Supporting Information

A Cloud-Based Metabolite and Chemical Prioritization System for the Biology/Disease-driven Human Proteome Project

Kun-Hsing Yu[1,2], Tsung-Lu Michael Lee[3], Yu-Ju Chen[4], Christopher Ré[5], Samuel C. Kou[2], Jung-Hsien Chiang[6,], Michael Snyder[7,*,‡], Isaac S. Kohane[1,*,‡]*

[1] Department of Biomedical Informatics, Harvard Medical School

[2] Department of Statistics, Harvard University

[3] Department of Information Engineering, Kun Shan University, Taiwan

[4] Institute of Chemistry, Academia Sinica, Taiwan

[5] Department of Computer Science, Stanford University

[6] Department of Computer Science and Information Engineering, National Cheng Kung University, Taiwan

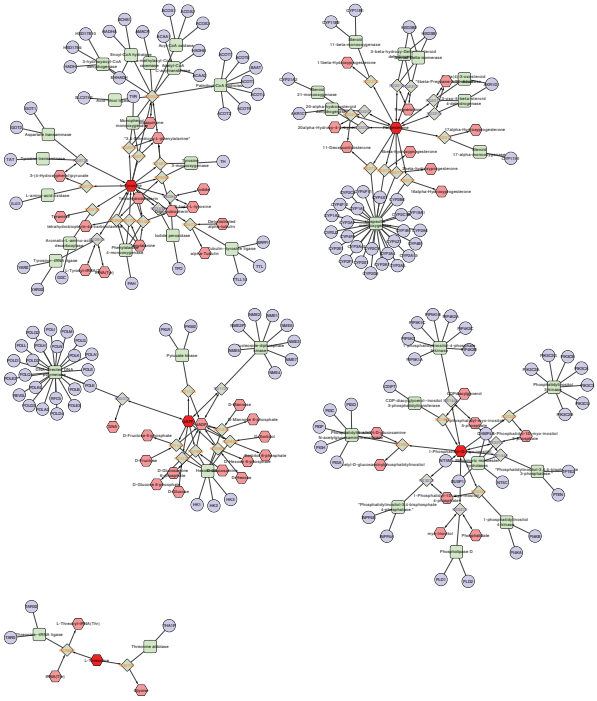
[7] Department of Genetics, Stanford University

Table of Contents:

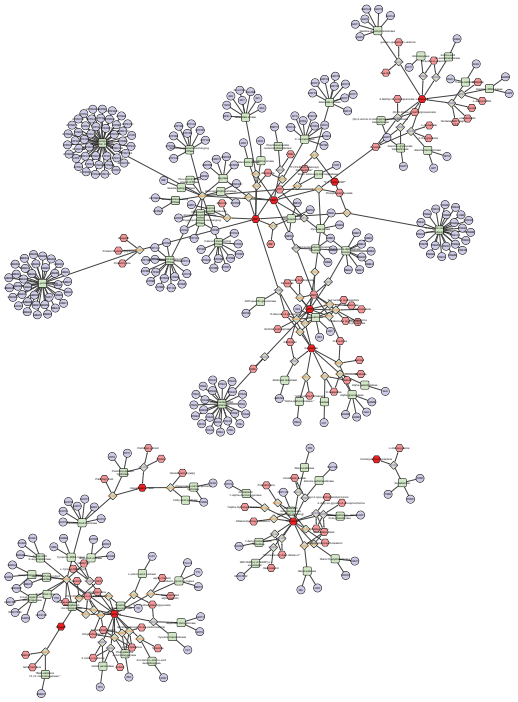
Figure S-1. Metscape visualized the pathways involved with the prioritized metabolites. Metabolites associated with cancers, diabetes, glycoproteomics and the musculoskeletal system are shown.

Figure S-1

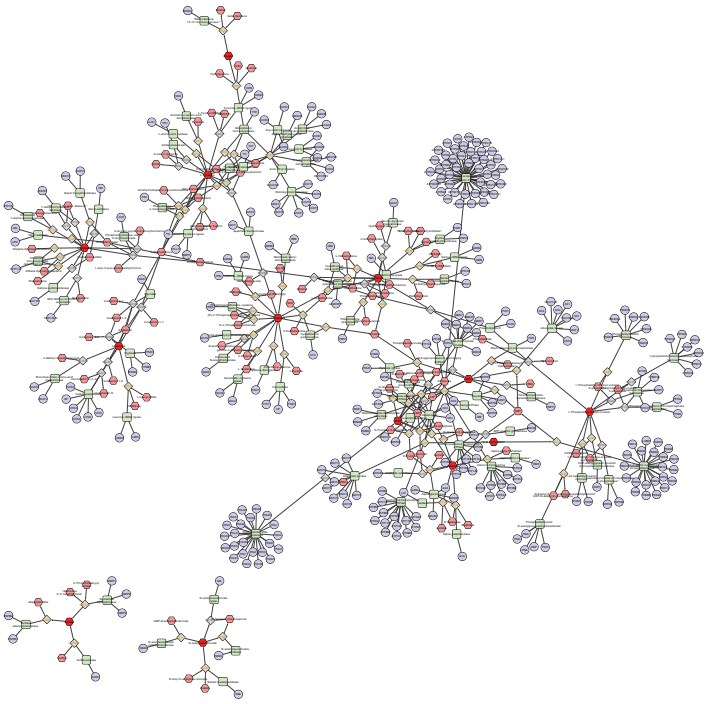
Cancers



Diabetes



Glycoproteomics



Musculoskelatal

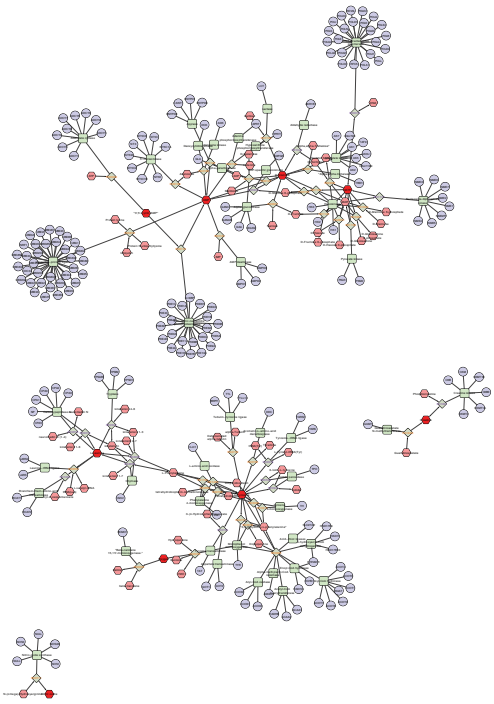


Figure S-2. The quantitative network statistics of the gene-metabolite interaction and gene-chemical interaction networks. (A) The centralization, efficiency, transitivity, and connectedness score of the gene-metabolite interaction networks. The gene-metabolite interaction networks of the four selected B/D-HPP targeted areas had high connectedness and moderate centralization, efficiency, and transitivity scores. (B) The centralization, efficiency, transitivity, and connectedness score of the gene-chemical interaction networks. Comparing with the gene-metabolite interaction networks, the gene-chemical interaction networks had slightly lower connectedness and centralization scores. The relative sparsity of the gene-chemical interaction networks contributed to the higher efficiency scores and varied transitivity scores.

Figure S-2

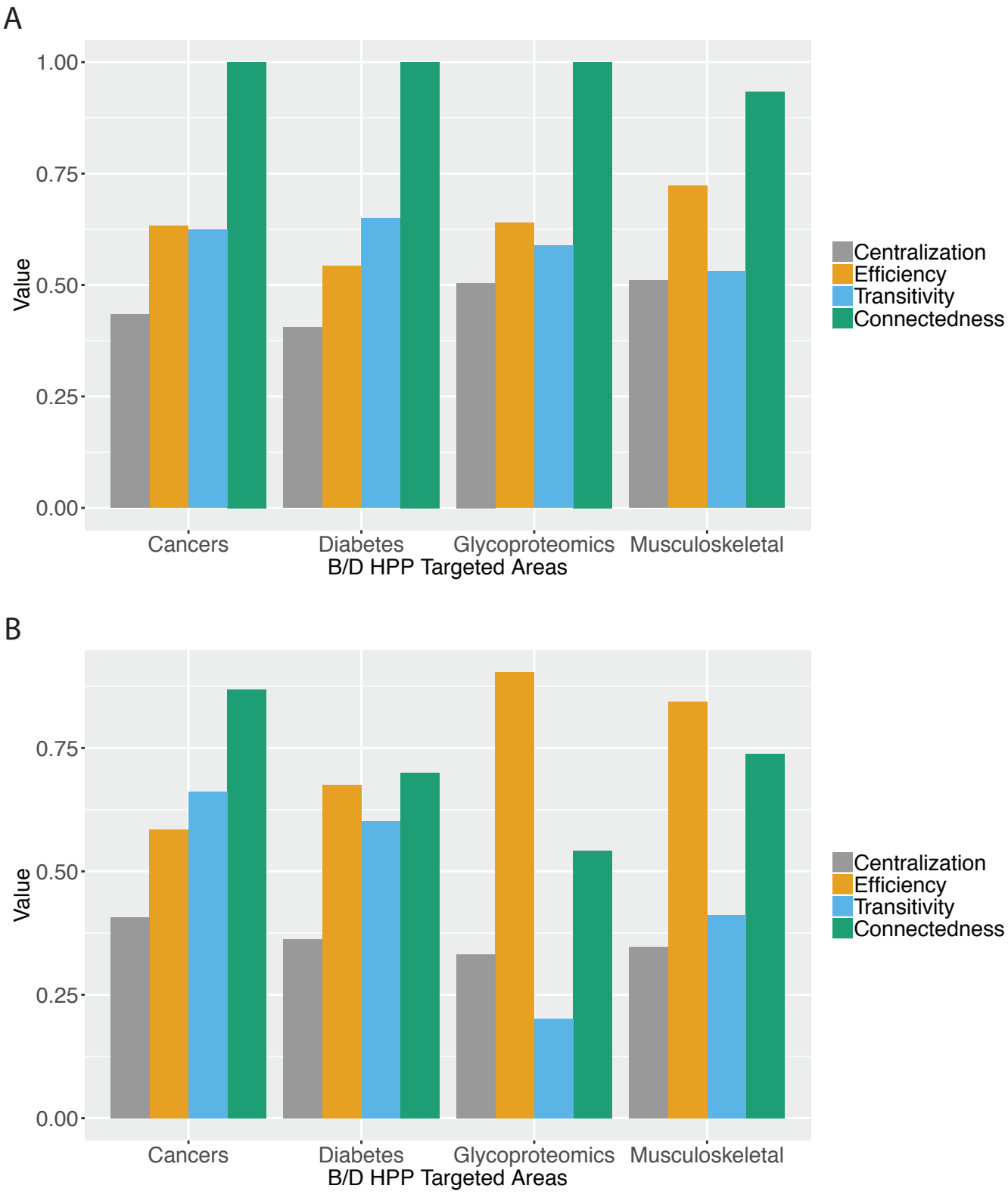

A
1.00
0.75
0.50
0.25
0.00
Value
Cancers
Diabetes
Glycoproteomics
Musculoskeletal
B/D HPP Targeted Areas
Centralization
Efficiency
Transitivity
Connectedness
B
0.75
0.50
0.25
0.00
Value
Cancers
Diabetes
Glycoproteomics
Musculoskeletal
B/D HPP Targeted Areas
Centralization
Efficiency
Transitivity
Connectedness

Figure S-3. Multidimensional scaling (MDS) visualization of the connections among B/D-HPP targeted fields and their associated chemicals. B/D-HPP fields with higher correlation in their associated chemicals' PURPOSE scores have shorter distances on the graph.

Figure S-3

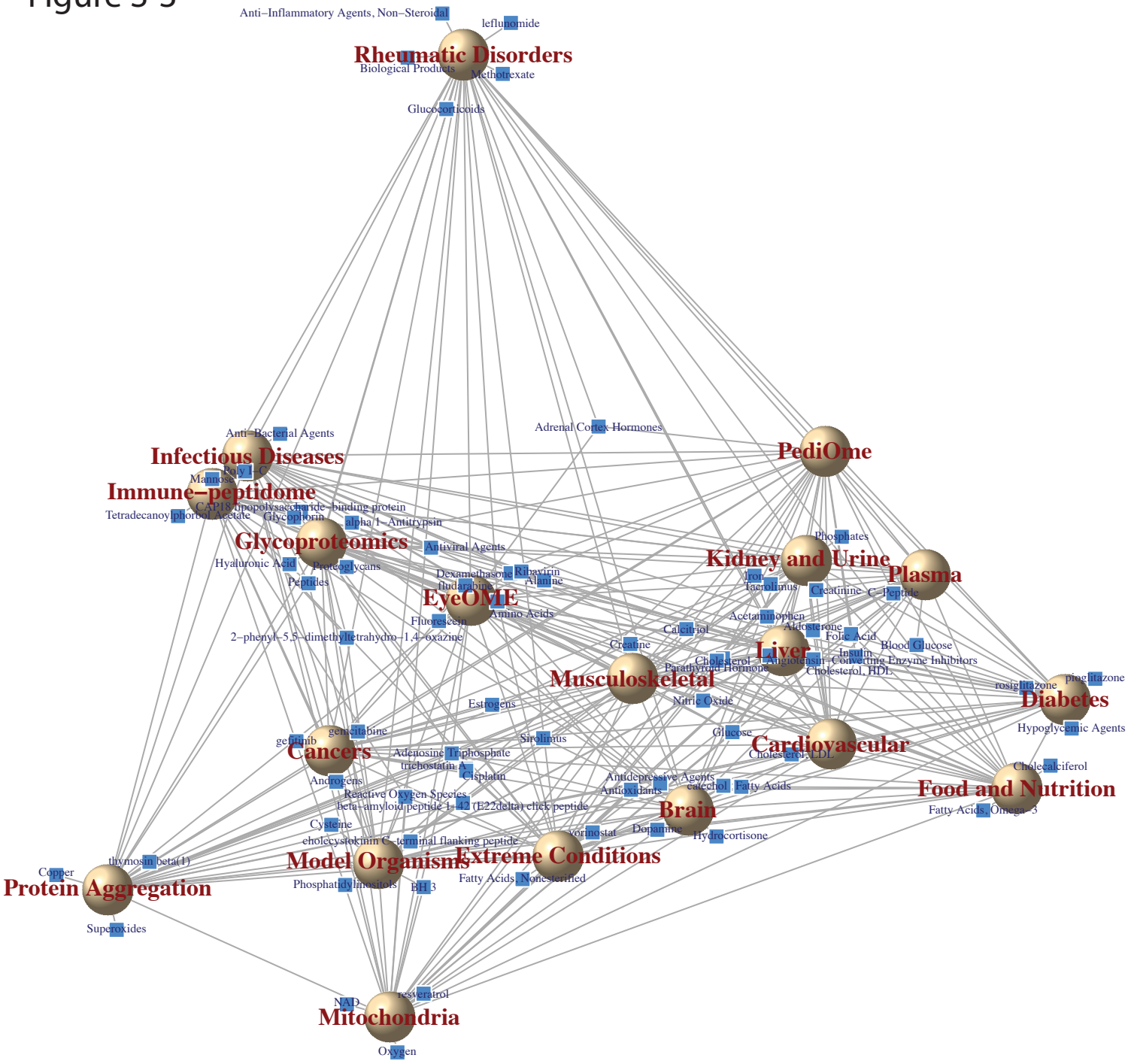

Anti–Inflammatory Agents, Non–Steroidal
leflunomide
Rheumatic Disorders
Biological Products
Methotrexate
Glucocorticoids
Adrenal Cortex Hormones
Anti–Bacterial Agents
Infectious Diseases
PediOme
Poly I-C
Mannose
Immune–peptidome
Tetradecanoylphorbol Acetate
CAP18 Lipopolysaccharide–binding protein
Glycophorin
alpha 1–Antitrypsin
Glycoproteomics
Antiviral Agents
Hyaluronic Acid
Proteoglycans
Peptides
Phosphates
Kidney and Urine
Plasma
Dexamethasone
Ribavirin
fludarabine
Alanine
Iron
Tacrolimus
Creatinine
C–Peptide
EyeOME
Amino Acids
Fluorescein
Acetaminophen
Aldosterone
Folic Acid
2–phenyl–5,5–dimethyltetrahydro–1,4–oxazine
Calcitriol
Creatine
Liver
Blood Glucose
Insulin
Cholesterol
Angiotensin–Converting Enzyme Inhibitors
Parathyroid Hormone
Cholesterol, HDL
Musculoskeletal
pioglitazone
rosiglitazone
Estrogens
Nitric Oxide
Diabetes
Hypoglycemic Agents
gemcitabine
gefitinib
Cancers
Sirolimus
Glucose
Cardiovascular
Adenosine Triphosphate
trichostatin A
Cholesterol, LDL
Cisplatin
Androgens
Cholecalciferol
Antidepressive Agents
Antioxidants
Catechol
Fatty Acids
Food and Nutrition
Reactive Oxygen Species
beta–amyloid peptide 1–42 (E22delta)
Brain
Fatty Acids, Omega–3
Cysteine
Dopamine
vorinostat
Hydrocortisone
cholecystokinin C–terminal flanking peptide
thymosin beta(1)
Model Organisms
Extreme Conditions
Copper
Protein Aggregation
Phosphatidylinositols
BH3
Fatty Acids, Nonesterified
Superoxides
resveratrol
NAD
Mitochondria
Oxygen

Table S-1. The PubMed search terms for the B/D-HPP targeted fields.	
B/D HPP Targeted Fields	**PubMed Search Terms**
Brain	brain
Cancers	cancer
Cardiovascular	cardiovascular
Diabetes	diabetes
EyeOME	eye OR ocular
Food and nutrition	food OR nutrition OR nutrients
Glycoproteomics	glycoproteins
Immune-peptidome	immune OR immune system
Infectious diseases	infectious OR infection
Kidney and urine	kidney OR urine
Liver	liver OR hepatic
Mitochondria	mitochondria
Model organisms	rat OR mouse
Musculoskeletal	muscle OR bone OR musculoskeletal
PediOme	pediatric OR newborn OR infant OR toddler OR child OR adolescent
Plasma	plasma OR serum
Protein aggregation	protein aggregation NOT platelet aggregation
Rheumatic disorders	rheumatic
Extreme conditions	hot OR cold OR alkaline condition OR acidic condition OR hypersaline OR radiation