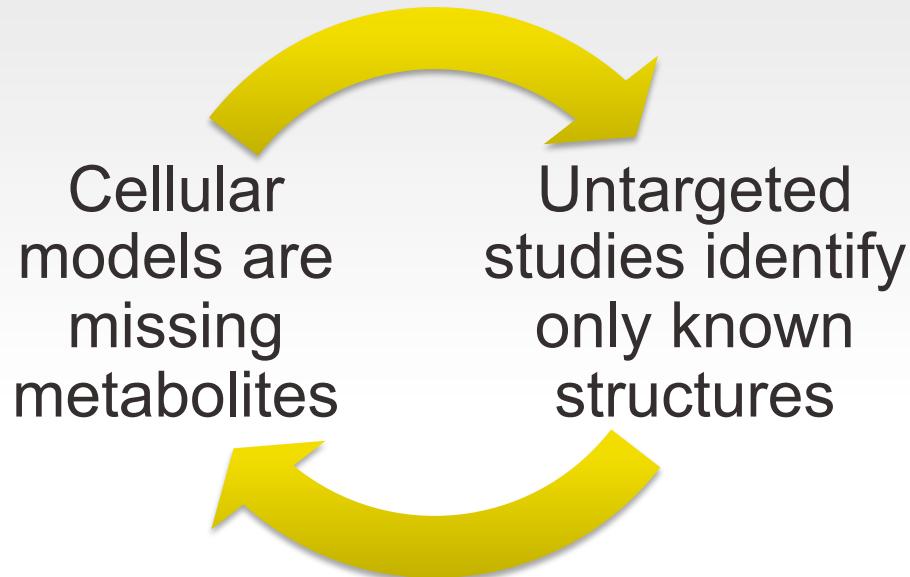
Northwestern | ENGINEERING

3. Dromms & Styczynski 2012

4. Prosser et al. 2014

We are currently unable to measure a majority of the metabolome

- There is no universal method for measuring small molecules¹
- Generally only possible to confidently identify 5-20% of features in LC MS/MS dataset²



1. O. Fiehn et al. 2011

2. K. Scheubert et al. 2013

Complete Enumeration



Model expansion



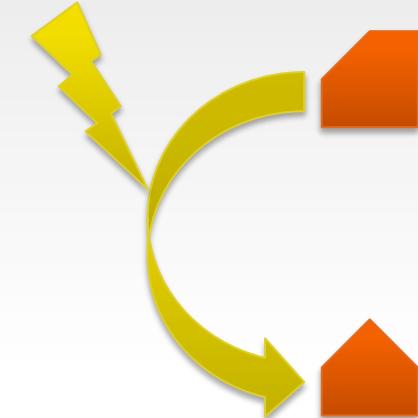
There are several sources of unknown metabolites



Enzyme Promiscuity

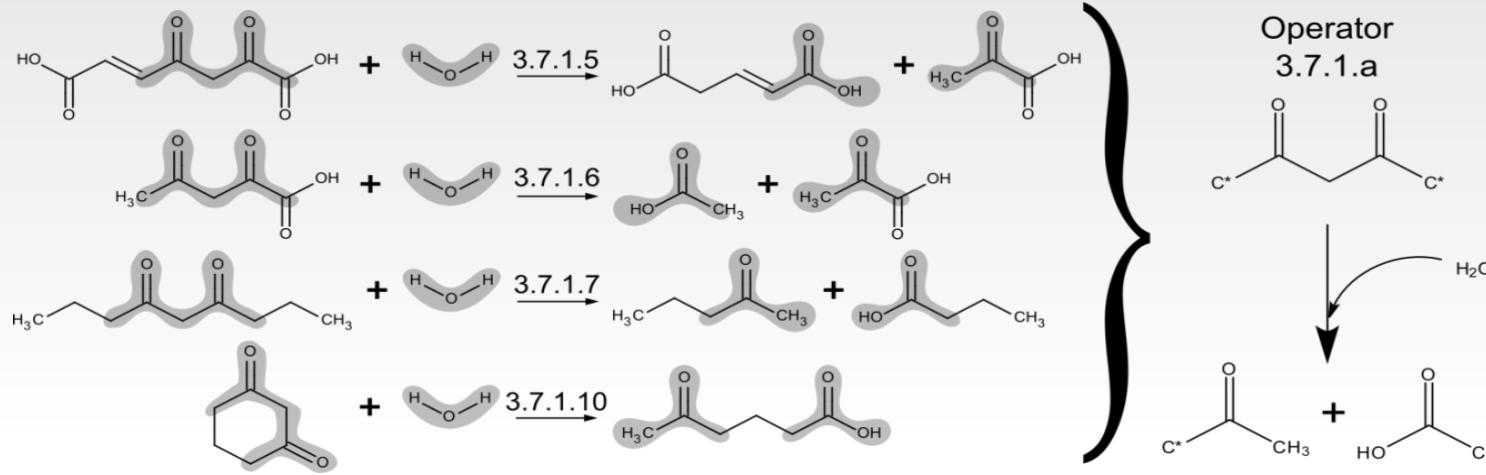


Unannotated Enzymes



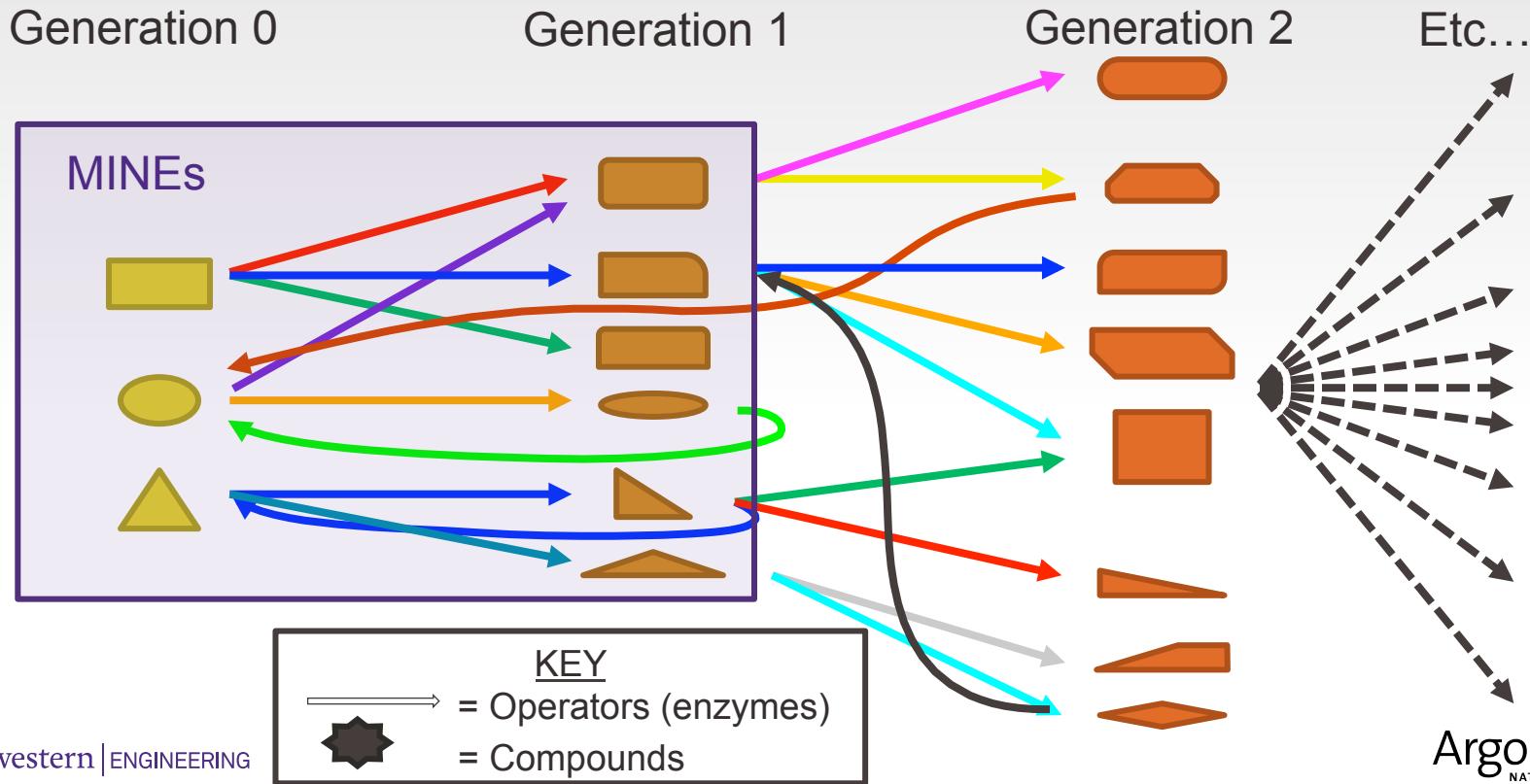
Spontaneous reactions

Chemical reactions are abstracted to build reaction rules

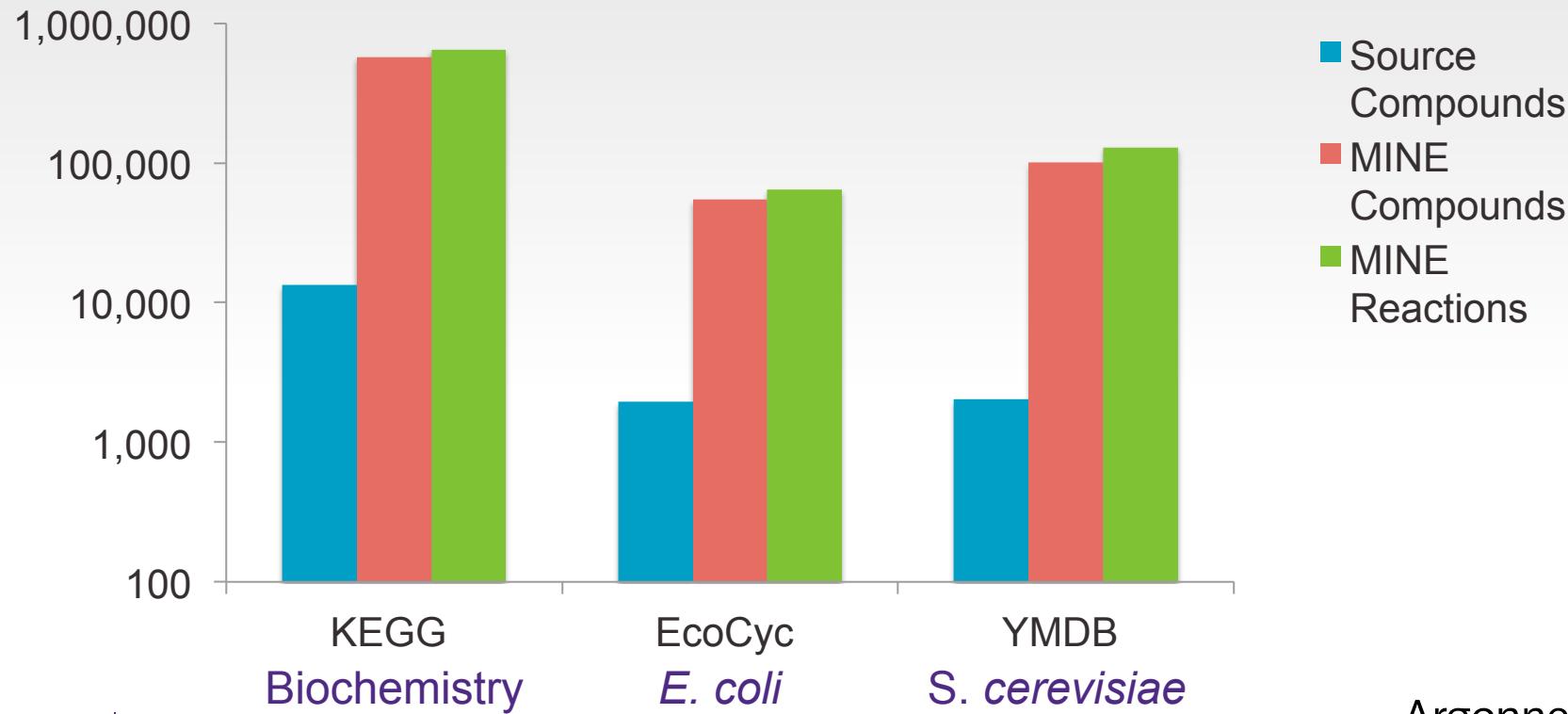


- More than 210 operators generalized from the Enzyme Commission classification system
- Performs substructure matching and compound transformation

Network Generation



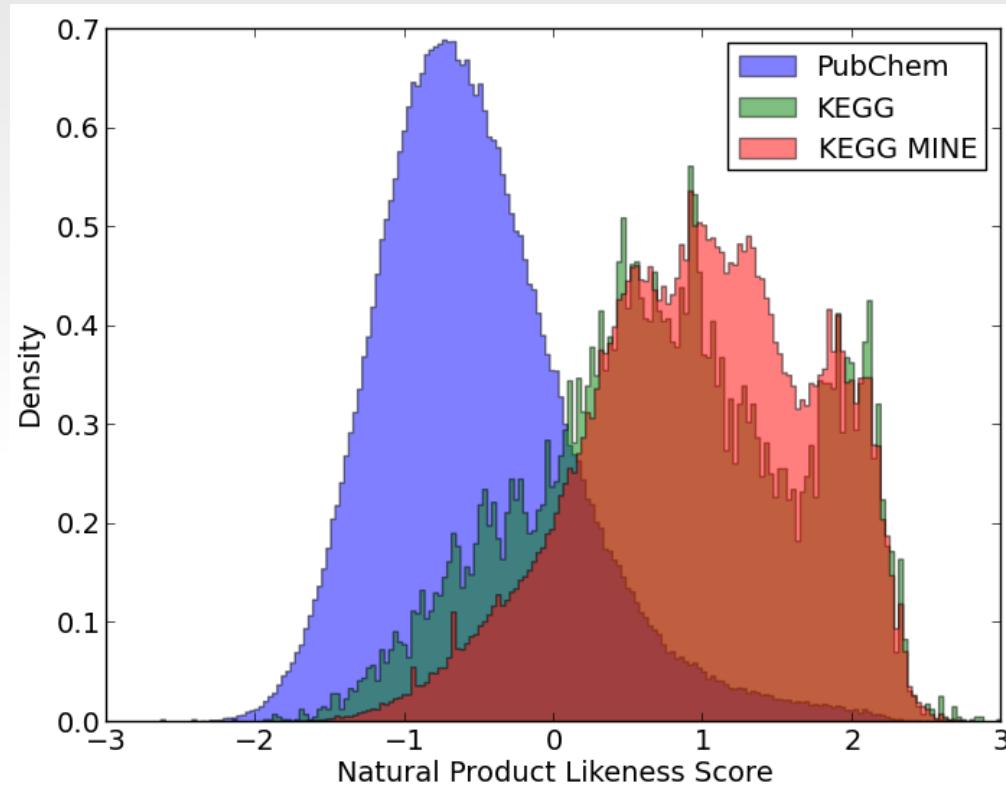
The MINEs are about 50X bigger than their sources



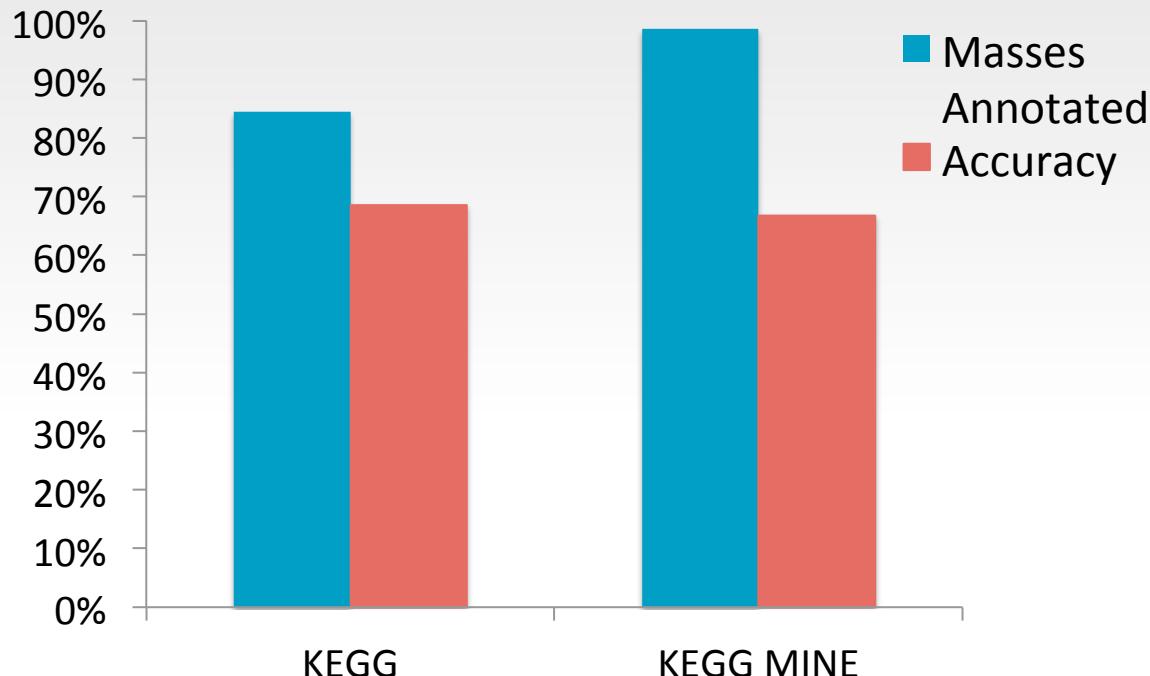
The MINEs are enriched for natural products

More Synthetic

More Natural



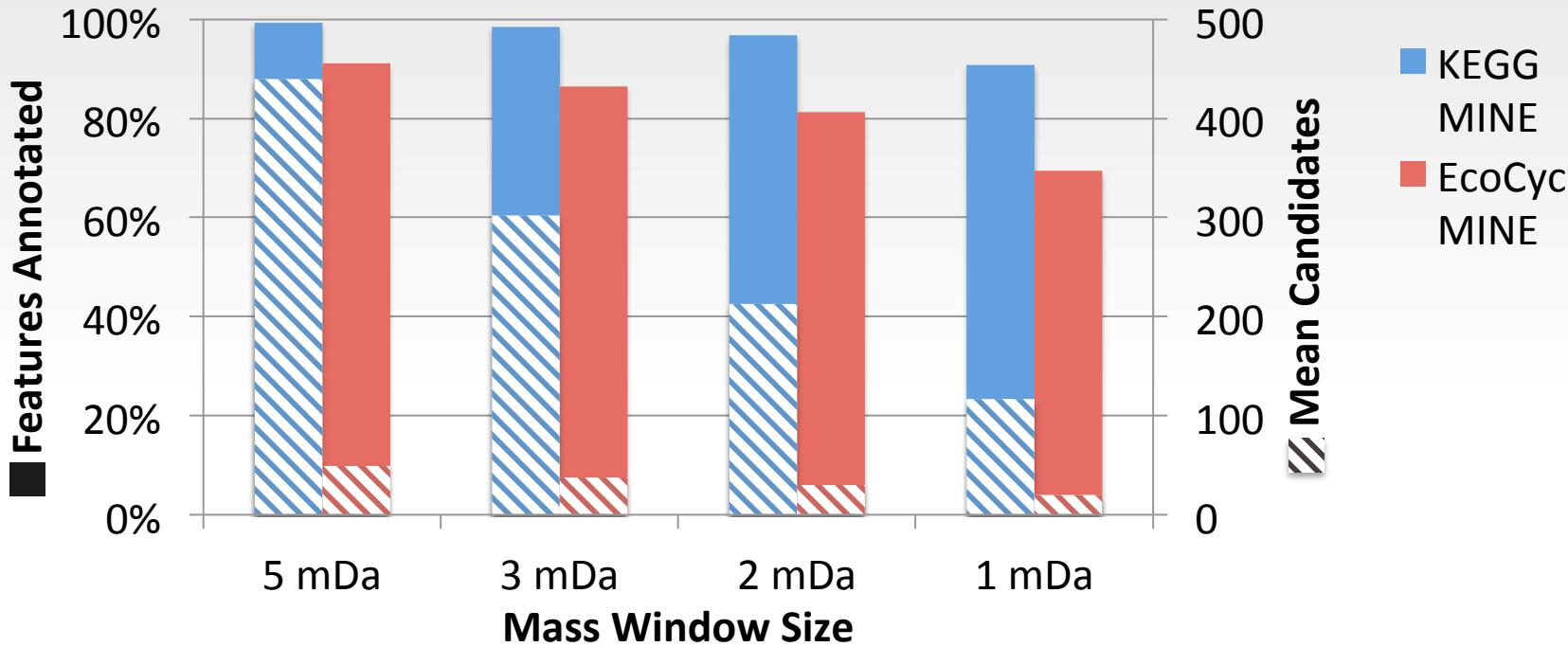
MINEs annotate more features than source databases while retaining accuracy



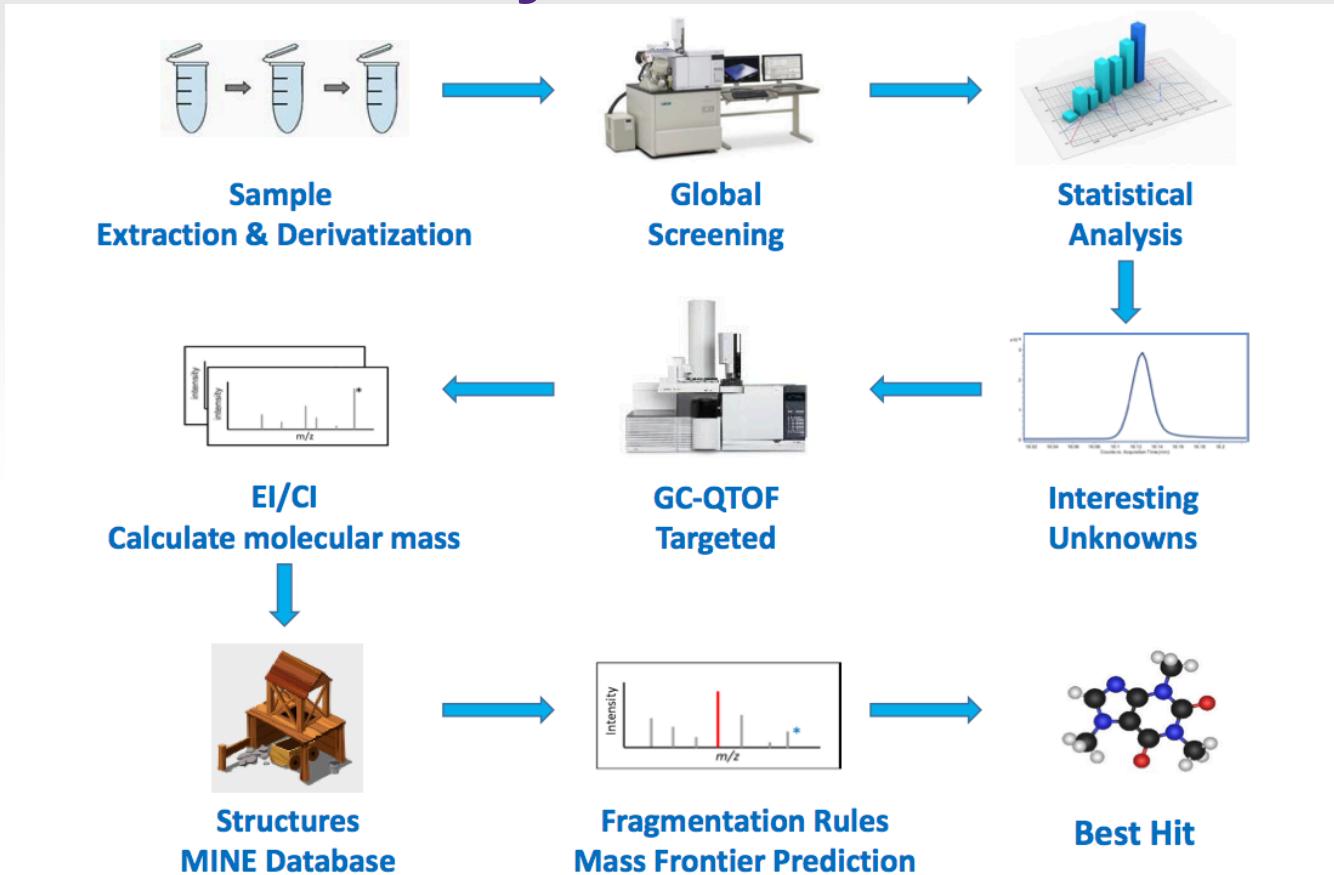
Testing set of 667 unique compounds with ESI spectra was compiled from MassBank

Search conducted with 2 mDa precision
 $[M]^+$, $[M+H]^+$,
 $[M+Na]^+$, $[M-H]^-$ and
 $[M+CH_3COO]^-$

Targeted databases dramatically reduce the number of candidates per feature

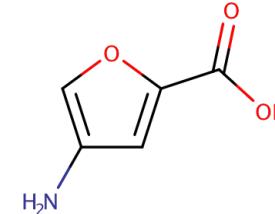
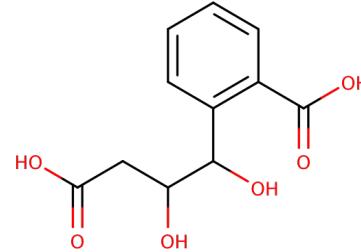
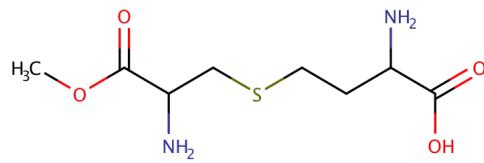
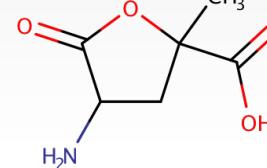
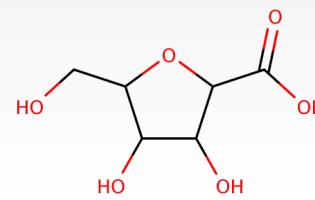
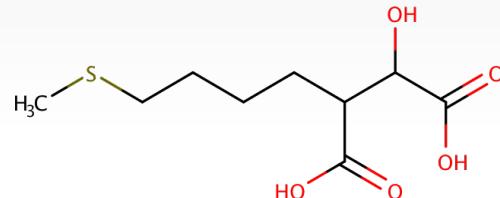
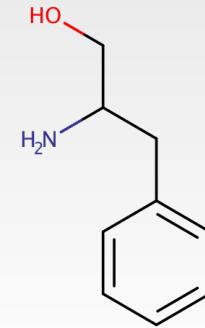
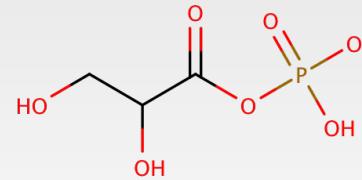
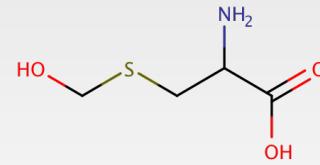
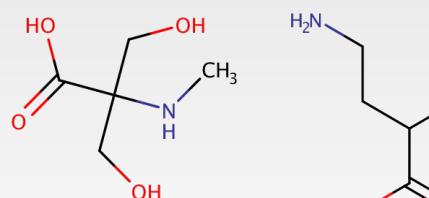


MINE databases have been integrated into a metabolite discovery workflow

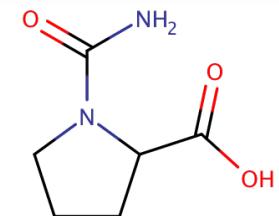
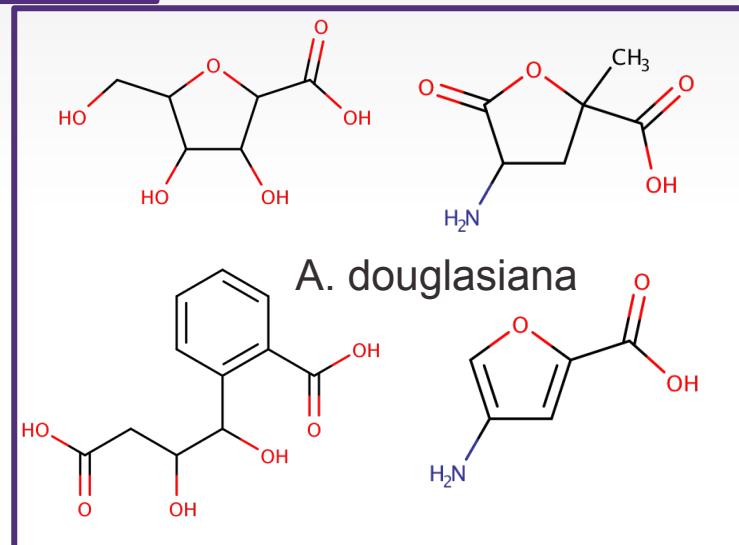
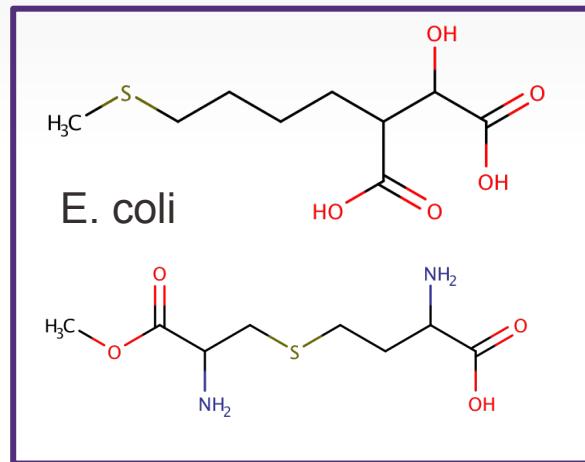
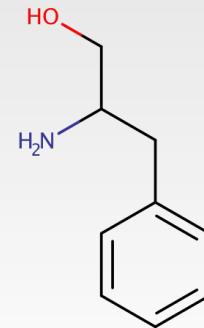
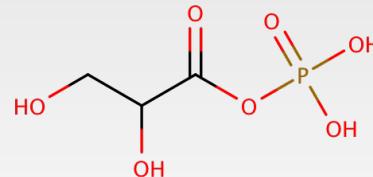
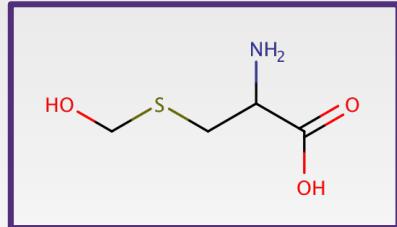
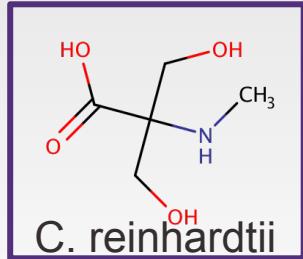


Credit: Zijuan Lai

Novel MINE compounds have been annotated by LC-MS²

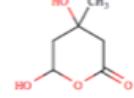
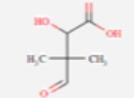
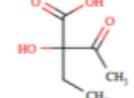


Annotated MINE Products



MINEs are freely available through a web app & API

Database 139 of 139 total hits Download Results First Previous 1 2 3 4 5 Next Last

M/Z	Adduct	Formula	Compound	MINE ID	logP	Kovats RI	NP Likeness
<input type="button" value="Search"/>	<input type="button" value="Search"/>	<input type="button" value="Search"/>	<input type="button" value="Search"/>	<input type="button" value="Search"/>	<input type="button" value="Sort by"/>	<input type="button" value="Sort by"/>	<input type="button" value="Sort by"/>
164.0937301	[M+NH4]+	C6H10O4		567037	-0.7	1183.8-1360.8	0
3-Hydroxy-5-oxohexanoate							
164.0937301	[M+NH4]+	C6H10O4		200425	-0.58	1255.9-1432.9	1.9863
Mevalonate							
164.0937301	[M+NH4]+	C6H10O4		49280	-0.17	1151.7-1328.7	1.6389
(R)-4-Dehydropantoate							
164.0937301	[M+NH4]+	C6H10O4		19516	0.24	1113.2-1290.2	0.9497
2-Aceto-2-hydroxybutanoate							

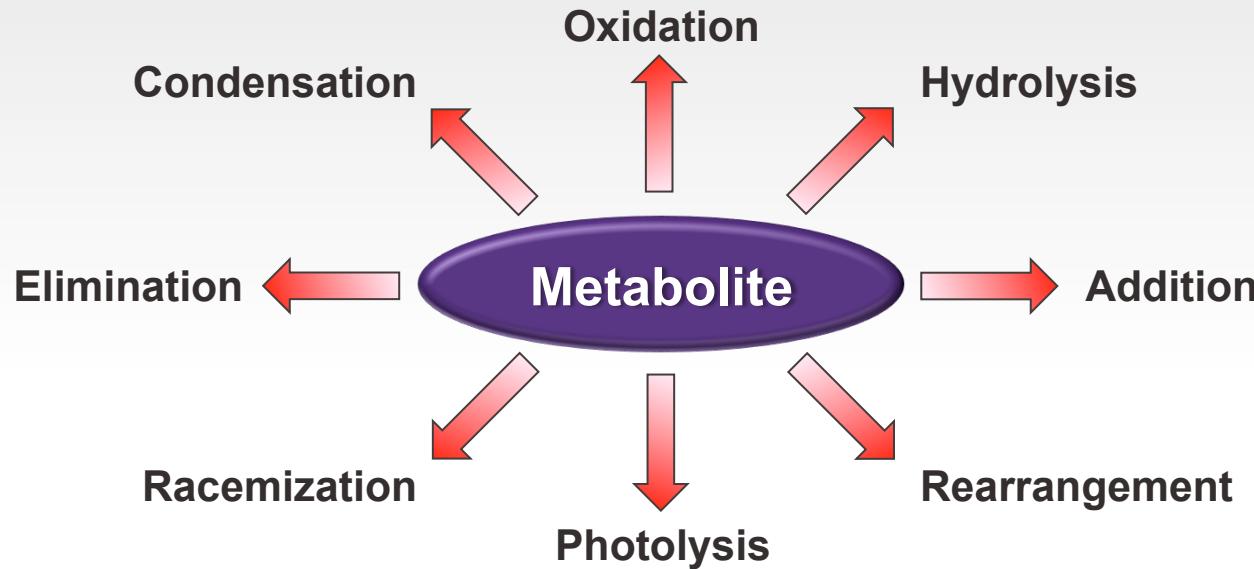
KEGG Genome Reconstructions

Escherichia

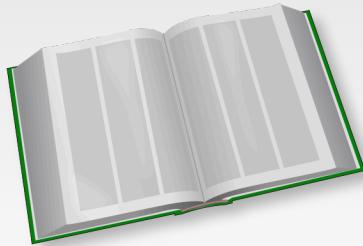
coli

Bacteria
Prokaryotes
Biological
Escherichia
Gammaproteobacteri
Campylobacter

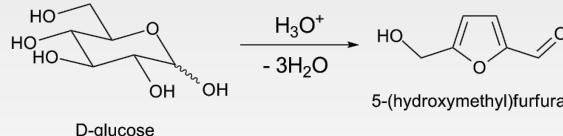
Metabolites undergo many transformations that are not mediated by enzymes



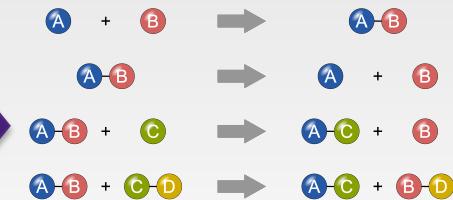
Predicting spontaneous metabolic reactions



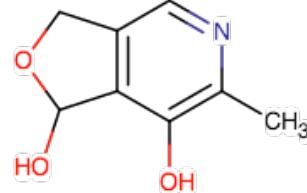
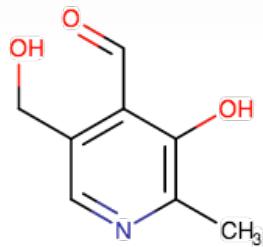
166 Publications



281 Spontaneous reactions



106 Reaction Rules

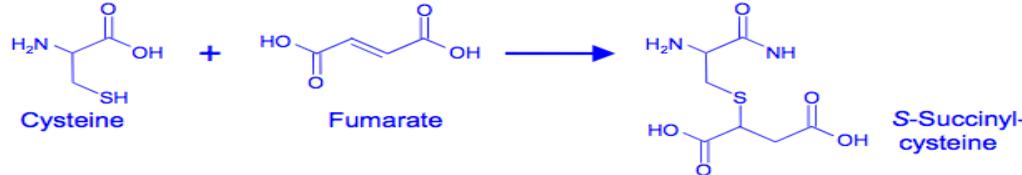


>5,000 reactive metabolites

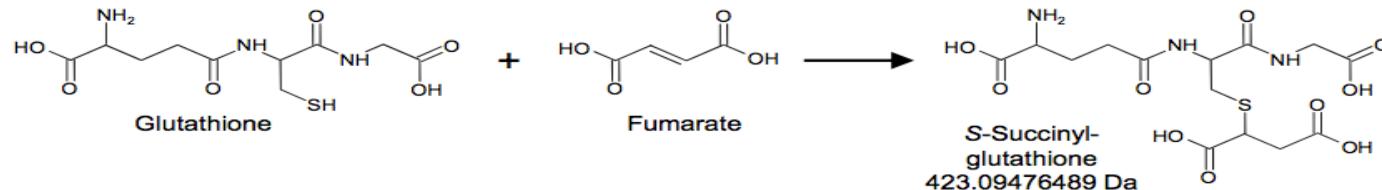
~72,000 reaction products

Two examples of from Literature

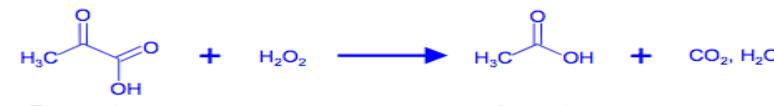
Basis



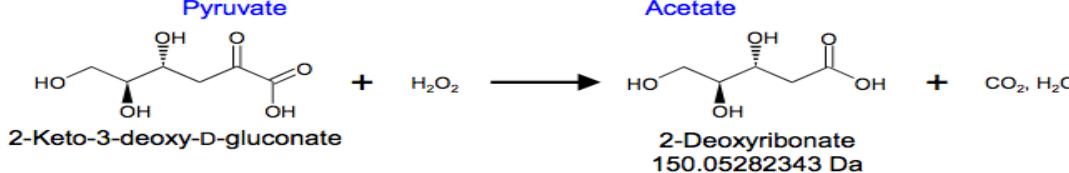
Prediction



Basis



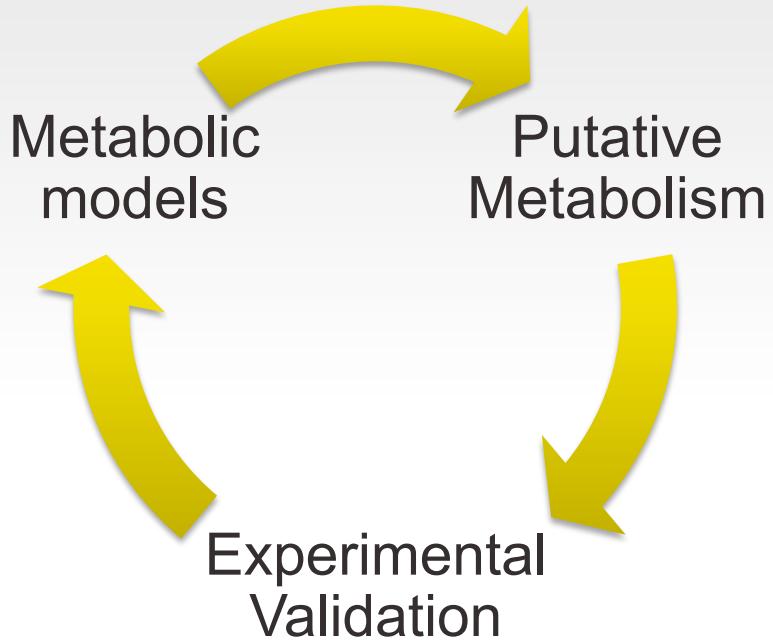
Prediction



Zheng et al. 2015, Sullivan et al. 2013
Lefevere et al. 1989

A method for exploring metabolic possibilities

- Freely available:
 - minedatabase.mcs.anl.gov
 - github.com/JamesJeffryes/MINE-API
- Active development:
 - MS² spectrum matching
 - Chiral representations
 - Integration into ModelSEED



Thank you!

- Keith Tyo (NU)
 - Dante Pertusi
 - Matt Moura
 - Trang Vu
- Linda Broadbelt (NU)
 - Andrew Stine
- Christopher Henry
(Argonne Nat'l Lab)
 - Ric Colestani
- Oliver Fiehn (UC-Davis)
 - Mona El-Badawi
 - Zijuan Lai
 - Tobias Kind
- Andrew Hanson (UF)
 - Claudia Lerma
 - Tom Niehaus
 - Oceane Frelin
 - Antje Thamm



Questions?

minedatabase.mcs.anl.gov