# PATHWAY ANALYSIS FOR METABOLOMICS

Alex Sanchez

# INTRODUCTION AND OBJECTIVES

# INTRODUCING OURSELVES


Statistics & Bioinformatics and Nutrition & Metabolomics groups @ UB

# SESSION OBJECTIVES

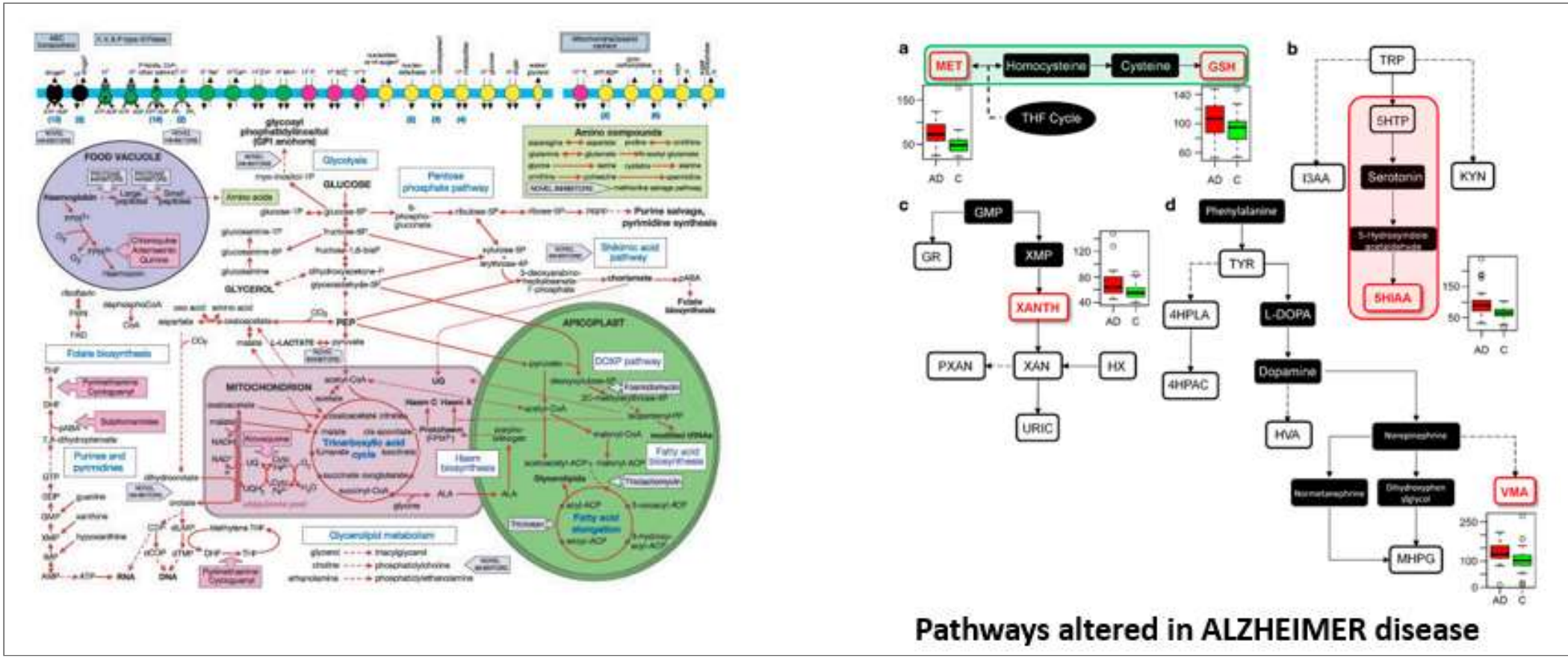
- Overview of Patwhay Analysis for Metabolomics
- Introduce its components and
- Go through some methods with some detail
- Discuss some limitations and provide recomendations.
- Introduce some tools for Pathway Analysis
- Get a practical grasp of how to apply it.

# SESSION OUTLINE

1. Introducion and objectives
2. Metabolite lists: What do they mean
3. Information sources to support interpretation
4. Methods and Tools to extract information
5. The limitations of PwA. Some recommendations
6. Software tools for PwA
7. Practical session

# HEALTH, DISEASE AND PATHWAYS

- Metabolism is a complex network of chemical reactions within the confines of a cell that can be analyzed in self-contained parts called pathways.
- We often assume that “normal” metabolism is what happens in healthy state or, that disease can be associated with some type of alteration in metabolism.



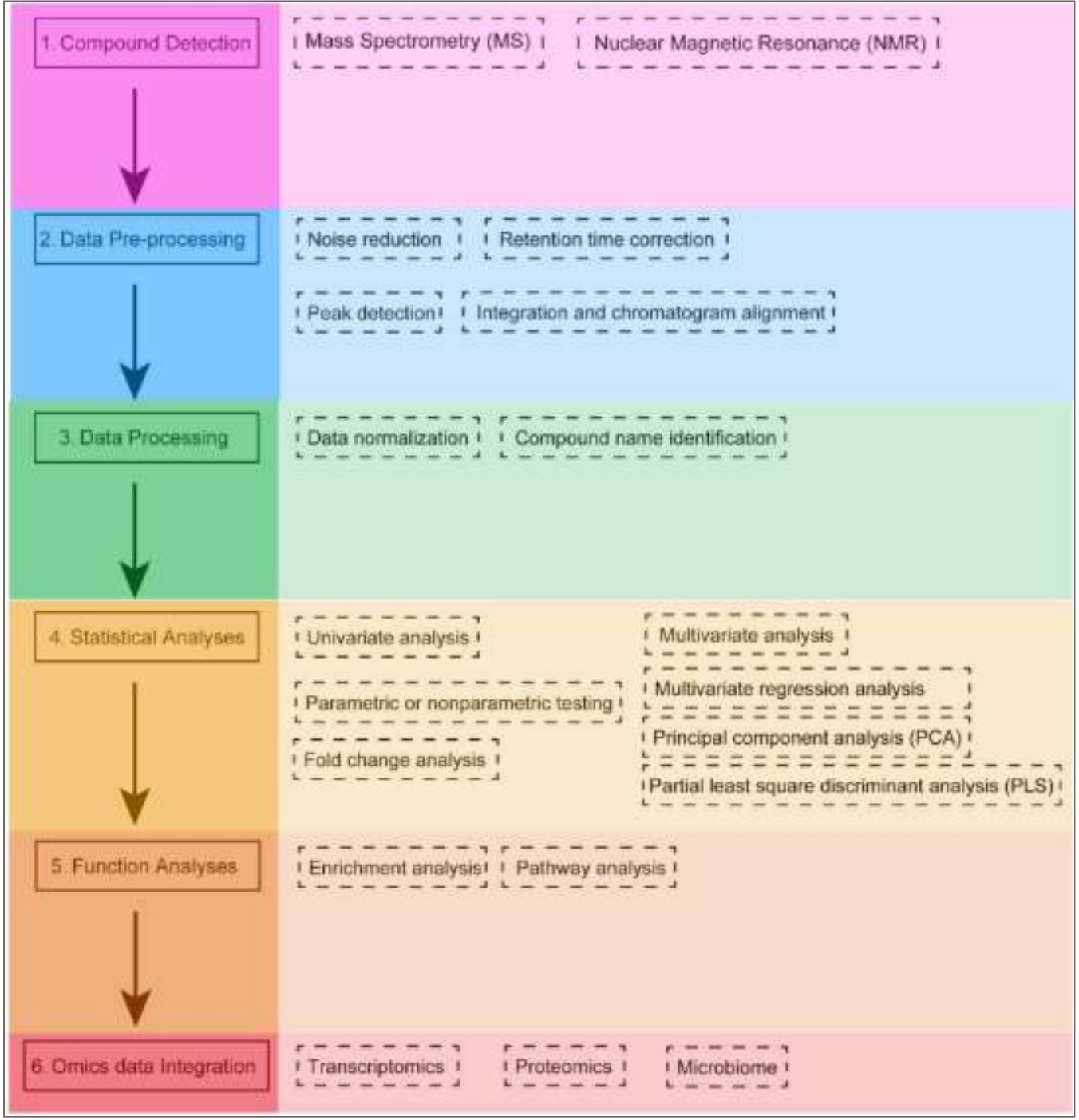
Characterization of disease attempted studying how this disrupts pathways

# SO WHAT IS PATHWAY ANALYSIS?

- … any analytic technique that benefits from biological pathway or molecular network information to gain insight into a biological system. (Creixell et alt., Nature Methods 2015 (12 (7))
- Pathway Analysis methods rely on high throughput information provided by omics technologies to:
  - Contextualize findings to help understand biological processes
  - Identify fetures associated with a disease
  - Predict drug targets
  - Understand how to intervene in disease1
  - Conduct target literature searches
  - Integrate diverse biological information

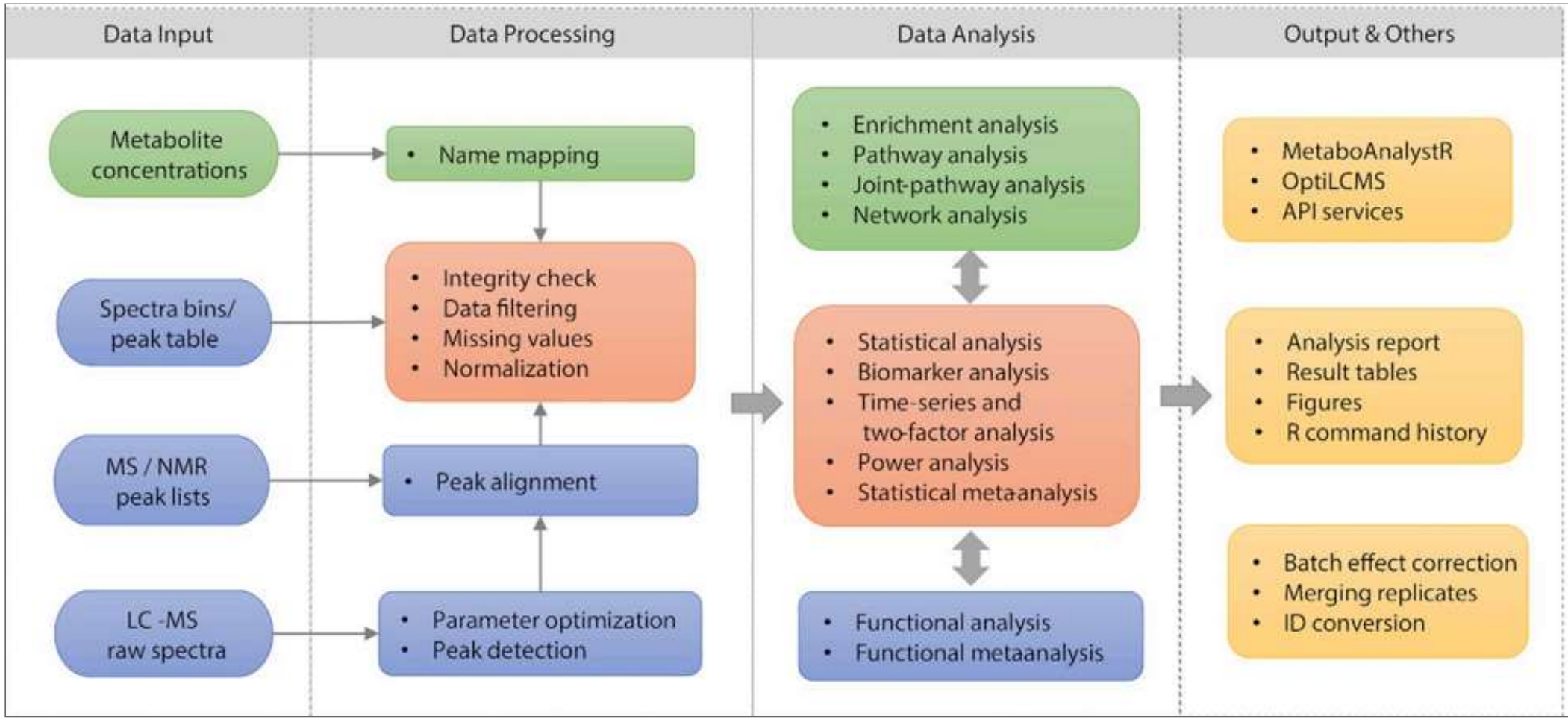
# FROM SAMPLES TO *FEATURES* LISTS

# BIOINFORMATICS WORKFLOWS


1. Compound Detection
Mass Spectrometry (MS)
Nuclear Magnetic Resonance (NMR)
2. Data Pre-processing
Noise reduction
Retention time correction
Peak detection
Integration and chromatogram alignment
3. Data Processing
Data normalization
Compound name identification
4. Statistical Analyses
Univariate analysis
Multivariate analysis
Parametric or nonparametric testing
Multivariate regression analysis
Fold change analysis
Principal component analysis (PCA)
Partial least square discriminant analysis (PLS)
5. Function Analyses
Enrichment analysis
Pathway analysis
6. Omics data Integration
Transcriptomics
Proteomics
Microbiome

# A Metabolomics Worflow Example

# FROM SAMPLES TO *FEATURES* LISTS (2)




Metabolomics Workflows in MetaboAnalyst 5.0

# ANALYSIS YIELD METABOLITES LISTS

| Metabolite |
|---|
| Amino acid |
| 5-oxoproline (pyroglutamic acid) |
| 7-Methylguanine |
| Creatinine |
| Histidine |
| Kynurenic acid |
| L-Tryptophan |
| N-(2-Furoyl)glycine |
| N-Acetylneuraminic acid |
| Spermidine |
| Organic compounds |
| (±)-Sulfobutanedioic acid |
| D-Tagatose |
| D-Xylulose |
| Glutaminyl-Gamma-Glutamate |
| L-Galacto-2-heptulose |
| N-Acetylgalactosamine 6-sulfate |
| Phenol sulphate |
| Trigonellinamide |
| Tyrosine |
| Salicyluric acid |
| Carbohydrates |
| Gluconic acid |
| Sorbitol |
| Xenobiotics |

An unordered list of metabolite IDs

| Metabolite | Fold change | *p*-value | AUC | *p*-value |
|---|---|---|---|---|
| N-(2-Furoyl)glycine | 13.83 | 0.001 | 0.902 | 0.001 |
| Histidine | 2.61 | 0.005 | 0.799 | 0.005 |
| D-Tagatose | 2.47 | 0.031 | 0.732 | 0.031 |
| Gluconic acid | 1.88 | 0.014 | 0.656 | 0.146 |
| Sorbitol | 1.60 | 0.038 | 0.763 | 0.014 |
| (±)-Sulfobutanedioic acid | 1.58 | 0.031 | 0.732 | 0.031 |
| Phenol sulphate | 1.58 | 0.042 | 0.719 | 0.042 |

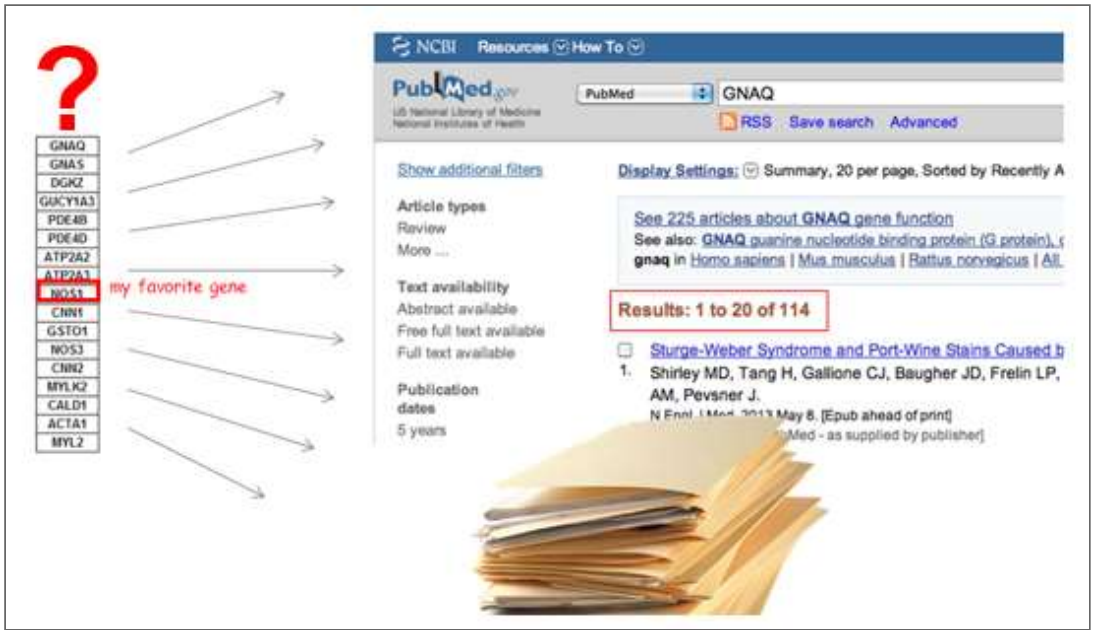
Fold changes and AUC of metabolites whose concentrations were significantly increased in the patients with breast cancer compared to the healthy controls

- Metabolites lists are diverse:
  - Truncated vs All the features analyzed
  - Ordered vs unordered
  - Only IDs vs IDs with difference measures

# THE *WHERE TO, NOW?* QUESTION

Once a list of feature is obtained it can be studied on a one-by-one basis

- Select some features for biochemical validation,
- Map individual features to specific pathways,
- Perform functional assays,
- Do a literature search …



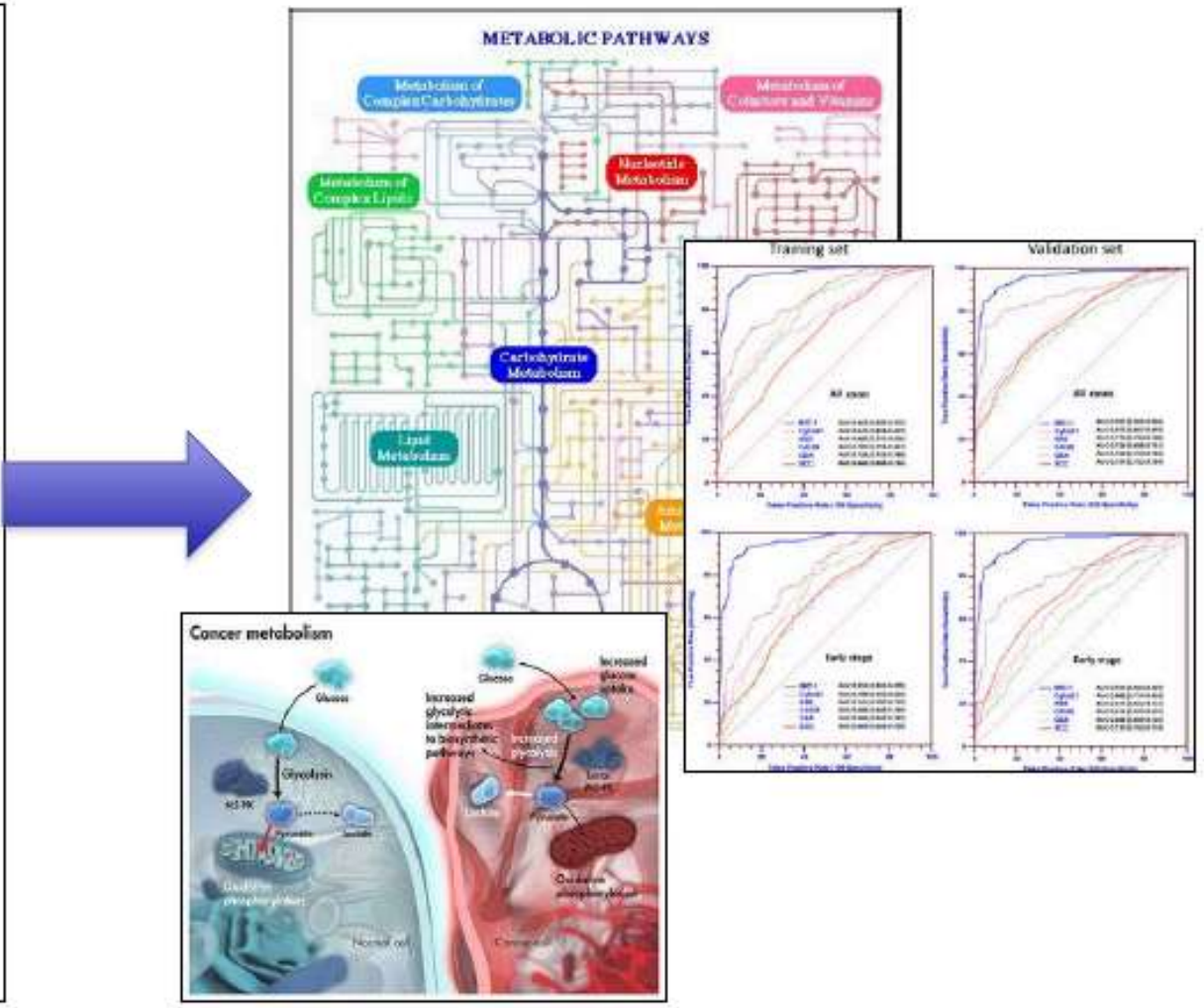

- This will yield useful information, but
  - It may be slow and resource-consuming
  - It does not account for **interaction** between features.

## AND HERE COMES PATHWAY ANALYSIS

- Pathway Analysis studies the list as a whole.
- With this aim it combines:
  - The list of features, with
  - Pre-existing sources of information related to them
- And, after some processing, it yields
  - some type of scores about
  - groups of features appearing to be significantly related with the process being studied.

# HOW CAN WE INTERPRET THESE LISTS?

| Compound | Retention Time (min) | Conc. in Urine (µM) | Compound | Retention Time (min) | Conc. in Urine (µM) |
|---|---|---|---|---|---|
| Dns-o-phospho -L-serine | 0.92 | <D.L.* | **Dns-Ile** | 6.35 | 25 |
| Dns-o-phospho -L-tyrosine | 0.96 | <D.L. | **Dns-3-aminosalicylic acid** | 6.44 | 0.5 |
| Dns-adenosine monophosphate | 0.99 | <D.L. | **Dns-pipecolic acid** | 6.50 | 0.5 |
| **Dns-o-phosphoethanolamine** | 1.06 | 16 | **Dns-Leu** | 6.54 | 54 |
| **Dns-glucosamine** | 1.08 | 22 | **Dns-cystathionine** | 6.54 | 0.3 |
| Dns-o-phospho -L-threonine | 1.09 | <D.L. | **Dns-Leu-Pro** | 6.60 | 0.4 |
| Dns-5-dimet hylamino purine | 1.20 | <D.L. | **Dns-5-hydroxylysine** | 6.65 | 1.6 |
| **Dns-3-methyl -histidine** | 1.22 | 80 | **Dns-Cystine** | 6.73 | 160 |
| **Dns-taurine** | 1.25 | 634 | **Dns-N-norleucine** | 6.81 | 0.1 |
| **Dns-carnosine** | 1.34 | 28 | Dns-5-hydroxydopamine | 7.17 | <D.L. |
| **Dns-Arg** | 1.53 | 36 | **Dns-dimethylamine** | 7.33 | 299 |
| **Dns-Asn** | 1.55 | 133 | **Dns-5-HIAA** | 7.46 | 18 |
| **Dns-hypotaurine** | 1.58 | 10 | **Dns-umbelliferone** | 7.47 | 1.9 |
| **Dns-homocarnosine** | 1.61 | 3.9 | Dns-2,3-diaminopropionic acid | 7.53 | <D.L. |
| Dns-guanidine | 1.62 | <D.L. | **Dns-L-ornithine** | 7.70 | 15 |
| **Dns-Gln** | 1.72 | 633 | **Dns-4-acetyamidophenol** | 7.73 | 51 |
| **Dns-allantoin** | 1.83 | 3.8 | **Dns-procaine** | 7.73 | 0.9 |
| **Dns-L-citrulline** | 1.87 | 2.9 | **Dns-homocystine** | 7.76 | 3.3 |
| **Dns-1 (or 3 -)-methylhistamine** | 1.94 | 1.9 | **Dns-acetaminophen** | 7.97 | 82 |
| **Dns-adenosine** | 2.06 | 2.6 | **Dns-Phe-Phe** | 8.03 | 0.4 |
| Dns-methylguanidine | 2.20 | <D.L. | **Dns-5-methyo xysalicylic acid** | 8.04 | 2.1 |
| **Dns-Ser** | 2.34 | 541 | **Dns-Lys** | 8.16 | 154 |
| **Dns-aspartic acid amide** | 2.44 | 26 | Dns-aniline | 8.17 | <D.L. |
| **Dns-4-hydroxy -proline** | 2.55 | 2.3 | **Dns-Leu-Phe** | 8.22 | 0.3 |
| **Dns-Glu** | 2.57 | 21 | **Dns-His** | 8.35 | 1550 |
| **Dns-Asp** | 2.60 | 90 | Dns-4-thalysine | 8.37 | <D.L. |
| **Dns-Thr** | 3.03 | 157 | Dns-benzylamine | 8.38 | <D.L. |
| Dns-epinephrine | 3.05 | <D.L. | **Dns-l-ephedrine** | 8.50 | 0.6 |
| **Dns-ethanolamine** | 3.11 | 471 | **Dns-tryptamine** | 8.63 | 0.4 |
| **Dns-aminoadipic acid** | 3.17 | 18 | Dns-pyridoxamine | 8.84 | <D.L. |
| **Dns-Gly** | 3.43 | 2510 | Dns-2-methyl -benzylamine | 9.24 | <D.L. |
| **Dns-Ala** | 3.66 | 590 | **Dns-5-hydroxytriptophan** | 9.25 | 0.12 |
| **Dns-aminolevulinic acid** | 3.97 | 30 | **Dns-1,3 -diaminopropane** | 9.44 | 0.23 |
| **Dns-r-amino -butyric acid** | 3.98 | 4.6 | **Dns-putrescine** | 9.60 | 0.5 |
| **Dns-p-amino -hippuric acid** | 3.98 | 2.9 | **Dns-1,2 -diaminopropane** | 9.66 | 0.1 |
| **Dns-5-hydro xymethyluriid** | 4.58 | 1.9 | **Dns-tyrosinamide** | 9.79 | 29 |
| **Dns-tryptophanamide** | 4.70 | 5.5 | **Dns-dopamine** | 10.06 | 140 |
| Dns-isoguanine | 4.75 | <D.L. | **Dns-cadaverine** | 10.08 | 0.06 |
| **Dns-5-aminopentanoic acid** | 4.79 | 1.5 | **Dns-histamine** | 10.19 | 0.4 |
| **Dns-sarcosine** | 4.81 | 7.2 | **Dns-3-methoxy -tyramine** | 10.19 | 9.2 |
| **Dns-3-amino -isobutyrate** | 4.84 | 85 | **Dns-Tyr** | 10.29 | 321 |
| **Dns-2-aminobutyric acid** | 4.89 | 17 | Dns-cysteamine | 10.43 | <D.L. |




From Lists to Biology

# ONTOLOGIES, DATABASES AND METABOLITE SETS

# THE ELEMENTS OF PATHWAYS ANALYSIS

- Loosely speaking, to do Pathway Analysis one needs:
  - A list of features, characterizing a process.
  - A source of information about these features.
  - An algorithm to highlight relevant information by linking *list* and *source*.
  - A tool implementing the algorithm.
- In this section, we focus on *sources of information* and on *how to provide it to the algorithms*.

# METABOLITES DATABASES

Some common databases in Metabolomics

# THE HUMAN METABOLOME DB

.

The Human Metabolome Database

- Detailed information about human metabolites, their structures, pathways, origins, concentrations, functions and reference spectra
- HMDB has 248,855 metabolites, 132,335 pathways, 3.1 million MS and NMR spectra, metabolite biomarker data on >600 diseases
- A resource established to provide reference metabolite values for human disease, human exposures & population health
- Captures both targeted and untargeted metabolomics (and exposomics) data

# THE FOOD CONSTITUENT DATABASE

.

The Food Constituent Database

- Database of 70,000+ compounds found in 727 foods and their effects on flavour, aroma, colour and human health
- Comprehensive concentration information to ID foods that are rich in particular micronutrients
- Links chemistry to food types (biological species) to flavour, aroma, colour and human health
- Supports sequence, spectral, structure and text searches

# THE KEGG DB

Kyoto Encyclopedia of Genes and Genomes

- The “Go-to” Metabolic Pathway Database
- Has 535 “canonical” pathway diagrams or maps covering 5994 organisms for a total of 604,808 pathways
- ~170 metabolic pathways covering 18,553 compounds, includes many disease pathways (80), protein signaling (70) pathways, and biological process pathways (70)
- Metabolic pathways are highly schematized and mostly limited to catabolic and anabolic processes

# SMALL MOLECULE PATHWAY DATABASE

.

The Small Molecule Pathway Database (SMPDB)

- Nearly 48,900 hand-drawn small molecule pathways – 404 drug action pathways – 20,251 metabolic disease pathways – 27,876 metabolic pathways – 160+ signaling and other pathways
- Depicts organs, cell compartments, organelles, protein locations, and protein quaternary structures
- Maps gene chip & metabolomic data
- Converts gene, protein or chemical lists to pathways or disease diagnoses

# FROM DATABASES TO METABOLITE SETS

## MSEA Workflow

# METABOLITES SET LIBRARIES

Overview of MSEA’s metabolite set libraries

# ANALYSIS METHODS

# LIMITATIONS AND RECOMMENDATIONS

# PATHWAY ANALYSIS TOOLS

# A USE CASE WITH METABOANALYST