Computational Approaches for Big Data Analytics & Integration

March 16, 2017

Jianguo (Jeff) Xia, PhD

McGill University, QC Canada

Outline

- Overview of big omics data
- Approaches in big data analytics
- Approaches in big data integration
- Web-based tools for common omics data analysis & integration

The Promises of Big Omics Data

Comprehensive molecular profiles

- Global systems overview
- Less biased
- Robust to noise?

What we can obtain from this type of data?

- Patterns & trends
  - Group behaviors
  - Collective functions
  - Hypothesis generation
- Mechanisms & knowledge
  - Mechanistic understanding
  - Pathways
  - Networks

Reality: the "Happy" Middle

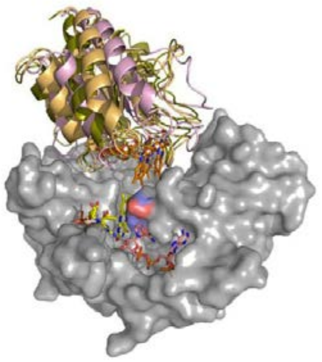
Reductionism

Knowledge of everything

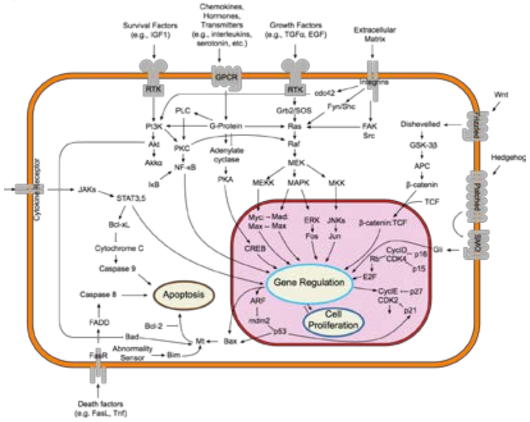
Individual Molecule

Big Data Analytics

Systems Biology



- Optimal simplicity
- Empowering
- Achievable

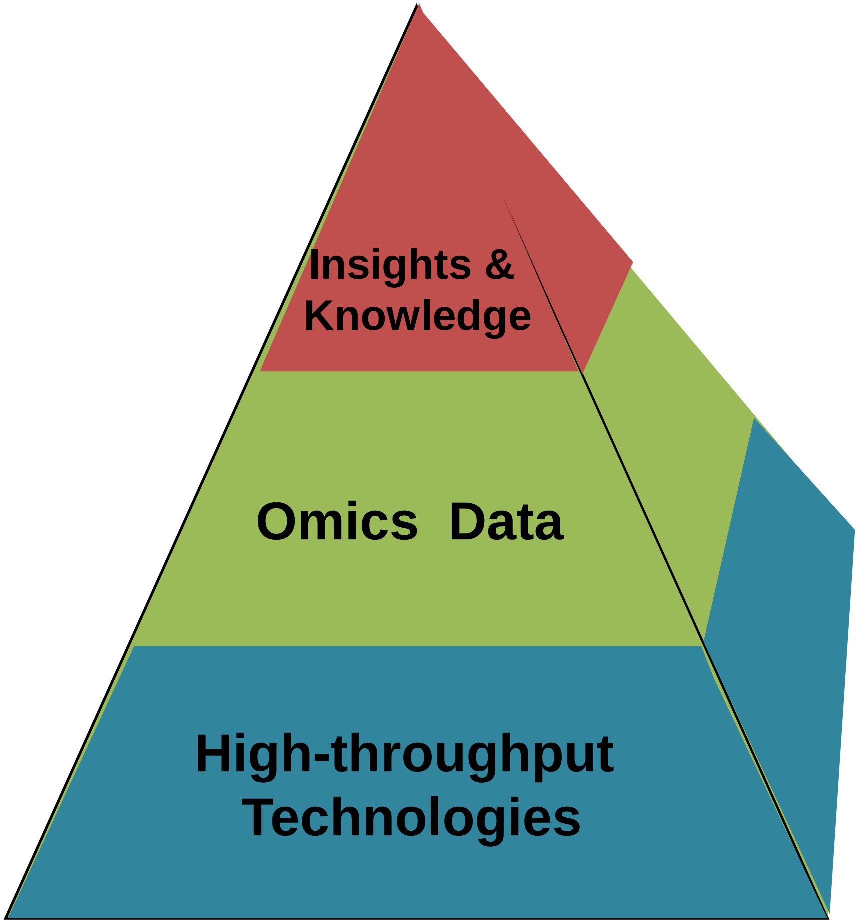


Mechanism

Patterns & Trends

Knowledge

The Big Omics Data



Workflow of Omics Data Analytics

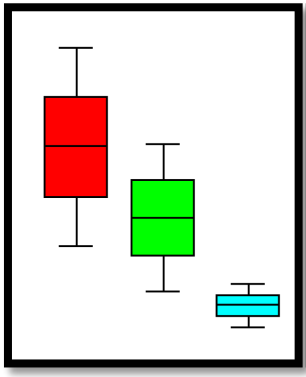
Raw Data processing

Platform-specific

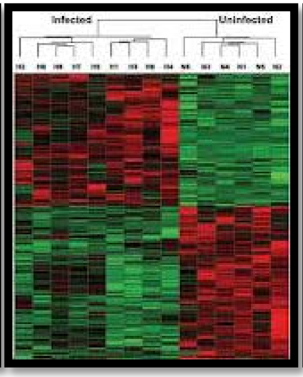
- **RNAseq**
- **ChIP-seq**
- **Mass spectrometry**

Statistics & visualization

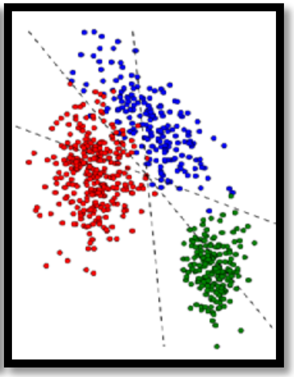
Comparison



Clustering



Classification



Functional interpretation

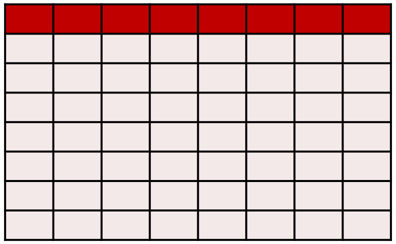
Omics-specific

- **Enrichment analysis**
- **Pathway analysis**
- **Network analysis**

Two distinct big data challenges

- **Size challenge** (raw data)
  - Raw reads, spectra, images
  - Large (100s MB ~ GB)
  - Large storage and computing resources

- **Complexity challenge** (feature table)
  - Feature table (abundance, intensities)
  - Small (100s KB ~ MB)
  - High-dimensional, missing values
  - Data integration usually starts here

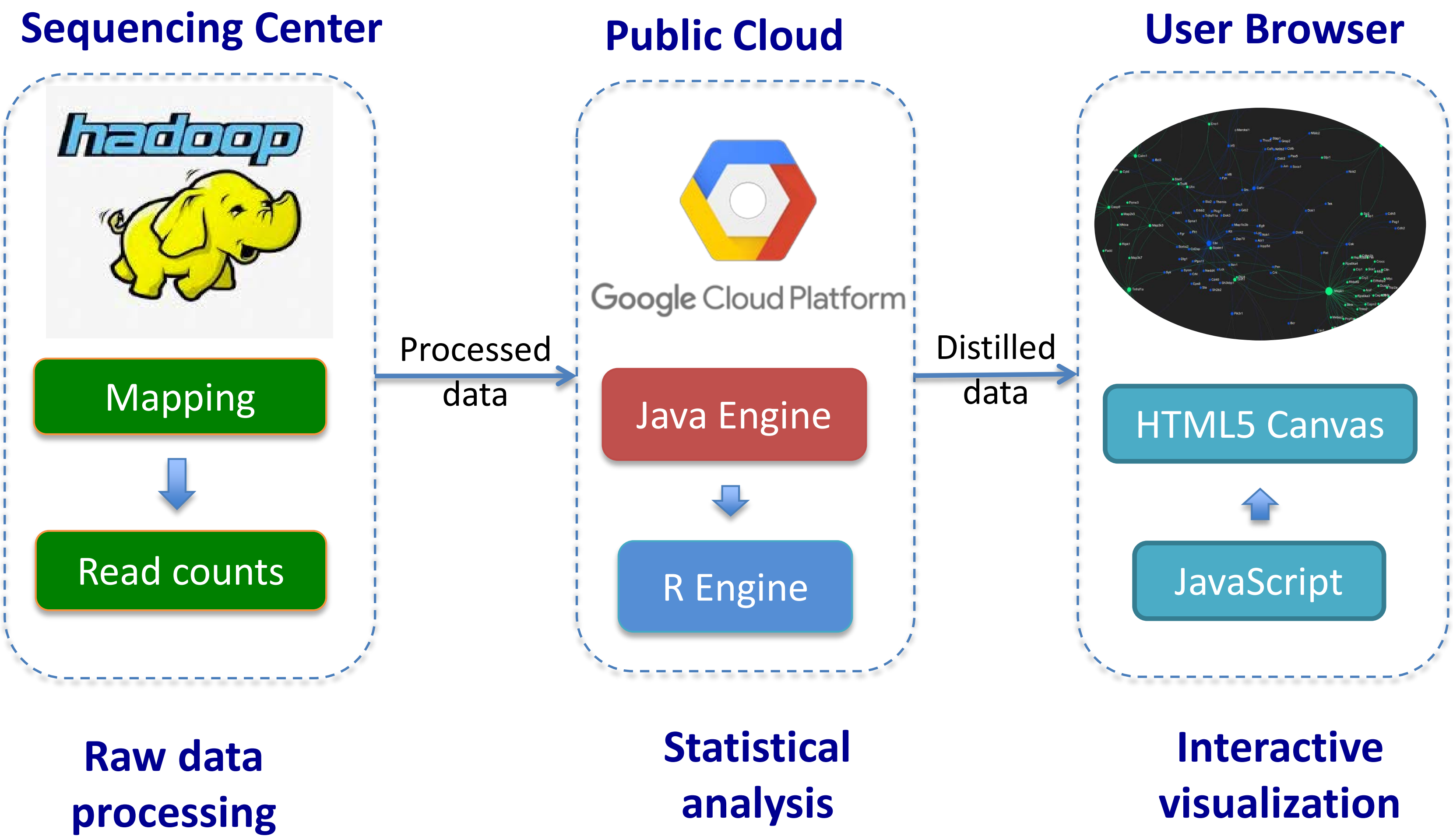


→ **Divide and conquer**

Dealing with big omics data

- **Bringing analysis to (raw) data**
  - Too big to transfer around
  - Done **locally** at the same place generating the data (i.e. omics centers)
  - Following standardized operation protocols (SOP)
- **Bringing (processed) data to analysis**
  - Usually done by individual researchers
  - Can be easily uploaded to tools deployed on **cloud**
  - Highly domain-specific and individualized

Local + Cloud + Browser



Sequencing Center

Mapping

Read counts

Raw data processing

Processed data

Public Cloud

Google Cloud Platform

Java Engine

R Engine

Statistical analysis

Distilled data

User Browser

HTML5 Canvas

JavaScript

Interactive visualization

Big Data Analytics

Big Data Analytics

Three Keys Factors

1. Data distillation
   - Not all data are useful
2. Coupling statistics with visualization
   - Support for user exploration
3. Connecting with prior knowledge
   - Validation previous findings & identification of novelties

Data distillation

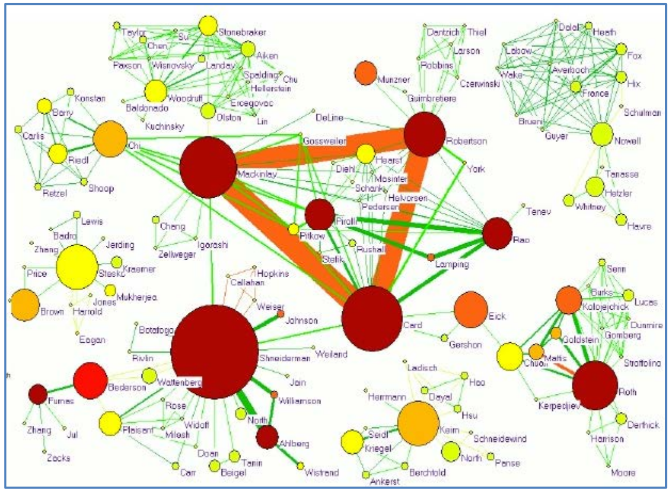
Raw data → processed data → informative data

- Only a small portion of molecules will respond to perturbations.
- The majority will still maintain their “normal” states (homeostasis)

Statistics + Visualization

Bringing data closer to domain experts

- Big omics data is complex and heterogeneous
- Statistics often fail to capture the data characteristics
- Disruptive discovery are often driven by the hunches and leaps of faith of the researcher by interacting with the data

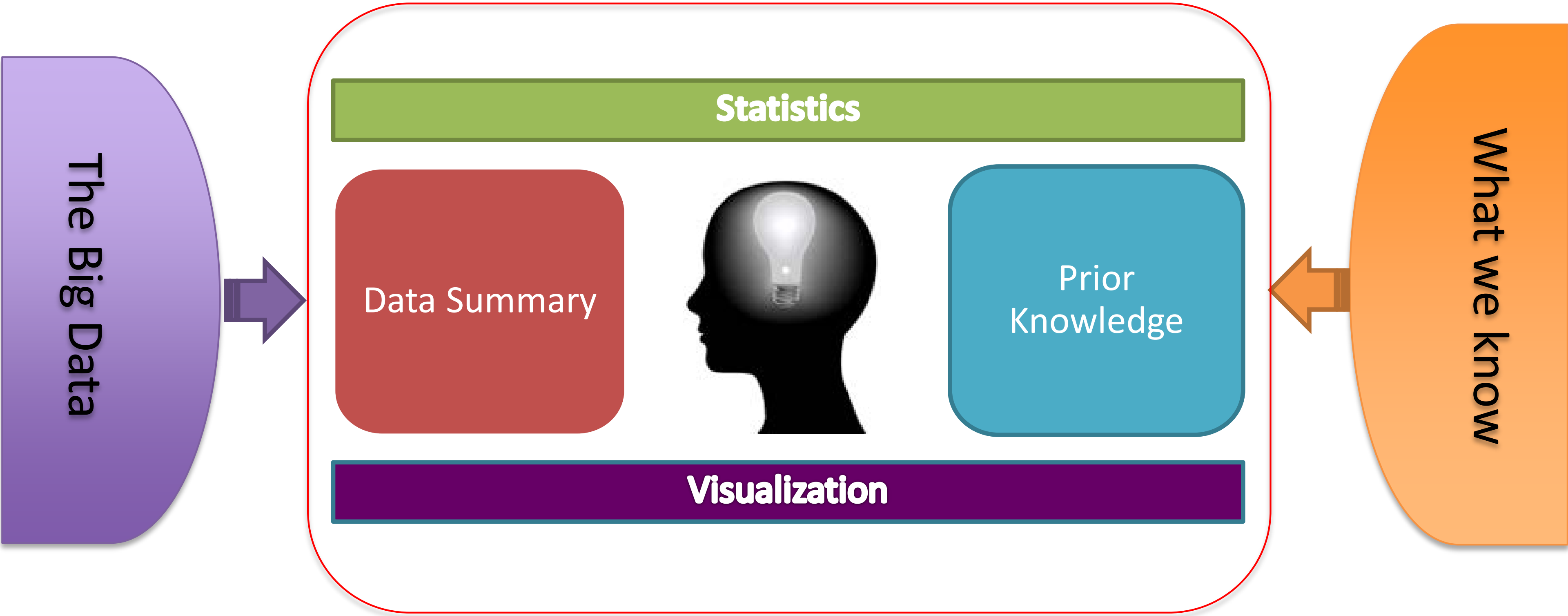


Statistics + Domain Knowledge

Connecting new data with prior knowledge

- Encode knowledge into computable forms
  - Gene sets, pathways, networks
- Evaluating new data within the context of our knowledge
  - Enrichment analysis
  - Pathway analysis
  - Network analysis

Big data analytics in a nutshell



The Big Data

Statistics

Data Summary

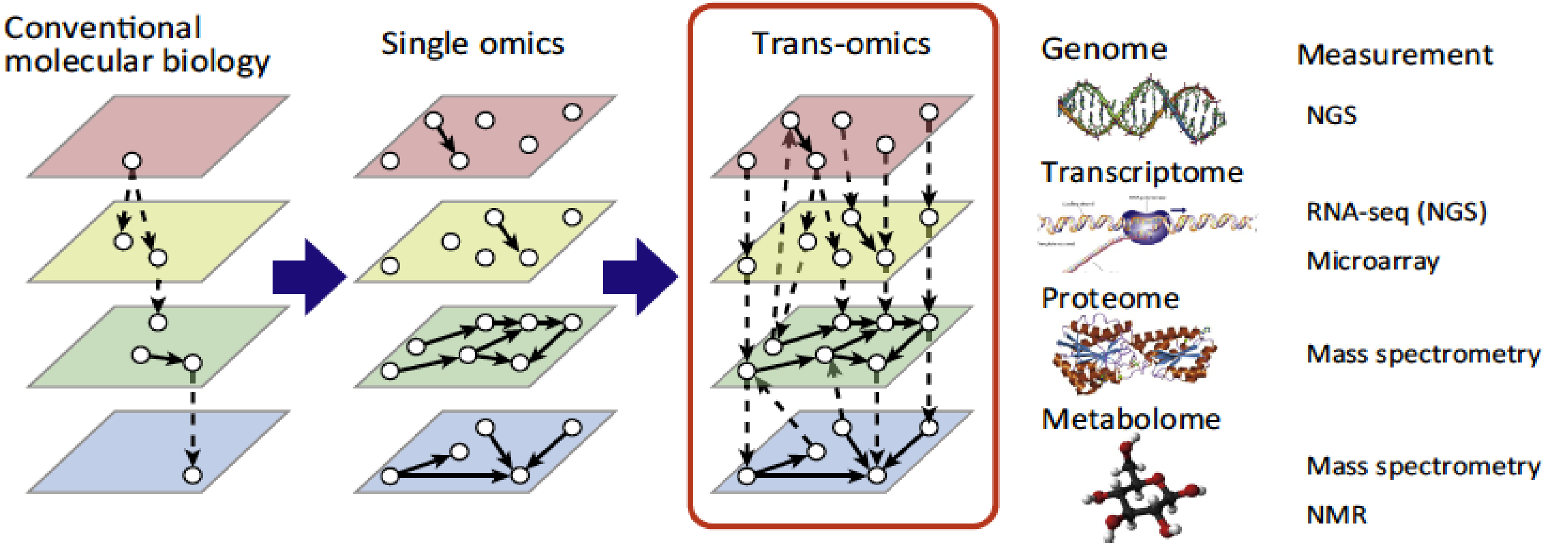
Prior Knowledge

Visualization

What we know

Omics Data Integration

This is our goal ....



Trends in Biotechnology, April 2016, Vol. 34, No. 4

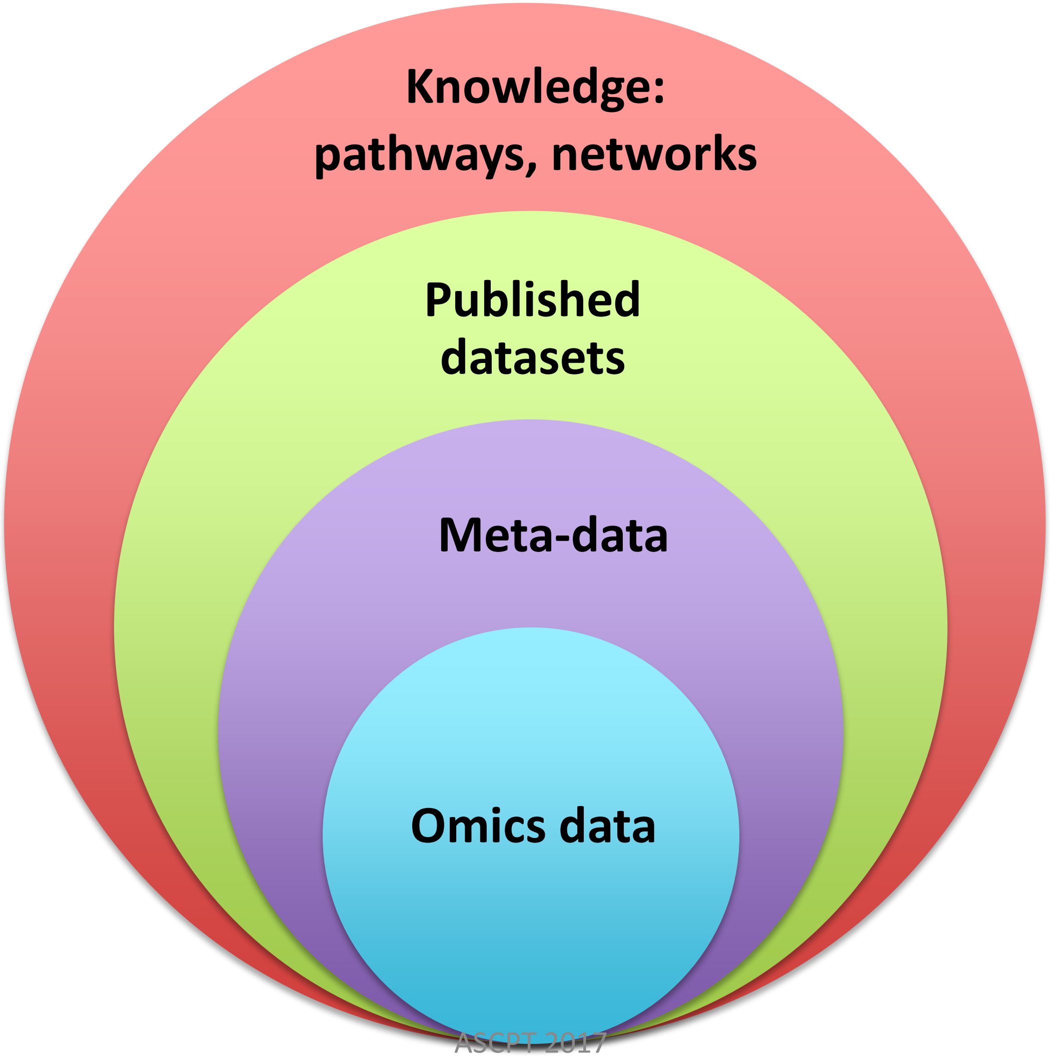
Mechanism ← Patterns & Trends → Knowledge

Big Data Analytics

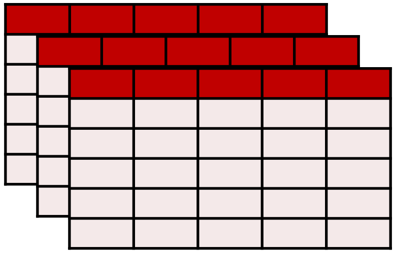
Current Status

- Done for simple organisms under well-controlled experiments
  - E. coli, yeast …
  - Time series to resolve false patterns
- For human studies, the limiting factor is more technological than computational
  - Longitudinal omics data collections for a large cohorts
  - Data sharing with public for algorithm benchmarking

The ‘ecosystem’ of big omics data



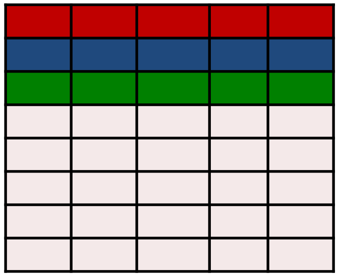
Data integration - three common scenarios



One disease;

One omics level;

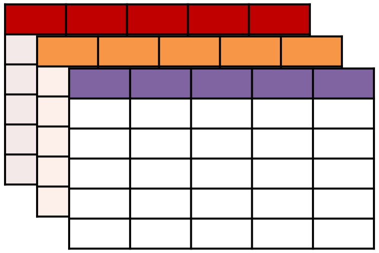
Multiple datasets;



One omics data;

Multiple clinical parameters:

- Diagnosis
- Age, Gender, Ethnicity, Smoking ….



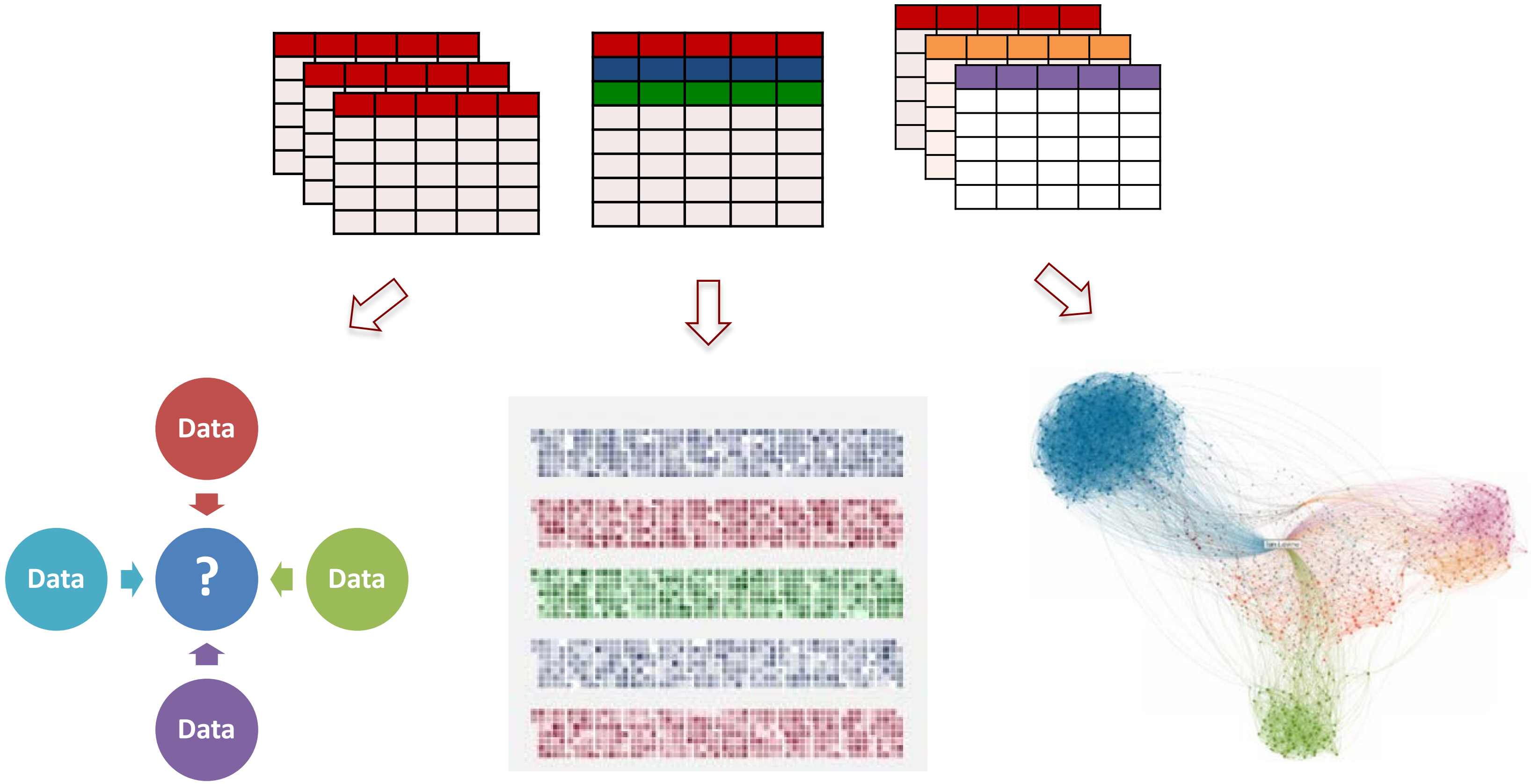
One disease;

Multiple omics level;

Multiple datasets;

Computational Approaches

Omics Data



Statistical integration **Visual integration** **Network integration**

Web-based Tools for Big Omics Data Analysis & Integration

- MetaboAnalyst (since 2009)
  - www.metaboanalyst.ca
  - Metabolomics data analysis & integration
- NetworkAnalyst (since 2012)
  - www.networkanalyst.ca
  - Transcriptomics data analysis & integration
- miRNet (since 2015)
  - www.mirnet.ca
  - miRNA data analysis & integration
- MicrobiomeAnalyst (since 2017)
  - www.microbiomeanalyst.ca
  - Microbiome data analysis & integration

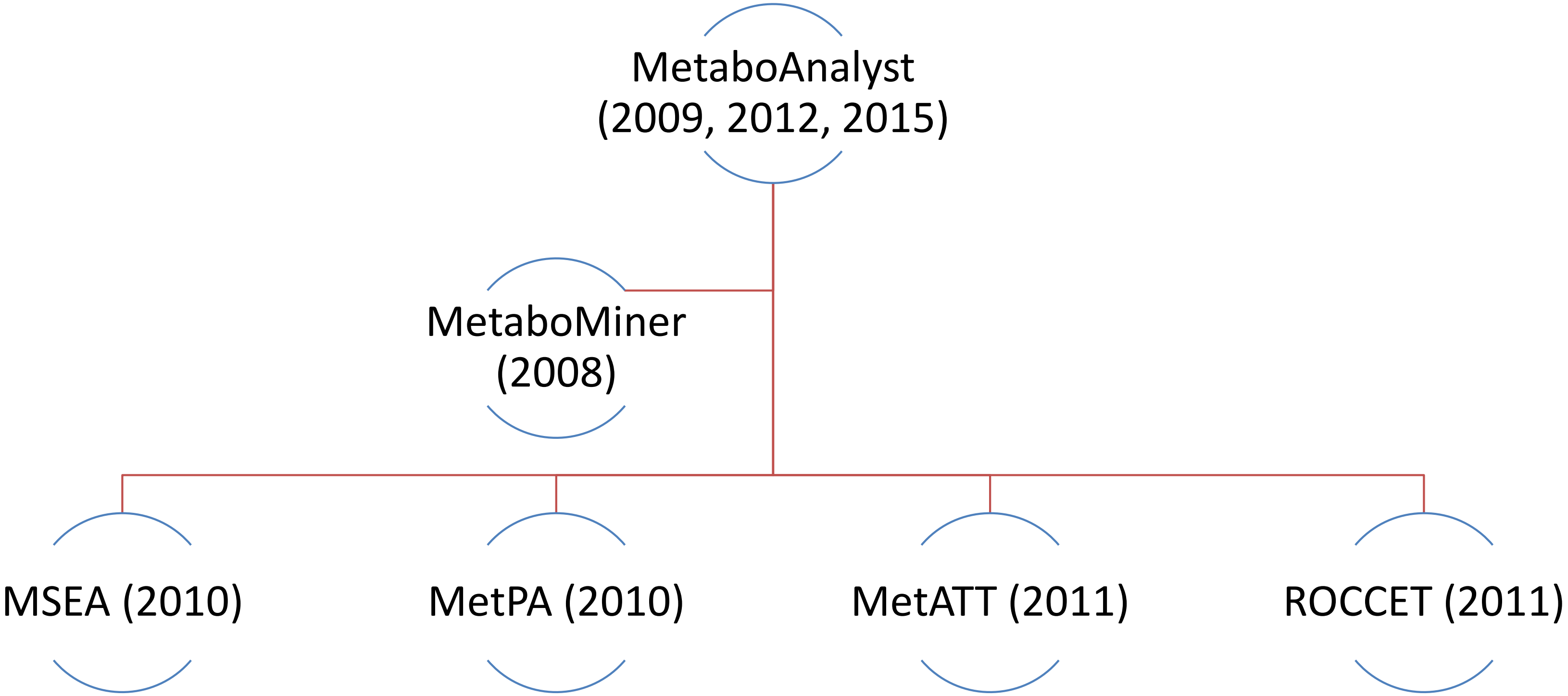
2009 - CURRENT

Metabolomics & MetaboAnalyst

http://www.metaboanalyst.ca

- Metabolomics (and integration with genomics)
- Real-time interactive data analysis
- 100,000 users, > 6,000 jobs submitted per day

A Roadmap of MetaboAnalyst



MetaboAnalyst (2009, 2012, 2015)

MetaboMiner (2008)

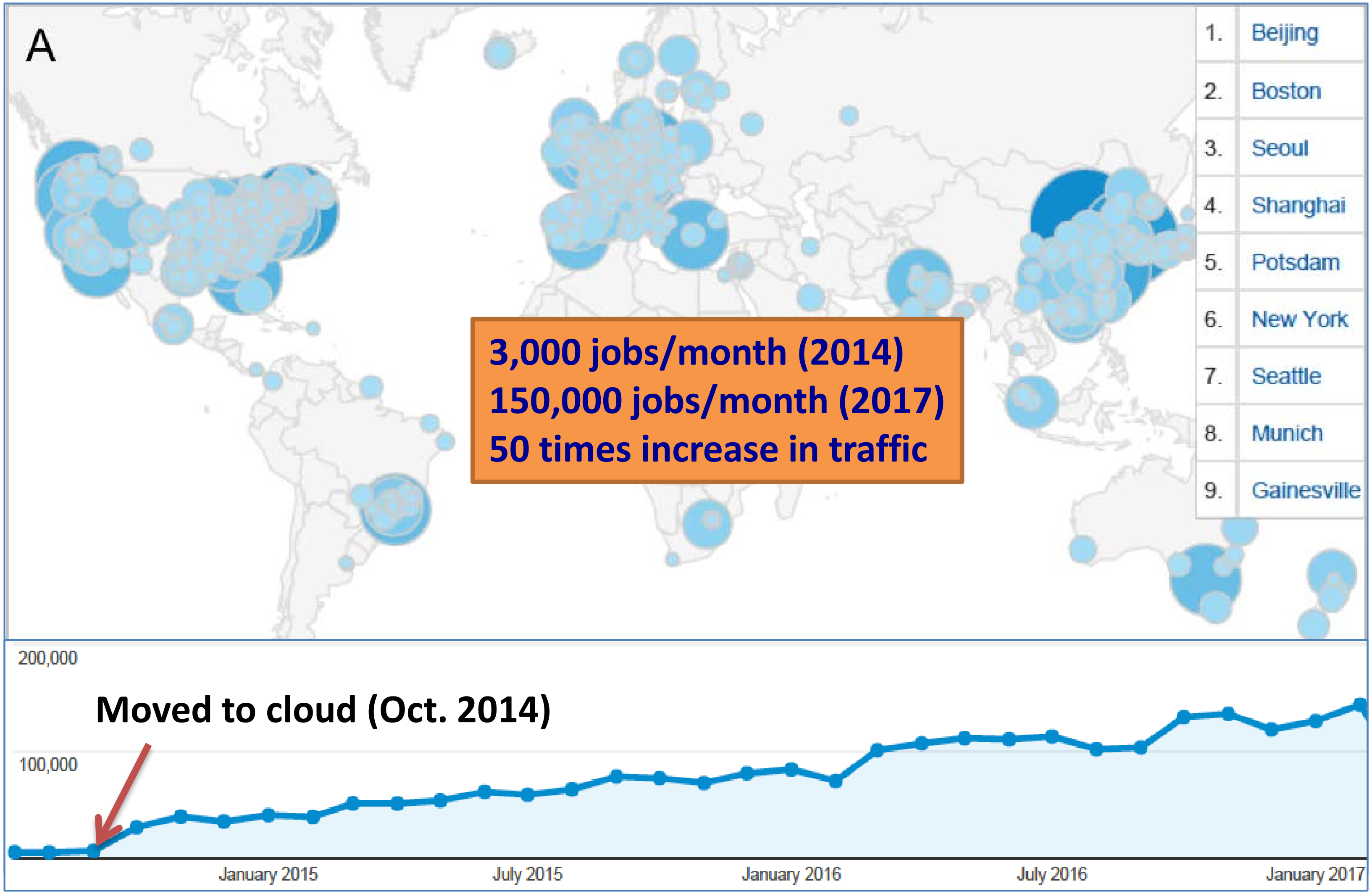
MSEA (2010)

MetPA (2010)

MetATT (2011)

ROCCET (2011)

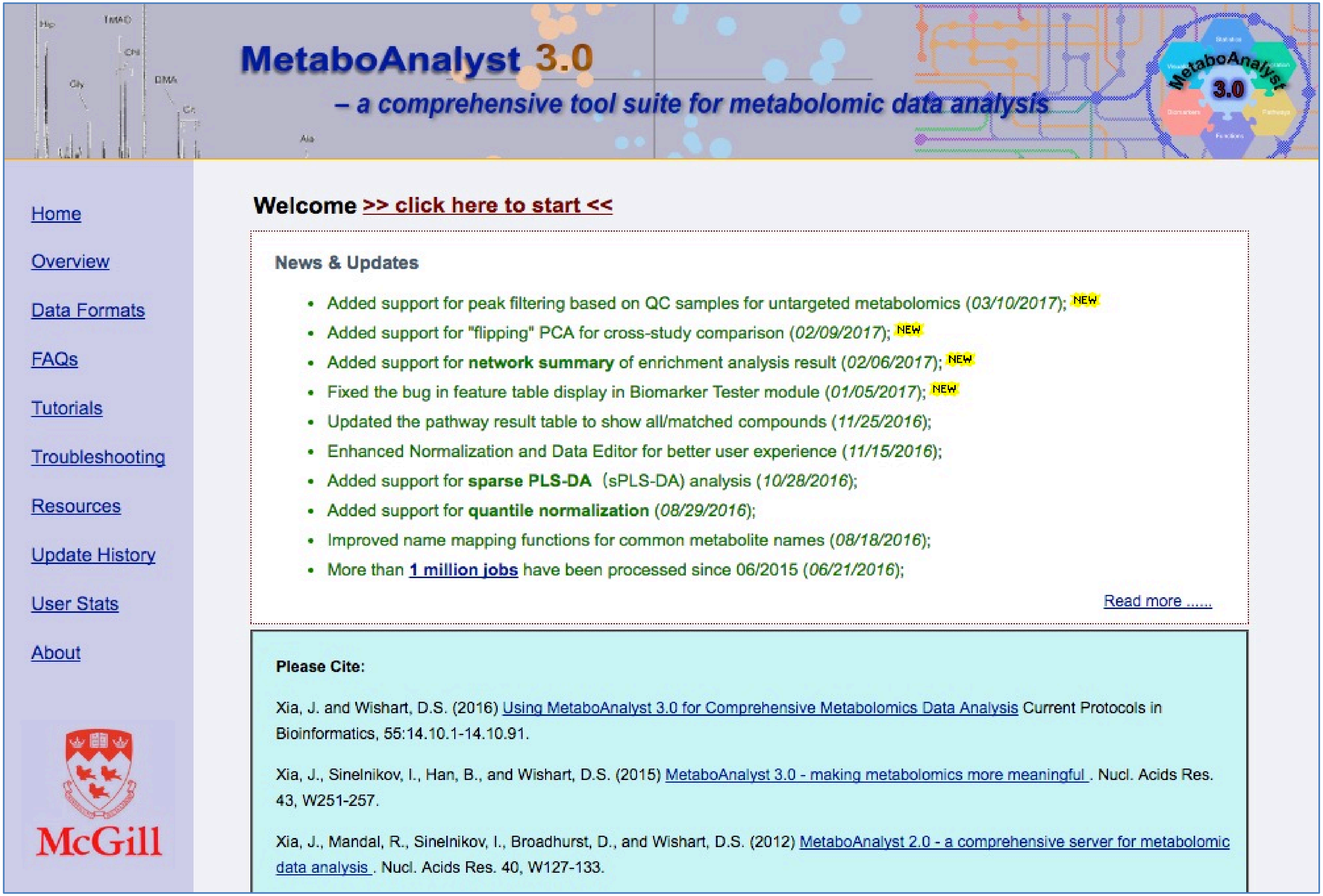
The power of cloud



1.	Beijing
2.	Boston
3.	Seoul
4.	Shanghai
5.	Potsdam
6.	New York
7.	Seattle
8.	Munich
9.	Gainesville

3,000 jobs/month (2014)
150,000 jobs/month (2017)
50 times increase in traffic

Moved to cloud (Oct. 2014)



MetaboAnalyst 3.0

– a comprehensive tool suite for metabolomic data analysis

Welcome >> click here to start <<

News & Updates

- Added support for peak filtering based on QC samples for untargeted metabolomics (*03/10/2017*); NEW
- Added support for "flipping" PCA for cross-study comparison (*02/09/2017*); NEW
- Added support for **network summary** of enrichment analysis result (*02/06/2017*); NEW
- Fixed the bug in feature table display in Biomarker Tester module (*01/05/2017*); NEW
- Updated the pathway result table to show all/matched compounds (*11/25/2016*);
- Enhanced Normalization and Data Editor for better user experience (*11/15/2016*);
- Added support for **sparse PLS-DA** (sPLS-DA) analysis (*10/28/2016*);
- Added support for **quantile normalization** (*08/29/2016*);
- Improved name mapping functions for common metabolite names (*08/18/2016*);
- More than **1 million jobs** have been processed since 06/2015 (*06/21/2016*);

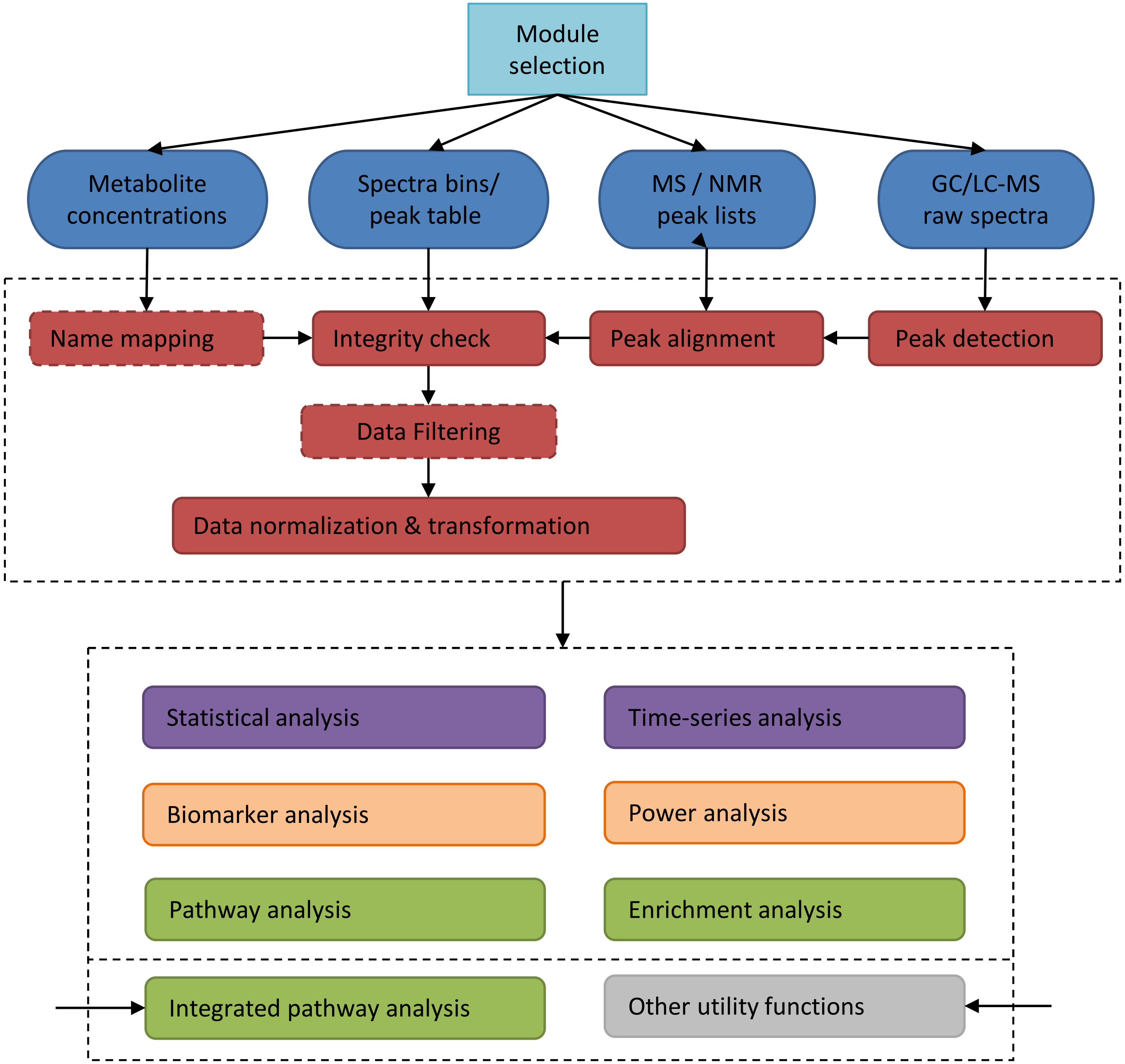
Read more ......

Please Cite:

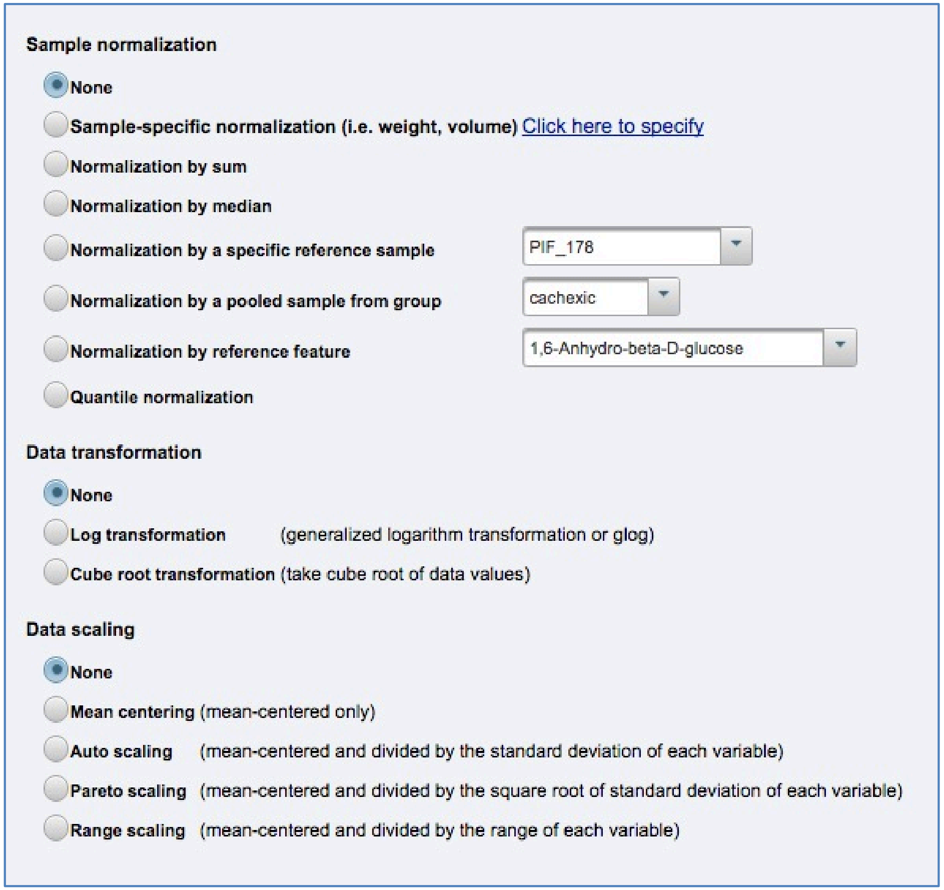
Xia, J. and Wishart, D.S. (2016) Using MetaboAnalyst 3.0 for Comprehensive Metabolomics Data Analysis Current Protocols in Bioinformatics, 55:14.10.1-14.10.91.

Xia, J., Sinelnikov, I., Han, B., and Wishart, D.S. (2015) MetaboAnalyst 3.0 - making metabolomics more meaningful . Nucl. Acids Res. 43, W251-257.

Xia, J., Mandal, R., Sinelnikov, I., Broadhurst, D., and Wishart, D.S. (2012) MetaboAnalyst 2.0 - a comprehensive server for metabolomic data analysis . Nucl. Acids Res. 40, W127-133.


Module selection
Metabolite concentrations
Spectra bins/ peak table
MS / NMR peak lists
GC/LC-MS raw spectra
Name mapping
Integrity check
Peak alignment
Peak detection
Data Filtering
Data normalization & transformation
Statistical analysis
Time-series analysis
Biomarker analysis
Power analysis
Pathway analysis
Enrichment analysis
Integrated pathway analysis
Other utility functions

Comprehensive Options for Data Analysis



Sample normalization

● None

○ Sample-specific normalization (i.e. weight, volume) Click here to specify

○ Normalization by sum

○ Normalization by median

○ Normalization by a specific reference sample PIF_178

○ Normalization by a pooled sample from group cachexic

○ Normalization by reference feature 1,6-Anhydro-beta-D-glucose

○ Quantile normalization

Data transformation

● None

○ Log transformation (generalized logarithm transformation or glog)

○ Cube root transformation (take cube root of data values)

Data scaling

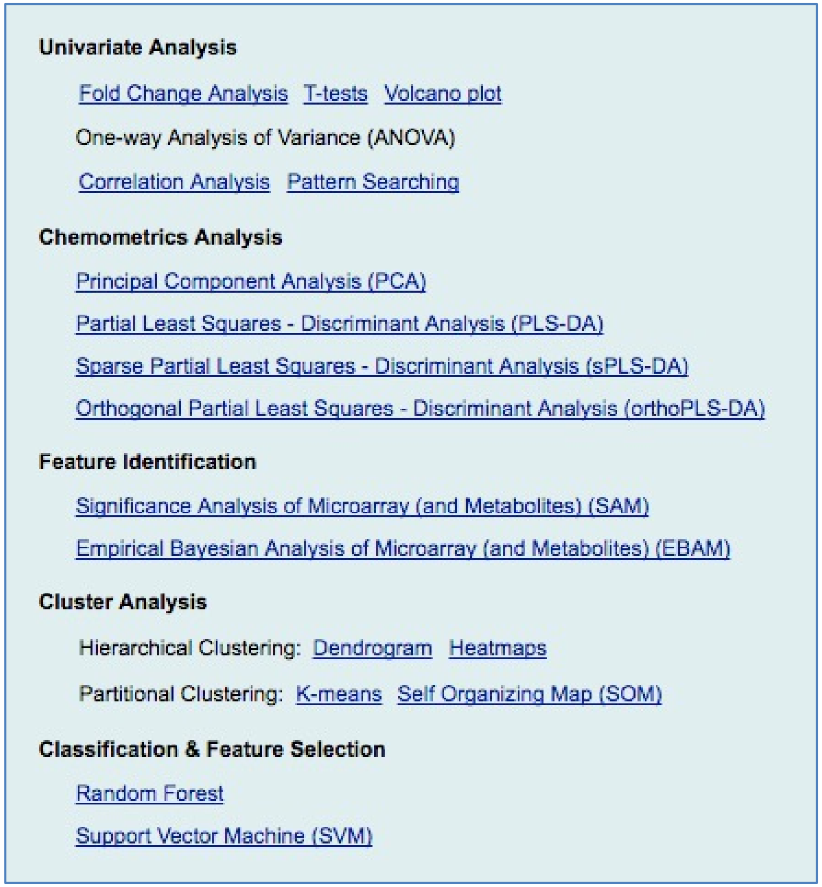
● None

○ Mean centering (mean-centered only)

○ Auto scaling (mean-centered and divided by the standard deviation of each variable)

○ Pareto scaling (mean-centered and divided by the square root of standard deviation of each variable)

○ Range scaling (mean-centered and divided by the range of each variable)



Univariate Analysis

Fold Change Analysis T-tests Volcano plot

One-way Analysis of Variance (ANOVA)

Correlation Analysis Pattern Searching

Chemometrics Analysis

Principal Component Analysis (PCA)

Partial Least Squares - Discriminant Analysis (PLS-DA)

Sparse Partial Least Squares - Discriminant Analysis (sPLS-DA)

Orthogonal Partial Least Squares - Discriminant Analysis (orthoPLS-DA)

Feature Identification

Significance Analysis of Microarray (and Metabolites) (SAM)

Empirical Bayesian Analysis of Microarray (and Metabolites) (EBAM)

Cluster Analysis

Hierarchical Clustering: Dendrogram Heatmaps

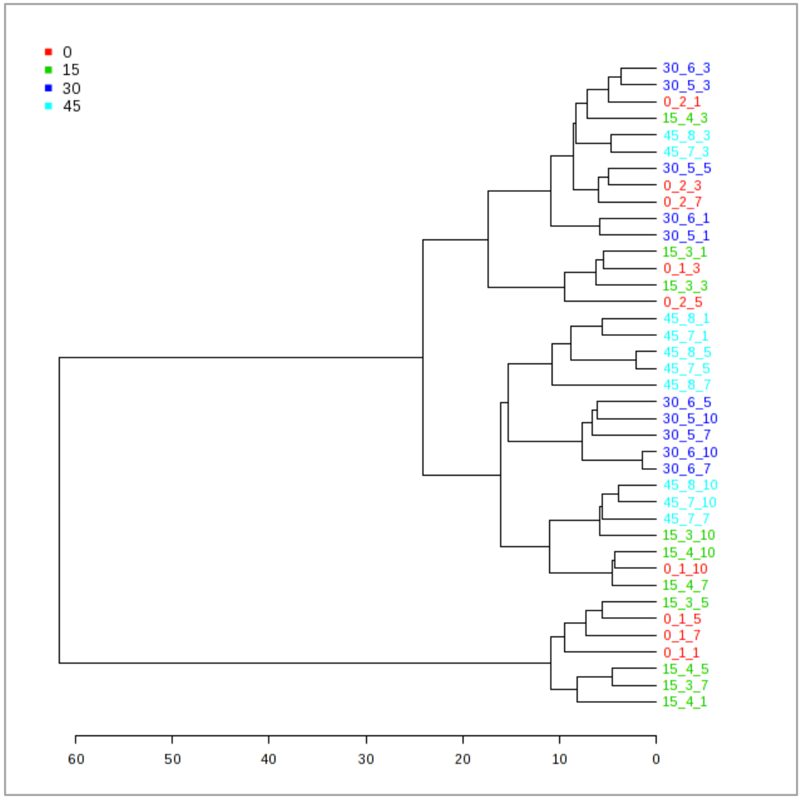
Partitional Clustering: K-means Self Organizing Map (SOM)

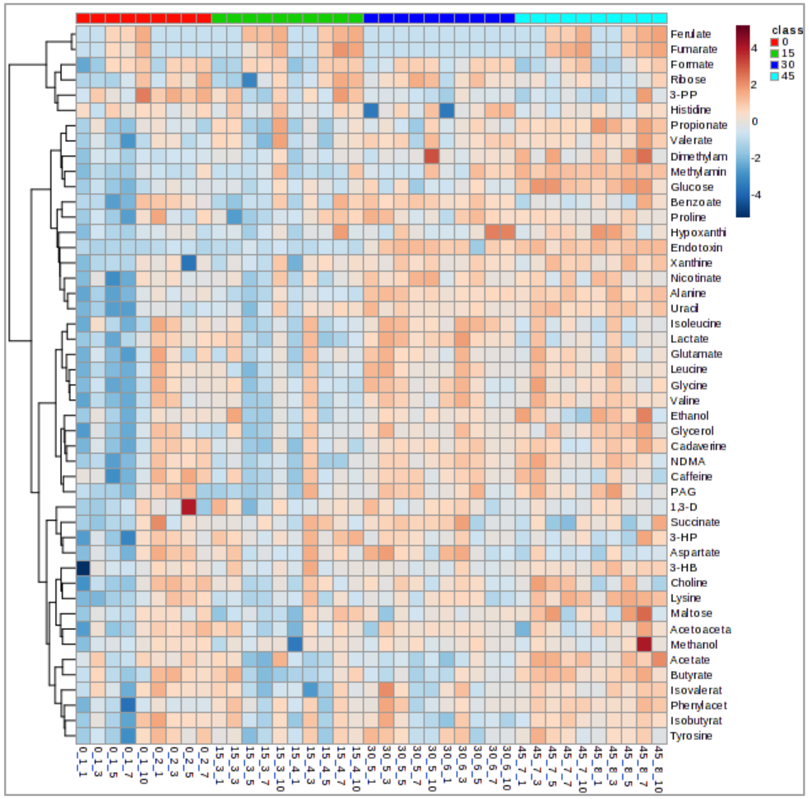
Classification & Feature Selection

Random Forest

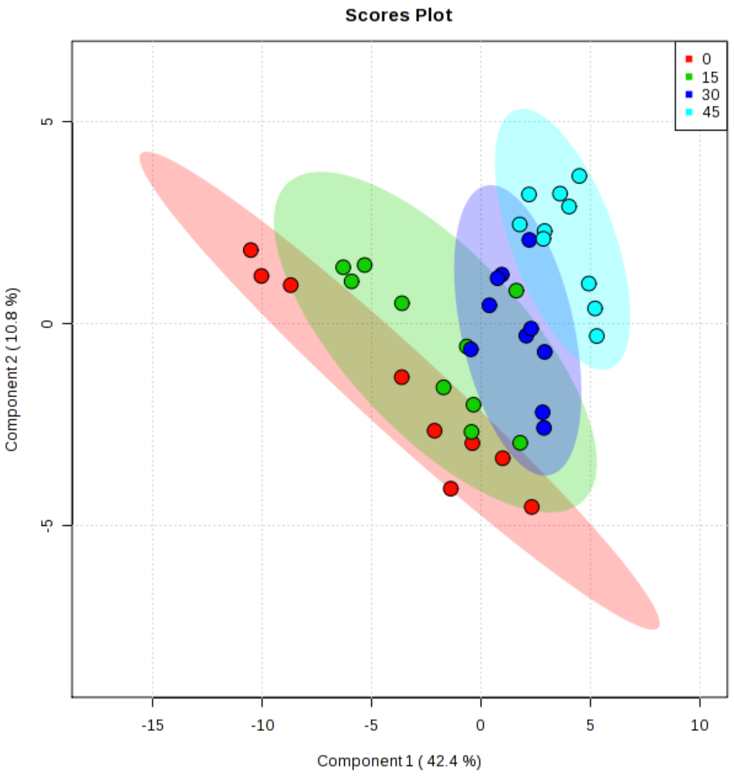
Support Vector Machine (SVM)

Clustering

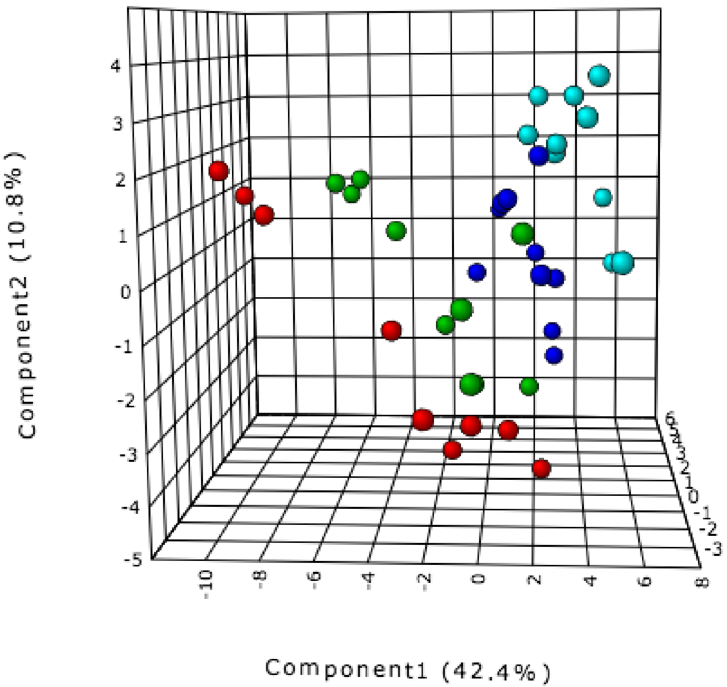

0
15
30
45
30_6_3
30_5_3
0_2_1
15_4_3
45_8_3
45_7_3
30_5_5
0_2_3
0_2_7
30_6_1
30_5_1
15_3_1
0_1_3
15_3_3
0_2_5
45_8_1
45_7_1
45_8_5
45_7_5
45_8_7
30_6_5
30_5_10
30_5_7
30_6_10
30_6_7
45_8_10
45_7_10
45_7_7
15_3_10
15_4_10
0_1_10
15_4_7
15_3_5
0_1_5
0_1_7
0_1_1
15_4_5
15_3_7
15_4_1
60
50
40
30
20
10
0


class
0
15
30
45
4
2
0
-2
-4
Ferulate
Fumarate
Formate
Ribose
3-PP
Histidine
Propionate
Valerate
Dimethylam
Methylamin
Glucose
Benzoate
Proline
Hypoxanthi
Endotoxin
Xanthine
Nicotinate
Alanine
Uracil
Isoleucine
Lactate
Glutamate
Leucine
Glycine
Valine
Ethanol
Glycerol
Cadaverine
NDMA
Caffeine
PAG
1,3-D
Succinate
3-HP
Aspartate
3-HB
Choline
Lysine
Maltose
Acetoaceta
Methanol
Acetate
Butyrate
Isovalerat
Phenylacet
Isobutyrat
Tyrosine
0_1_1
0_1_3
0_1_5
0_1_7
0_1_10
0_2_1
0_2_3
0_2_5
0_2_7
15_3_1
15_3_3
15_3_5
15_3_7
15_3_10
15_4_1
15_4_3
15_4_5
15_4_7
15_4_10
30_5_1
30_5_3
30_5_5
30_5_7
30_5_10
30_6_1
30_6_3
30_6_5
30_6_7
30_6_10
45_7_1
45_7_3
45_7_5
45_7_7
45_7_10
45_8_1
45_8_3
45_8_5
45_8_7
45_8_10

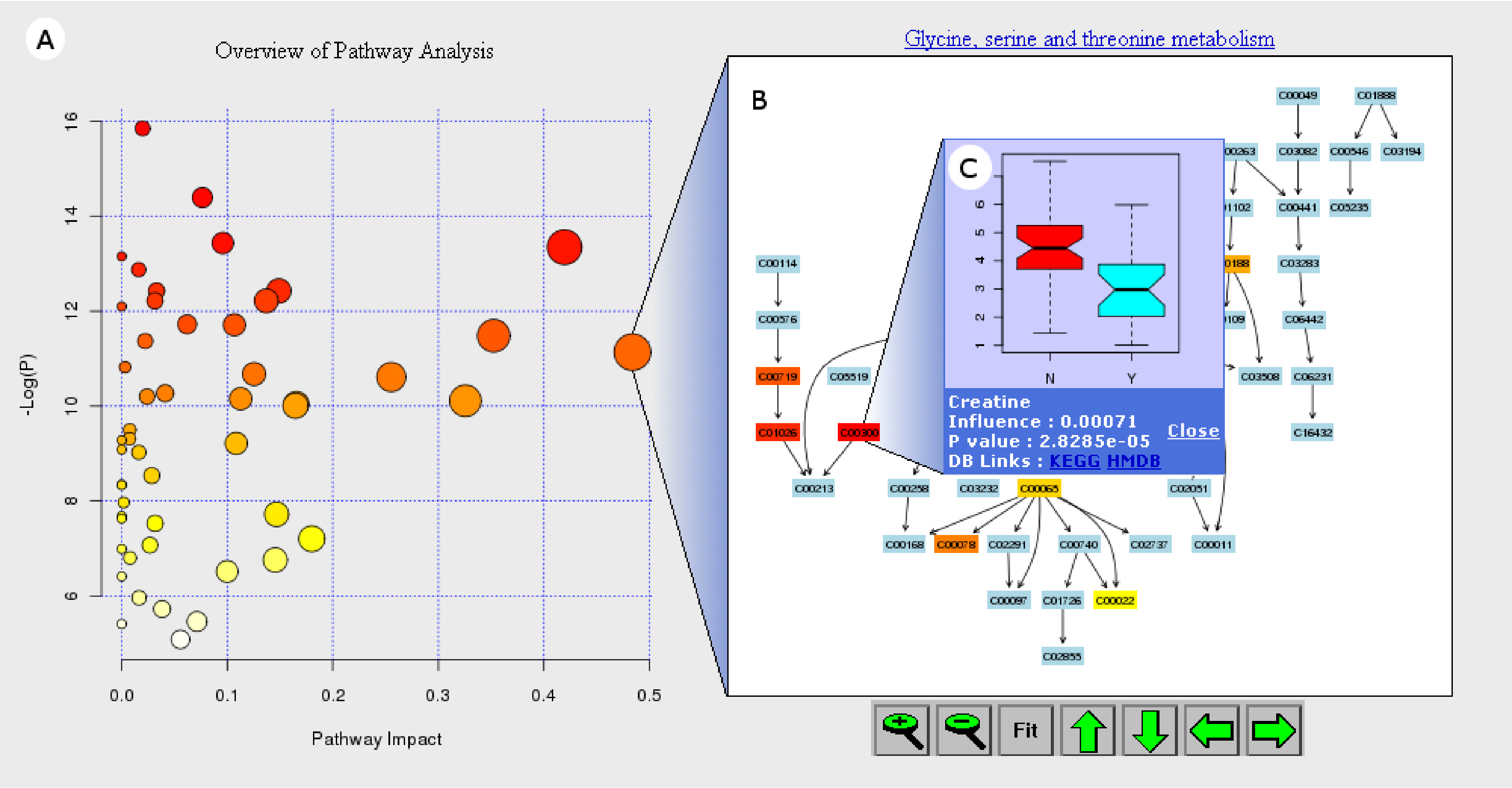
Multivariate Statistics



Scores Plot



Pathway Analysis

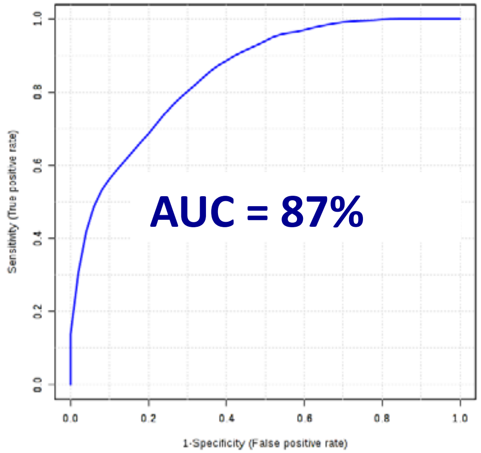


Creatine
Influence : 0.00071
P value : 2.8285e-05
DB Links : KEGG HMDB

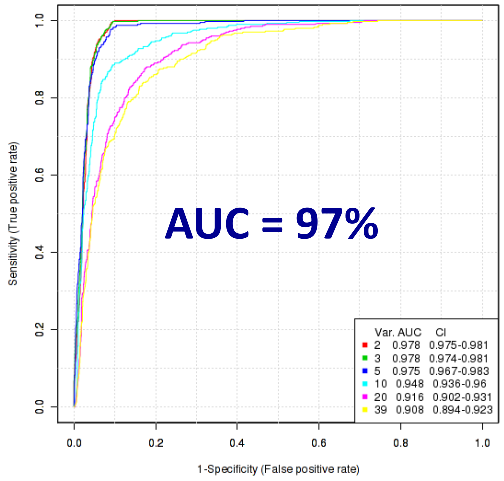
Close

Biomarker Analysis

- Highly relevant for translational studies
- Performance evaluation
  - Receiver operator characteristics (ROC) curve
  - Modern machine learning approaches
    - Cross validation
    - Permutation
    - Predicting new samples

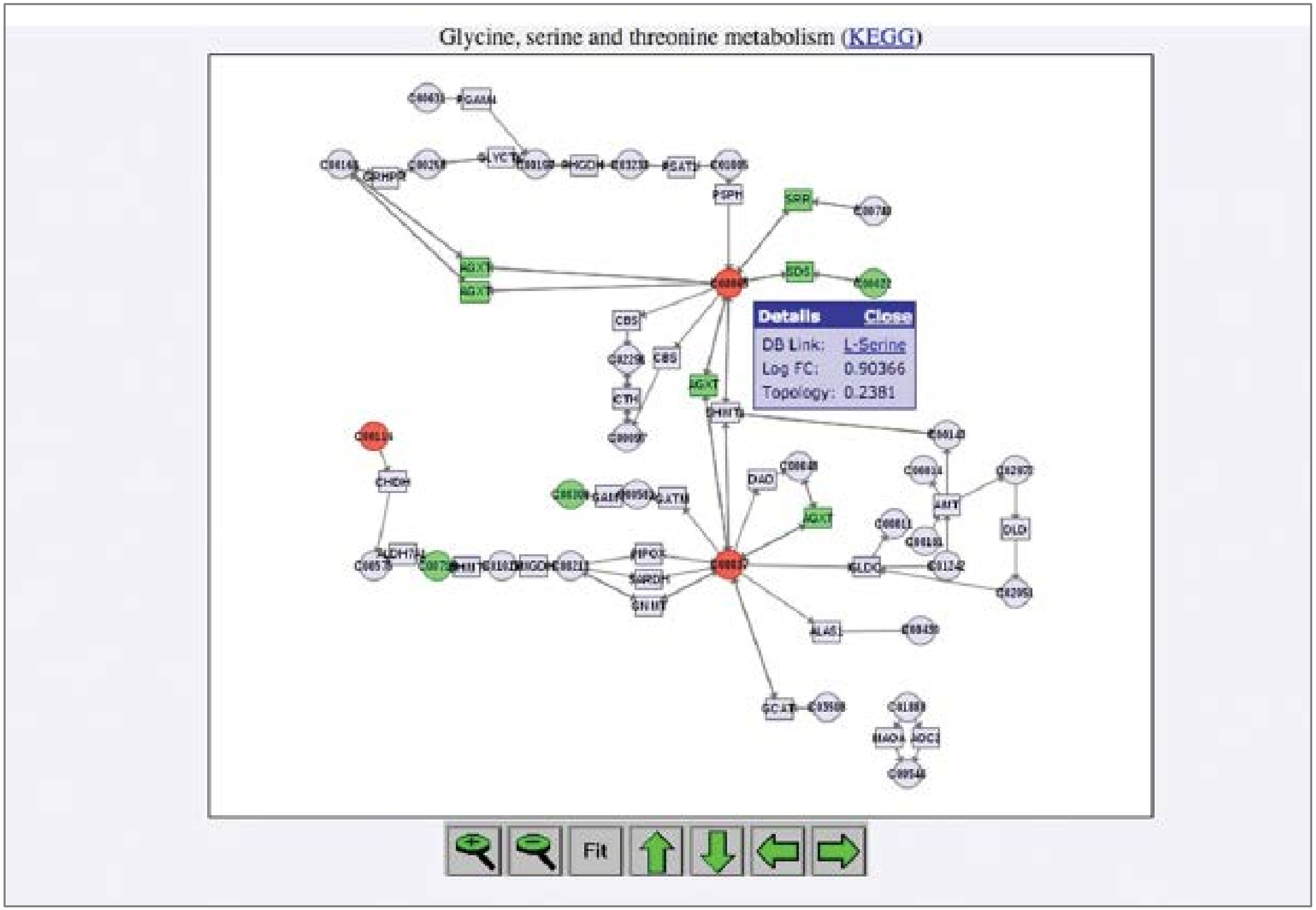


AUC = 87%



AUC = 97%

Gene-metabolite joint pathway analysis



Glycine, serine and threonine metabolism (KEGG)

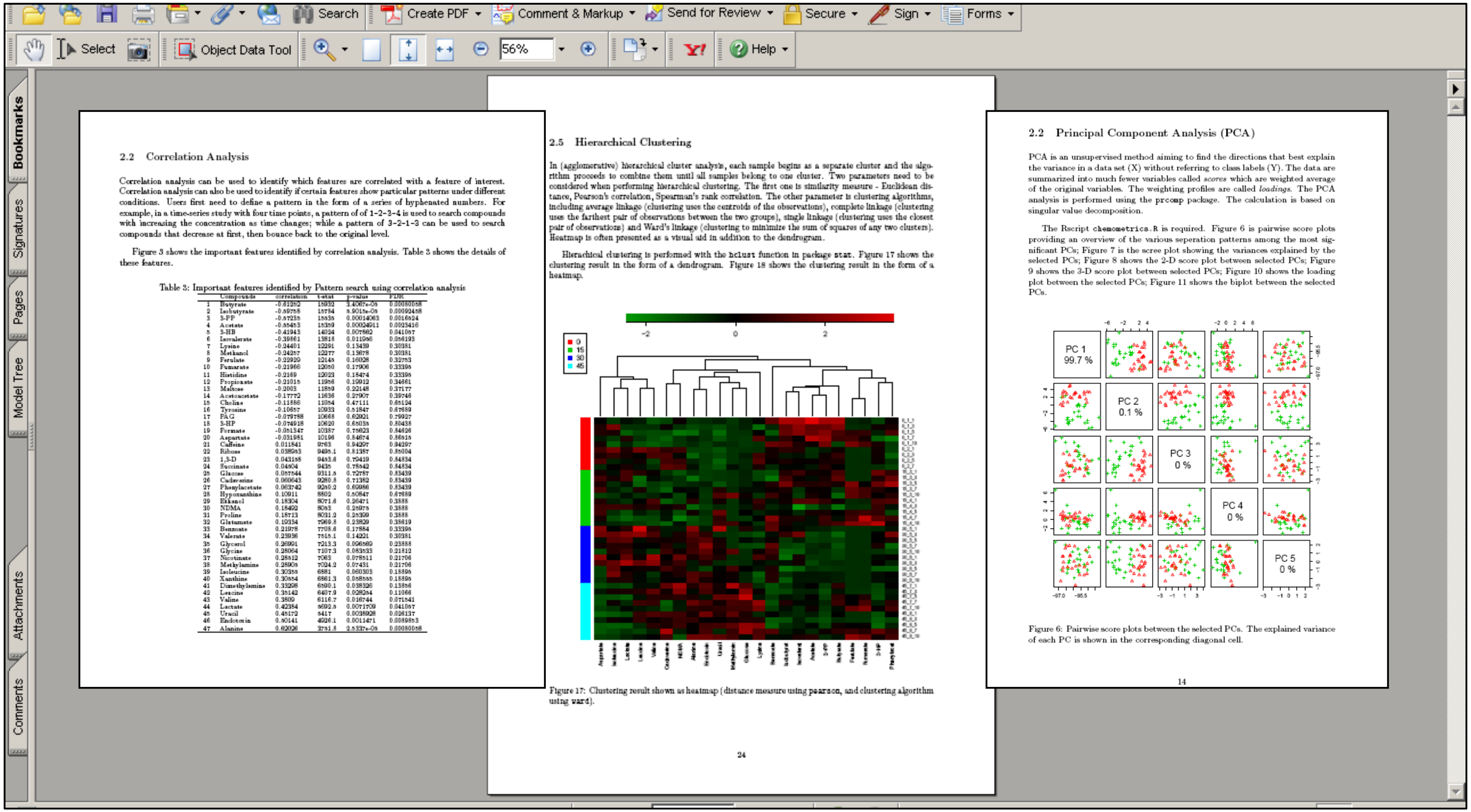
Details Close

DB Link: L-Serine

Log FC: 0.90366

Topology: 0.2381

Comprehensive report generation



2.2 Correlation Analysis

Correlation analysis can be used to identify which features are correlated with a feature of interest. Correlation analysis can also be used to identify if certain features show particular patterns under different conditions. Users first need to define a pattern in the form of a series of hyphenated numbers. For example, in a time-series study with four time points, a pattern of of 1-2-3-4 is used to search compounds with increasing the concentration as time changes; while a pattern of 3-2-1-3 can be used to search compounds that decrease at first, then bounce back to the original level.

Figure 3 shows the important features identified by correlation analysis. Table 3 shows the details of these features.

Table 3: Important features identified by Pattern search using correlation analysis

	Compounds	correlation	t-stat	p-value	FDR
1	Butyrate	-0.81282	18932	3.4067e-08	0.00080088
2	Isobutyrate	-0.89788	18784	8.9018e-08	0.00092438
3	3-PP	-0.87238	18838	0.00014063	0.0016824
4	Acetate	-0.88483	18389	0.00024911	0.0023416
5	3-HB	-0.41943	14024	0.007882	0.041087
6	Isovalerate	-0.39861	13818	0.011986	0.086193
7	Lysine	-0.24401	12291	0.13439	0.30381
8	Methanol	-0.24287	12277	0.13878	0.30381
9	Ferulate	-0.22929	12148	0.16028	0.32783
10	Fumarate	-0.21966	12080	0.17906	0.33398
11	Histidine	-0.2169	12023	0.18474	0.33398
12	Propionate	-0.21018	11988	0.19912	0.34661
13	Maltose	-0.2003	11889	0.22148	0.37177
14	Acetoacetate	-0.17772	11638	0.27907	0.39748
15	Choline	-0.11886	11084	0.47111	0.68124
16	Tyrosine	-0.10887	10933	0.81847	0.67889
17	PAG	-0.079788	10668	0.62921	0.79927
18	3-HP	-0.074918	10620	0.68038	0.80438
19	Formate	-0.081347	10387	0.78823	0.84828
20	Aspartate	-0.031981	10198	0.84674	0.86818
21	Caffeine	0.011841	9763	0.94297	0.94297
22	Ribose	0.038983	9498.1	0.81387	0.88004
23	1,3-D	0.043188	9483.8	0.79419	0.84834
24	Succinate	0.04804	9438	0.78842	0.84834
25	Glucose	0.087844	9311.8	0.72787	0.83439
26	Cadaverine	0.060643	9280.8	0.71382	0.83439
27	Phenylacetate	0.063742	9280.2	0.69986	0.83439
28	Hypoxanthine	0.10911	8802	0.80847	0.67889
29	Ethanol	0.18304	8071.8	0.26471	0.3888
30	NDMA	0.18492	8083	0.28978	0.3888
31	Proline	0.18713	8031.2	0.28399	0.3888
32	Glutamate	0.19334	7969.8	0.23829	0.38819
33	Benzoate	0.21978	7708.8	0.17884	0.33398
34	Valerate	0.23938	7818.1	0.14221	0.30381
35	Glycerol	0.26991	7213.3	0.096869	0.23888
36	Glycine	0.28064	7107.3	0.083833	0.21812
37	Nicotinate	0.28812	7083	0.078811	0.21708
38	Methylamine	0.28908	7024.2	0.07431	0.21708
39	Isoleucine	0.30388	6881	0.060303	0.18898
40	Xanthine	0.30884	6861.3	0.088888	0.18898
41	Dimethylamine	0.33298	6890.1	0.038326	0.13888
42	Leucine	0.38142	6407.9	0.028284	0.11068
43	Valine	0.3809	6116.7	0.016744	0.071841
44	Lactate	0.42384	8692.8	0.0071709	0.041087
45	Uracil	0.48172	8417	0.0038928	0.028137
46	Endotoxin	0.80141	4926.1	0.0011471	0.0089883
47	Alanine	0.82026	3781.8	2.8337e-08	0.00080088

2.5 Hierarchical Clustering

In (agglomerative) hierarchical cluster analysis, each sample begins as a separate cluster and the algorithm proceeds to combine them until all samples belong to one cluster. Two parameters need to be considered when performing hierarchical clustering. The first one is similarity measure - Euclidean distance, Pearson's correlation, Spearman's rank correlation. The other parameter is clustering algorithms, including average linkage (clustering uses the centroids of the observations), complete linkage (clustering uses the farthest pair of observations between the two groups), single linkage (clustering uses the closest pair of observations) and Ward's linkage (clustering to minimize the sum of squares of any two clusters). Heatmap is often presented as a visual aid in addition to the dendrogram.

Hierachical clustering is performed with the `hclust` function in package `stat`. Figure 17 shows the clustering result in the form of a dendrogram. Figure 18 shows the clustering result in the form of a heatmap.

Figure 17: Clustering result shown as heatmap (distance measure using pearson, and clustering algorithm using ward).

24

2.2 Principal Component Analysis (PCA)

PCA is an unsupervised method aiming to find the directions that best explain the variance in a data set (X) without referring to class labels (Y). The data are summarized into much fewer variables called *scores* which are weighted average of the original variables. The weighting profiles are called *loadings*. The PCA analysis is performed using the `prcomp` package. The calculation is based on singular value decomposition.

The Rscript `chemometrics.R` is required. Figure 6 is pairwise score plots providing an overview of the various seperation patterns among the most significant PCs; Figure 7 is the scree plot showing the variances explained by the selected PCs; Figure 8 shows the 2-D score plot between selected PCs; Figure 9 shows the 3-D score plot between selected PCs; Figure 10 shows the loading plot between the selected PCs; Figure 11 shows the biplot between the selected PCs.

Figure 6: Pairwise score plots between the selected PCs. The explained variance of each PC is shown in the corresponding diagonal cell.

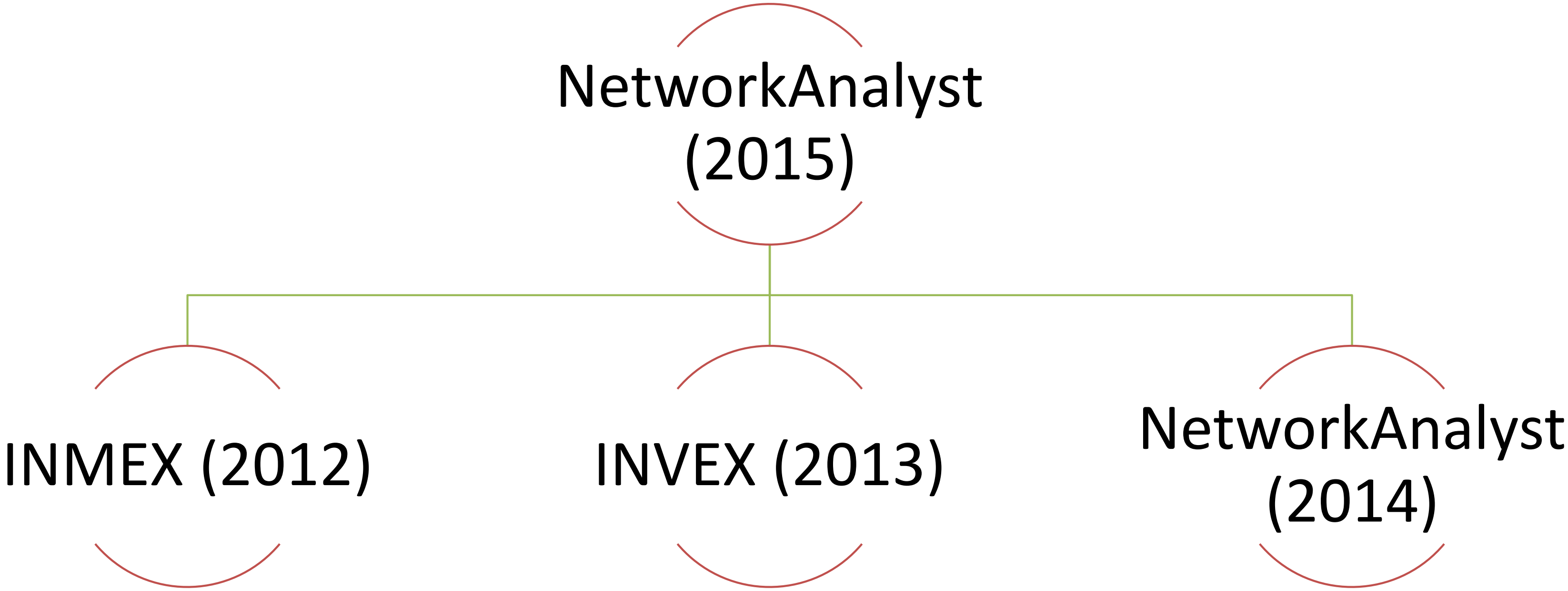
14

2012 - CURRENT

Transcriptomics & NetworkAnalyst
http://www.networkanalyst.ca

- Gene expression analysis (microarray & RNAseq)
- Statistical data integration (meta-analysis)
- Visual integration (heatmaps, Venn diagrams, 3D PCA/tSNE)
- Network integration (PPI, KEGG, miRanda)

Implementation Roadmap



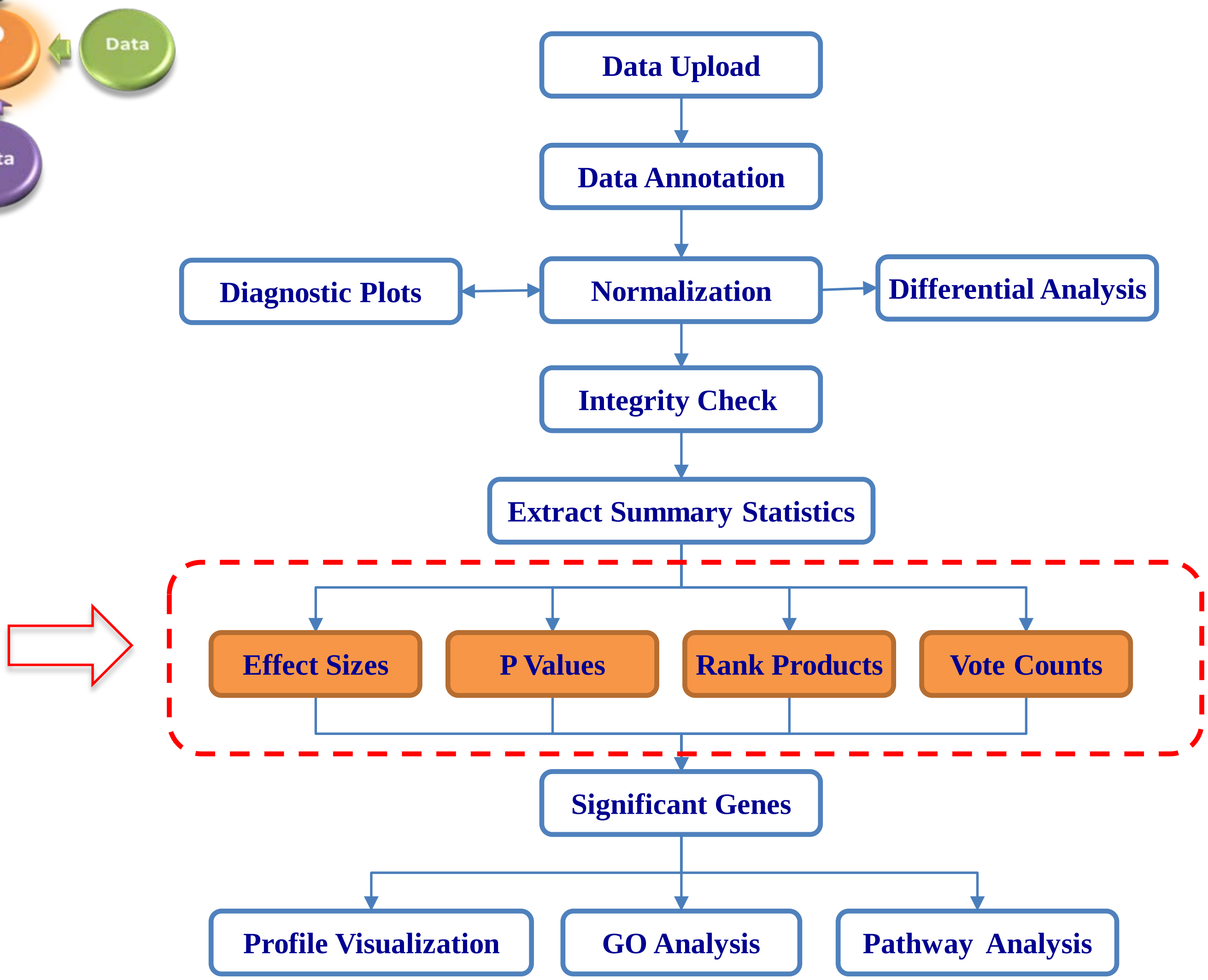
NetworkAnalyst (2015)

INMEX (2012)

INVEX (2013)

NetworkAnalyst (2014)

Statistical Integration



Data Upload

Data Annotation

Diagnostic Plots

Normalization

Differential Analysis

Integrity Check

Extract Summary Statistics

Effect Sizes

P Values

Rank Products

Vote Counts

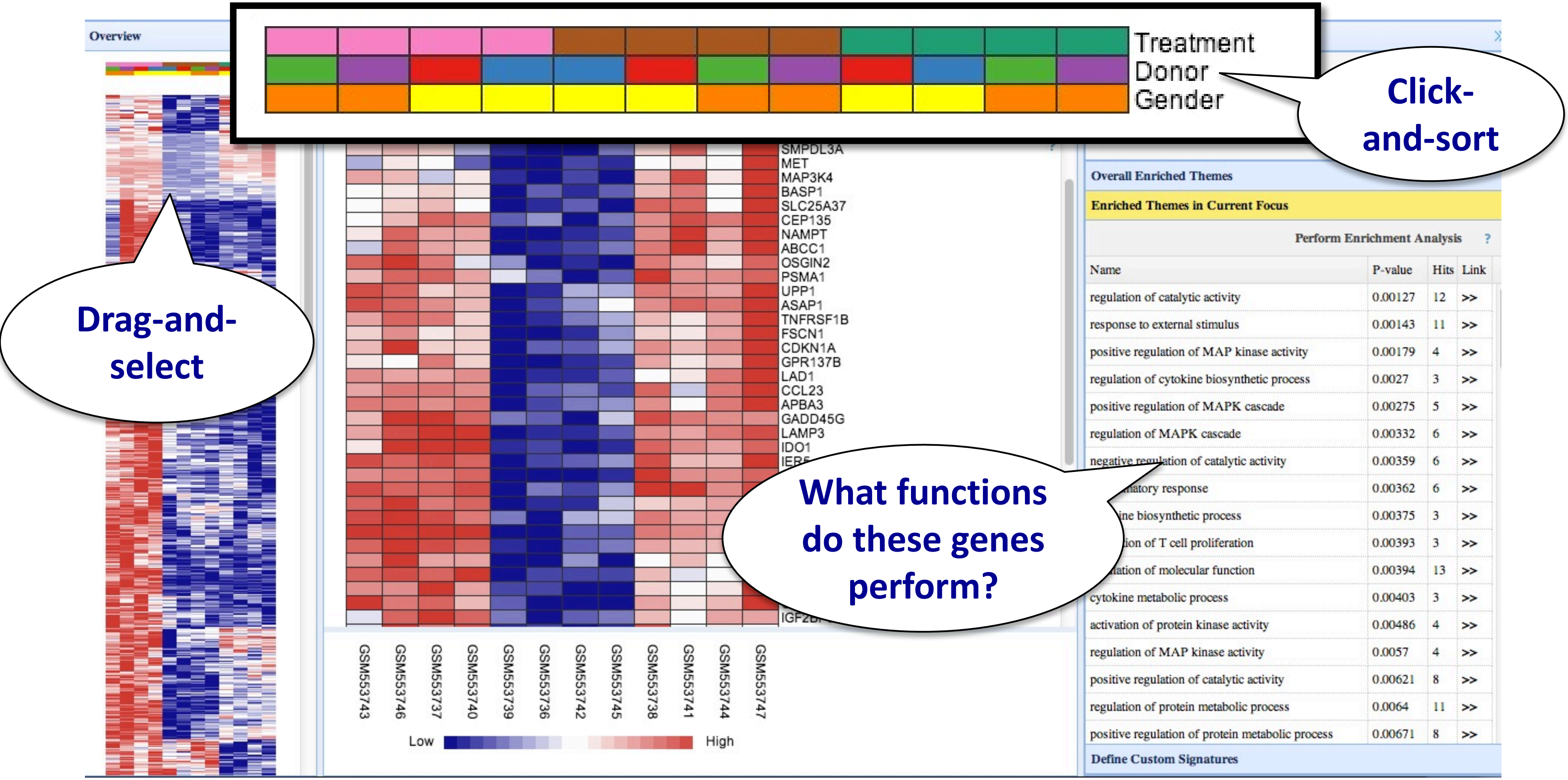
Significant Genes

Profile Visualization

GO Analysis

Pathway Analysis

Visual Integration



Click-and-sort

Drag-and-select

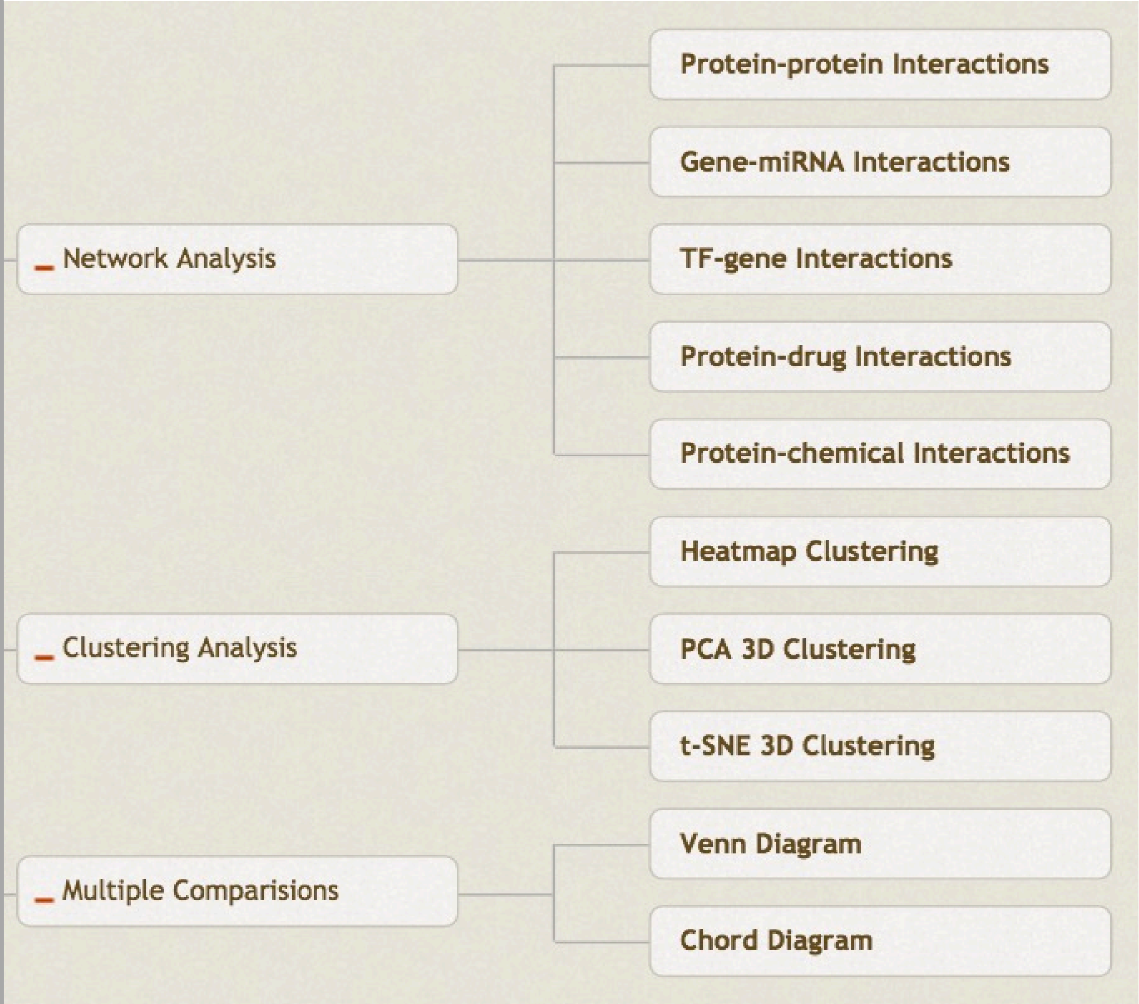
What functions do these genes perform?

Name	P-value	Hits	Link
regulation of catalytic activity	0.00127	12	>>
response to external stimulus	0.00143	11	>>
positive regulation of MAP kinase activity	0.00179	4	>>
regulation of cytokine biosynthetic process	0.0027	3	>>
positive regulation of MAPK cascade	0.00275	5	>>
regulation of MAPK cascade	0.00332	6	>>
negative regulation of catalytic activity	0.00359	6	>>
matory response	0.00362	6	>>
ine biosynthetic process	0.00375	3	>>
ion of T cell proliferation	0.00393	3	>>
ation of molecular function	0.00394	13	>>
cytokine metabolic process	0.00403	3	>>
activation of protein kinase activity	0.00486	4	>>
regulation of MAP kinase activity	0.0057	4	>>
positive regulation of catalytic activity	0.00621	8	>>
regulation of protein metabolic process	0.0064	11	>>
positive regulation of protein metabolic process	0.00671	8	>>

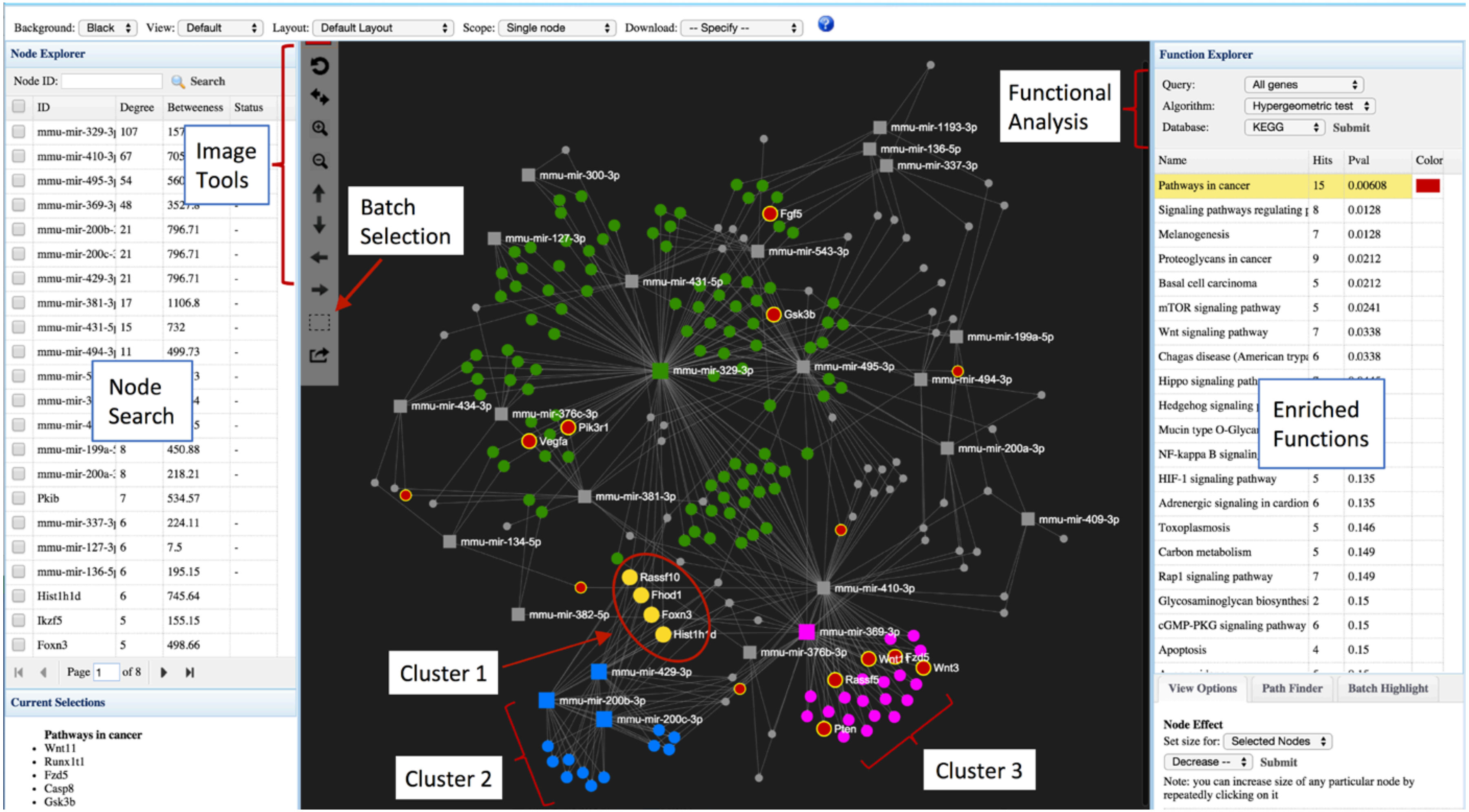
J. Xia. et al. (2013) Bioinformatics 29 (24), 3232-3234

Integrating with prior knowledge

- Protein-protein interactions
  - ➢ STRING
  - ➢ InnateDB
- Metabolic pathways
  - ➢ KEGG
  - ➢ Reactome
- Chemicals
  - ➢ DrugBank
  - ➢ CTD
- Gene regulations
  - ➢ Trans Factor
  - ➢ miRNAs



Network Visualization & Exploration



Background: Black View: Default Layout: Default Layout Scope: Single node Download: -- Specify --

Node Explorer

Node ID: Search

	ID	Degree	Betweeness	Status
□	mmu-mir-329-3	107	157[illegible]	
□	mmu-mir-410-3	67	705[illegible]	
□	mmu-mir-495-3	54	560[illegible]	
□	mmu-mir-369-3	48	352[illegible]	-
□	mmu-mir-200b-	21	796.71	-
□	mmu-mir-200c-	21	796.71	-
□	mmu-mir-429-3	21	796.71	-
□	mmu-mir-381-3	17	1106.8	-
□	mmu-mir-431-5	15	732	-
□	mmu-mir-494-3	11	499.73	-
□	mmu-mir-5	[illegible]	[illegible]3	-
□	mmu-mir-3	[illegible]	[illegible]4	-
□	mmu-mir-4	[illegible]	[illegible]5	-
□	mmu-mir-199a-	8	450.88	-
□	mmu-mir-200a-	8	218.21	-
□	Pkib	7	534.57	
□	mmu-mir-337-3	6	224.11	-
□	mmu-mir-127-3	6	7.5	-
□	mmu-mir-136-5	6	195.15	-
□	Hist1h1d	6	745.64	
□	Ikzf5	5	155.15	
□	Foxn3	5	498.66	

Page 1 of 8

Current Selections

Pathways in cancer

- Wnt11
- Runx1t1
- Fzd5
- Casp8
- Gsk3b

Image Tools

Node Search

Batch Selection

Functional Analysis

Cluster 1

Cluster 2

Cluster 3

Enriched Functions

Function Explorer

Query: All genes

Algorithm: Hypergeometric test

Database: KEGG Submit

Name	Hits	Pval	Color
Pathways in cancer	15	0.00608	🟥
Signaling pathways regulating	8	0.0128	
Melanogenesis	7	0.0128	
Proteoglycans in cancer	9	0.0212	
Basal cell carcinoma	5	0.0212	
mTOR signaling pathway	5	0.0241	
Wnt signaling pathway	7	0.0338	
Chagas disease (American trypa	6	0.0338	
Hippo signaling path	[illegible]	[illegible]	
Hedgehog signaling	[illegible]	[illegible]	
Mucin type O-Glycar	[illegible]	[illegible]	
NF-kappa B signalin	[illegible]	[illegible]	
HIF-1 signaling pathway	5	0.135	
Adrenergic signaling in cardion	6	0.135	
Toxoplasmosis	5	0.146	
Carbon metabolism	5	0.149	
Rap1 signaling pathway	7	0.149	
Glycosaminoglycan biosynthesi	2	0.15	
cGMP-PKG signaling pathway	6	0.15	
Apoptosis	4	0.15	

View Options | Path Finder | Batch Highlight

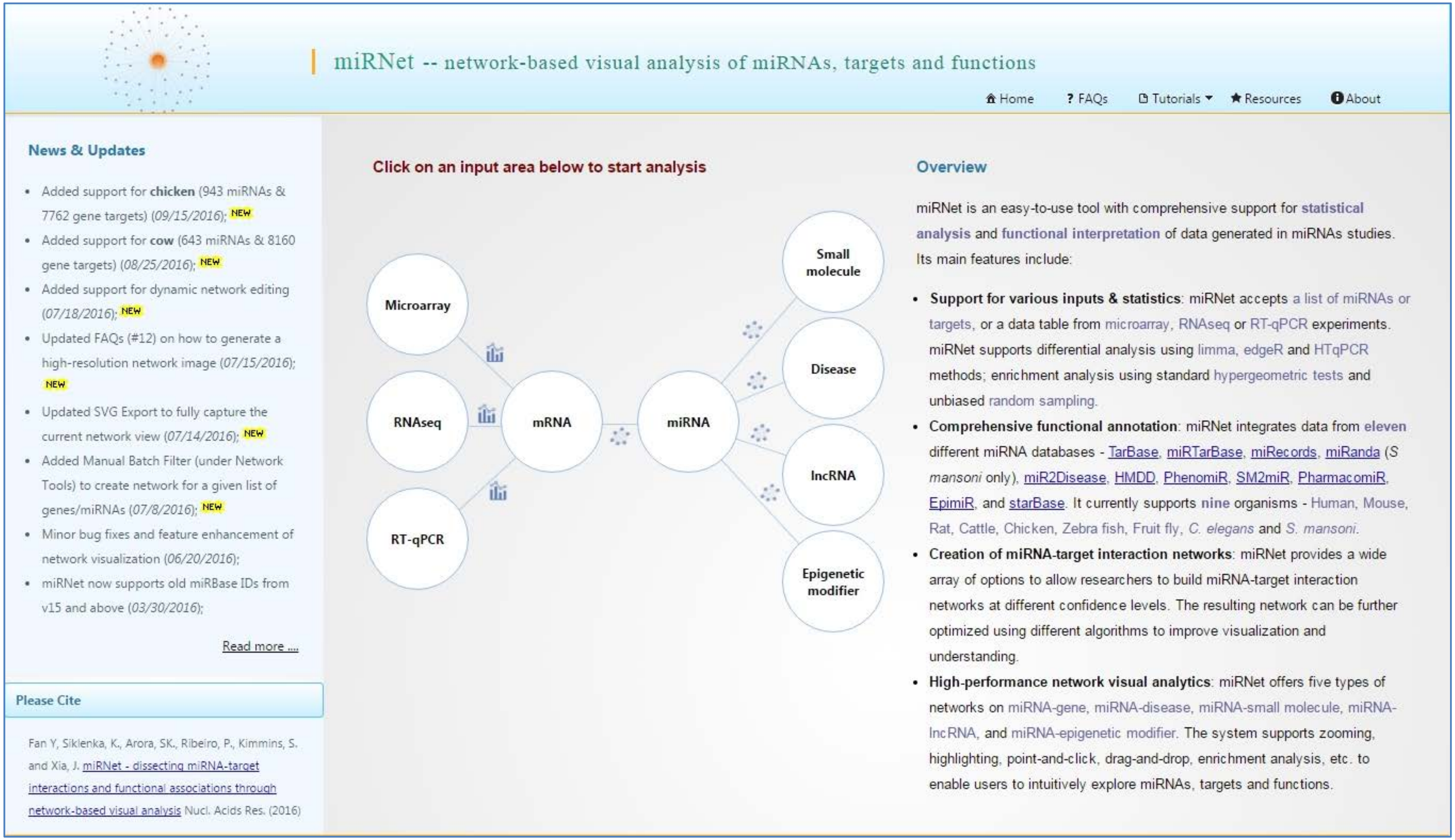
Node Effect

Set size for: Selected Nodes

Decrease -- Submit

Note: you can increase size of any particular node by repeatedly clicking on it

MicroRNA Data (www.mirnet.ca)



miRNet -- network-based visual analysis of miRNAs, targets and functions

News & Updates

- Added support for **chicken** (943 miRNAs & 7762 gene targets) *(09/15/2016)*; NEW
- Added support for **cow** (643 miRNAs & 8160 gene targets) *(08/25/2016)*; NEW
- Added support for dynamic network editing *(07/18/2016)*; NEW
- Updated FAQs (#12) on how to generate a high-resolution network image *(07/15/2016)*; NEW
- Updated SVG Export to fully capture the current network view *(07/14/2016)*; NEW
- Added Manual Batch Filter (under Network Tools) to create network for a given list of genes/miRNAs *(07/8/2016)*; NEW
- Minor bug fixes and feature enhancement of network visualization *(06/20/2016)*;
- miRNet now supports old miRBase IDs from v15 and above *(03/30/2016)*;

Read more ....

Please Cite

Fan Y, Siklenka, K., Arora, SK., Ribeiro, P., Kimmins, S. and Xia, J. miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis Nucl. Acids Res. (2016)

Click on an input area below to start analysis

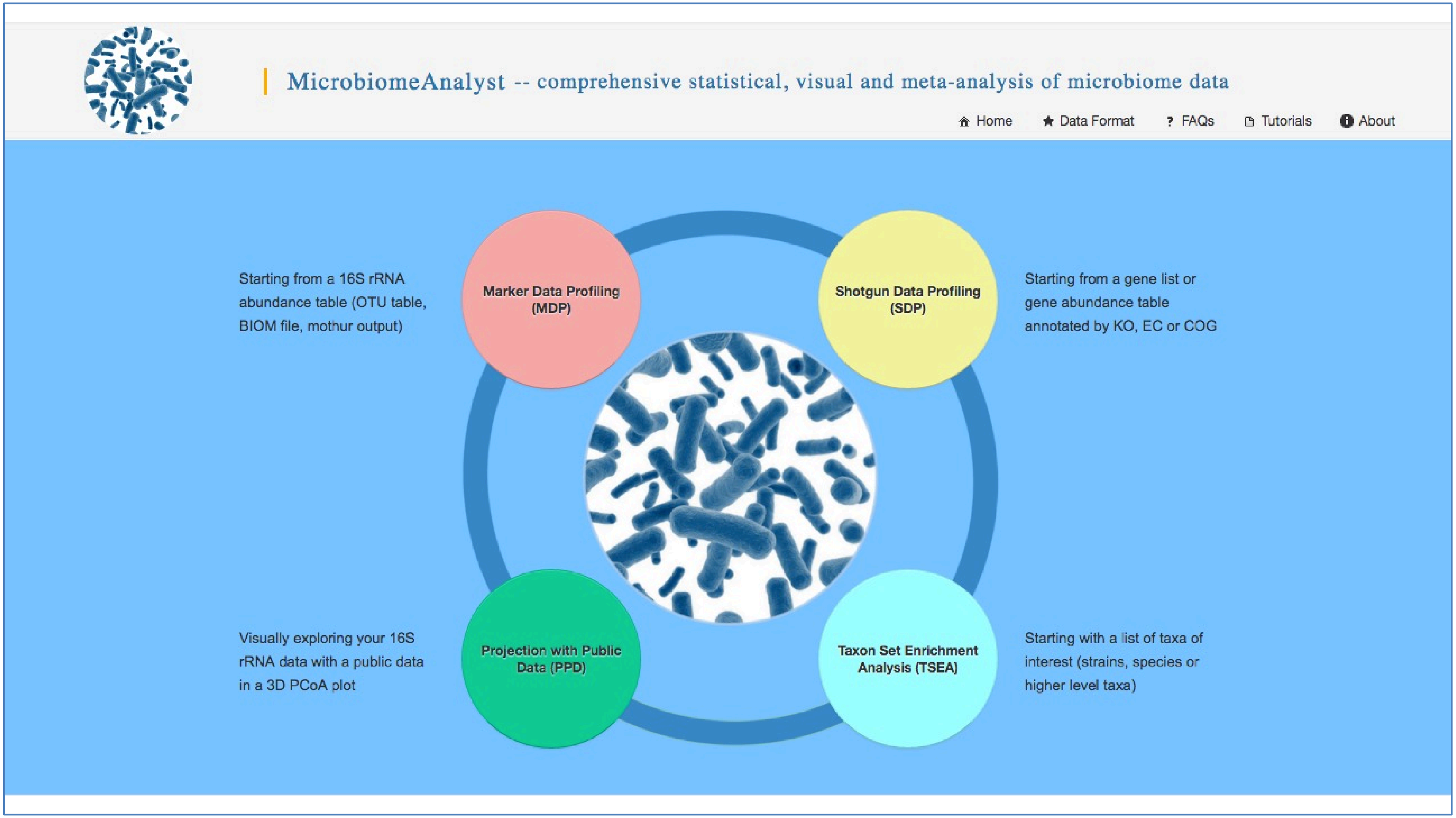
Overview

miRNet is an easy-to-use tool with comprehensive support for **statistical analysis** and **functional interpretation** of data generated in miRNAs studies. Its main features include:

- **Support for various inputs & statistics**: miRNet accepts a list of miRNAs or targets, or a data table from microarray, RNAseq or RT-qPCR experiments. miRNet supports differential analysis using limma, edgeR and HTqPCR methods; enrichment analysis using standard hypergeometric tests and unbiased random sampling.
- **Comprehensive functional annotation**: miRNet integrates data from **eleven** different miRNA databases - TarBase, miRTarBase, miRecords, miRanda (*S mansoni* only), miR2Disease, HMDD, PhenomiR, SM2miR, PharmacomiR, EpimiR, and starBase. It currently supports **nine** organisms - Human, Mouse, Rat, Cattle, Chicken, Zebra fish, Fruit fly, *C. elegans* and *S. mansoni*.
- **Creation of miRNA-target interaction networks**: miRNet provides a wide array of options to allow researchers to build miRNA-target interaction networks at different confidence levels. The resulting network can be further optimized using different algorithms to improve visualization and understanding.
- **High-performance network visual analytics**: miRNet offers five types of networks on miRNA-gene, miRNA-disease, miRNA-small molecule, miRNA-lncRNA, and miRNA-epigenetic modifier. The system supports zooming, highlighting, point-and-click, drag-and-drop, enrichment analysis, etc. to enable users to intuitively explore miRNAs, targets and functions.

Fan, Y et al. (2016) (doi: 10.1093/nar/gkw288)

Microbiome Data (www.microbiomeanalyst.ca)



MicrobiomeAnalyst -- comprehensive statistical, visual and meta-analysis of microbiome data

Home
Data Format
FAQs
Tutorials
About
Marker Data Profiling (MDP)
Shotgun Data Profiling (SDP)
Projection with Public Data (PPD)
Taxon Set Enrichment Analysis (TSEA)

Starting from a 16S rRNA abundance table (OTU table, BIOM file, mothur output)

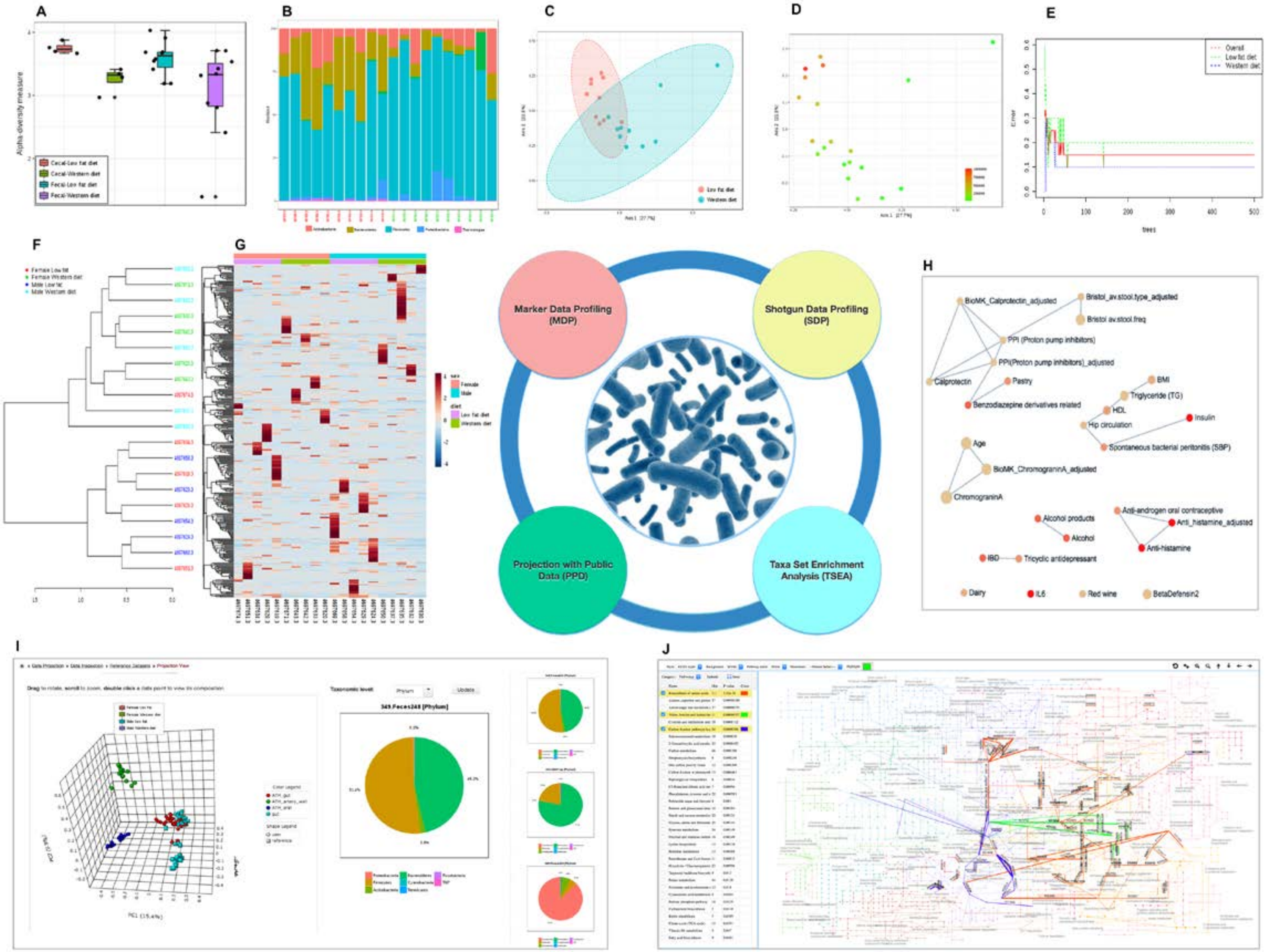
Starting from a gene list or gene abundance table annotated by KO, EC or COG

Visually exploring your 16S rRNA data with a public data in a 3D PCoA plot

Starting with a list of taxa of interest (strains, species or higher level taxa)

Dhariwal, A et al. (2017) (under review)

Microbiome profiling & integration



In the near future ……

Acknowledgements

- Xia Lab @ McGill University
- Wishart Lab @ Univ. Alberta
- Hancock Lab @ Univ. of British Columbia

McGill UNIVERSITY

NSERC CRSNG

Fonds de recherche
sur la nature
et les technologies
Québec

GenomeCanada