Visualization of Mass Spectrometry Related Data

ASMS Bioinformatics Interest Group

Arzu Tuğçe Güler, Amsterdam University Medical Centers, Amsterdam
Claire O’Donovan, European Bioinformatics Institute, Cambridge
Agenda

❖ Overview
❖ Visualization tools
❖ Metabolomics study examples
❖ Interactive discussion
Two different worlds?

Figure made by Arzu Tugce Guler created with Biorender.com
A picture is worth a thousand words

Anscombe’s Quartet

```r
> datasets::anscombe
  x1  x2  x3  x4  y1  y2  y3  y4
1  10  10  10  10  8.04  9.14  7.46  6.58
2   8   8   8   8  6.94  8.14  6.77  5.76
3  13  13  13  13  8.81  8.93  7.81  6.43
4   9   9   9   9  7.91  8.00  8.04  5.05
5  11  11  11  11  8.04  6.94  6.62  7.04
6  14  14  14  14  8.04  6.94  6.62  7.04
7   6   6   6   6  6.28  7.76  6.18  7.26
8  12  12  12  12  8.14  8.47  7.26  6.13
9  14  14  14  14  8.46  7.09  8.00  5.05
10  8  8  8  8  6.28  7.76  6.18  7.26
11  8  8  8  8  8.14  6.94  6.62  7.04

> fBasics::basicstats(anscombe)
  x1  x2  x3  x4  y1  y2  y3  y4
NAs 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000 0.00000
Sum 99.00000 99.00000 99.00000 99.00000 99.00000 99.00000 82.51000 82.51000 82.51000 82.51000
SE Mean 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 0.612541 0.612541 0.612541 0.612541
Skewness 0.00000 0.00000 0.00000 0.00000 2.466911 -0.048374 -0.978693 1.380120 -0.514319 1.240044
Kurtosis -1.528926 -1.528926 -1.528926 4.520661 -1.199123 -0.514319 1.240044 0.628751
```
Take home lesson: never trust summary statistics alone; always visualize your data!
Datasaurus (aka Anscombosaurus)

Take home lesson: “never trust summary statistics alone; always visualize your data!”

Visualization of mass spectrometry data

2. [https://phosphopedia.gs.washington.edu/PhosphoproteomicsAssay/index.xhtml](https://phosphopedia.gs.washington.edu/PhosphoproteomicsAssay/index.xhtml)
The -omics

1. Transcription
   - DNA → mRNA
2. Translation
   - mRNA → Protein

GENOMICS
TRANSCRIPTOMICS
PROTEOMICS
METABOLOMICS

Figure made by Arzu Tugce Guler
created with Biorender.com
Visualizations for mass spectrometry-based proteomics data

• Numeric data (with or without labels)
  e.g. volcano plots, heatmaps, boxplots, PCAs etc.

• Trees / networks
  e.g. ontologies, gene ontology enrichments, protein interaction networks etc.

• Images
  e.g. experiment workflows, anatomical visualizations, western blots etc.
Volcano plots
Heatmap
Boxplot
PCA

(Median) PCA of all samples
Western blots / gels, microscope images
Tools for visualizing numeric data

Excel*
- Quick and adaptable
- Especially good for internal sanity checks
Useful R libraries for more advanced, publication quality visualizations

There are currently 161 Proteomics and 118 Mass Spectrometry packages in Bioconductor version 3.15.

- ggplot2
- lattice
- shiny*
- plotly*
- MSstats
- protViz

* commercial
Perseus
Tools for icons / anatomical visualizations

- bioicons
Tools for icons / anatomical visualizations

- BioRender*
Tools for icons / anatomical visualizations

- COMICS
Tools for network visualizations

- Cytoscape
Tools for network visualizations

- VOSviewer
Interactive plots

Collaboration & sharing
Data analysis & interactive graphs
Smart manuscripts & interactivity
FiglinQ
Data reuse
Data findability & accessibility

- FiglinQ*

- Raw data
- Analyze data
- Visualize data
- Publish manuscript
- Write manuscript, embed visualization

Permanent, 2-way connection between data, data analysis, data visualization and published manuscripts
World access to all data lifecycle stages

*commercial / paid subscription
<p>| | | | | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Nampt</td>
<td>0.215198204</td>
<td>0.29555106</td>
<td>0.45293429</td>
<td>0.82390337</td>
<td>0.328380436</td>
<td>-1.574401259</td>
<td>-1.554394041</td>
<td>-1.576914805</td>
<td>0.492651793</td>
<td>0.964351356</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Atrx</td>
<td>0.05386484</td>
<td>-0.434552983</td>
<td>0.44359081</td>
<td>1.353461981</td>
<td>1.010728656</td>
<td>-1.887060666</td>
<td>-1.380543833</td>
<td>-1.896587133</td>
<td>0.535104999</td>
<td>0.719460905</td>
<td>0.667079515</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>Add1</td>
<td>0.198700441</td>
<td>-0.016361643</td>
<td>0.040212072</td>
<td>0.736406579</td>
<td>-0.233907223</td>
<td>-1.302153099</td>
<td>-1.288216168</td>
<td>-0.533631742</td>
<td>1.187776089</td>
<td>1.9056741</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Cdk5</td>
<td>0.03017006</td>
<td>-0.776342154</td>
<td>0.743665509</td>
<td>0.962791921</td>
<td>-1.002786564</td>
<td>-0.605718374</td>
<td>-1.676633332</td>
<td>-0.268917114</td>
<td>-0.013392712</td>
<td>0.29367843</td>
<td>2.172544255</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>Rpl9</td>
<td>0.727042794</td>
<td>-0.954574717</td>
<td>-0.081508435</td>
<td>1.052151918</td>
<td>-0.345926365</td>
<td>-1.121611767</td>
<td>-1.852335399</td>
<td>-0.614648105</td>
<td>0.554352823</td>
<td>0.243497774</td>
<td>1.326697227</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Erp</td>
<td>0.376109749</td>
<td>-0.508586509</td>
<td>-0.168736849</td>
<td>0.848609746</td>
<td>-0.409578513</td>
<td>-2.162549519</td>
<td>-1.431763503</td>
<td>0.322842355</td>
<td>1.075832727</td>
<td>0.27483266</td>
<td>0.755981233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>Mvp</td>
<td>-0.033950245</td>
<td>-0.813685237</td>
<td>0.580776485</td>
<td>1.300670505</td>
<td>-0.854547679</td>
<td>-0.713124435</td>
<td>-1.240190825</td>
<td>-0.788911879</td>
<td>0.886659361</td>
<td>1.467129707</td>
<td>0.064919449</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>Cams4</td>
<td>-0.247023597</td>
<td>0.264265597</td>
<td>1.223739294</td>
<td>0.792339027</td>
<td>-0.476846274</td>
<td>-1.636339212</td>
<td>1.205765357</td>
<td>0.609995553</td>
<td>0.807082923</td>
<td>1.749877453</td>
<td>-0.117308887</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Searching for (visualization) tools

- Essential scientific and technical information about software tools, databases and services for bioinformatics and the life sciences.
**GenBank taxonomy tools**

Utilities for manipulations and visualization tasks on GenBank taxonomic information.

- **Taxonomy**
- **GenBank taxonomy tools**

**Gitools**

A framework for analysis and visualization of multidimensional genomic data using interactive heat-maps.

- **Genomics**
- **Visualisation**
- **Desktop application**

**Reitools**

Tools for the analysis of enrichment-based epigenomic data. Features include summarization and visualization of epigenomic data across promoters according to gene expression context, finding regions of differential methylation/binding, BayMeth for quantifying methylation etc.

- **Epigenomics**
- **Differentially-methylated region identification**
- **Transcriptional regulatory element prediction**
- **Command-Line tool**
- **Library**
- **GPL-2.0**
- **BioConductor**
Colors & ambiguity
Colors & ambiguity
Mass spectrometry data in MetaboLights
the home for metabolomics experiments and derived information

Claire O’Donovan
EMBL-EBI Metabolomics Team
www.ebi.ac.uk
FAIR Guiding Principles

Provide a guideline for data producers and publishers to enhance the reusability of scientific data

Findable
Accessible
Interoperable
Re-useable

Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse.
Can metabolomics learn from the other omic techniques and not repeat many of the same mistakes that have been made in which a lot of data were collected but little information was gleaned?
The Metabolomics Standards Initiative (MSI)

The MSI is an academic policy provider, to support the development of open data and metadata formats for metabolomics.

- community-agreed reporting standards
- minimal information (MI) checklists
- data exchange formats to support the MIs reporting standards, allow data to be efficiently applied, shared and reused
Open Source Study Repository & Metabolite Knowledgebase

https://www.ebi.ac.uk/metabolights
MetaboLights user base

MaHPIC
Malaria Host-Pathogen Interaction Center (NIH)

Max-Planck

OIST
Okinawa Institute of Science and Technology Graduate University

Pacific Northwest
National Laboratory

Woods Hole
Oceanographic Institution

CHEAR
Children’s Health Exposure Analysis Resource

NIH

PHENOME CENTRE
BIRMINGHAM

UNIVERSITY OF CAMBRIDGE

National Phenome Centre

EMBL-EBI
Geographical distribution of submitted studies

MetaboLights study submission rates per year

BioRxiv 3.3%
PNAS 4.2%
Anal Chem 5.9%
MetaboCare 4.1%
Cell Press journals 3.0%
BMC-series journals 3.5%
J Proteome Res 7.3%
Frontiers journals 10.0%
Metabolomics 13.6%
Nature group 27.3%
PLoS 17.3%
MetaboLights Database

- Cross species, cross technique
- Comprehensive information collection
- Open access, discoverable through various portals
MetaboLights: Ontology mapping

Ontologies: Controlled vocabulary

<table>
<thead>
<tr>
<th>Area</th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species/Organism, Organism part</td>
<td>National Center for Biotechnology Information Organismal Classification (NCBITAXON), World Register of Marine Species (WoRMS), National Cancer Institute Thesaurus (NCIT), Environment Ontology (ENVO), Brenda Tissue and Enzyme Source Ontology (BTO)</td>
</tr>
<tr>
<td>Sample details/factors (e.g., disease state, treatment, timings)</td>
<td>National Cancer Institute Thesaurus (NCIT), Experimental Factor Ontology (EFO), Plant Ontology (PO), Unit of Measurement Ontology (UO), Human Phenotype Ontology (HP), Orphanet Rare Disease ontology (ORDO)</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>Metabolomics Standards Initiative Ontology (MSIO), Chemical Methods Ontology (CHMO), Mass Spectrometry Ontology (MS), Unit of Measurement Ontology (UO), MetaboLights (MTBLS)</td>
</tr>
<tr>
<td>Metabolites</td>
<td>Chemical Entities of Biological Interest (ChEBI) ... cross ref to other databases</td>
</tr>
</tbody>
</table>

**Diagram:**
- Query Term
- Metabolights:owl
- EBI Zooma service
- EBI OLS service
- Bioportal
- Manually curated terms
- Relative terms
- Metabolights approved terms
- Other institute approved terms
- NSIO: instruments
- CHMC: Chromatography
- EFO: Factors
- NCBITAXON: Taxonomy
- Metabolomics approved terms
- Search certain ontologies branches
- Search certain ontologies
- Output

**Logos:**
- NCBI
- WoRMS
- MetaboLights (MTBLS)
- Orphanet
- Human Phenotype Ontology
- ChEBI
- Plant Ontology
- EFO
- BTO
- Zooma
- Webulous

**Footer:**
- EMBL-EBI
MetaboLights: Browse or search studies

Utilization of Metabolomics to Identify Serum Biomarkers for Hepatocellular Carcinoma in Patients with Liver Cirrhosis

A lipidomic human population and translational feeding study of hepatic steatosis and de novo lipogenesis (Human plasma assay)

Plasma Lipidomics for the Identification of Antecedent Memory Impairment
MTBLS749: Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington’s disease subjects

Kim Kullima, Stephanie Herman

Huntington’s disease (HD) is a severe neurological disease leading to psychiatric symptoms, motor impairment and cognitive decline. The disease is caused by a CAG expansion in the huntingtin (HTT) gene, but how this translates into the clinical phenotype of HD remains elusive. Using liquid chromatography mass spectrometry, we analysed the metabolome of cerebrospinal fluid (CSF) from premanifest and manifest HD subjects as well as control subjects. Inter-group differences revealed that the tyrosine metabolism, including tyrosine, thyroxine, L-DOPA and dopamine, was significantly altered in manifest compared with premanifest HD. These metabolites demonstrated moderate to strong associations to measures of disease severity and symptoms. Tyrosine and dopamine also correlated with the five-year risk of onset in premanifest HD subjects. The phenylalanine and the purine metabolisms were also significantly altered, but associated less to disease severity. Decreased levels of lumichrome were commonly found in mutated HTT carriers and the levels correlated with the five-year risk of disease onset in premanifest carriers. These biochemical findings demonstrates that the CSF metabolome can be used to characterize molecular pathogenesis occurring in HD, which may be essential for future development of novel HD therapies.

PUBLICATIONS

Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington’s disease subjects

Herman Stephania, Niemelä Valter, Emami Khoosna PhD, Sun...
MTBLS749: Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington’s disease subjects

Kim Kultima, Stephanie Herman

Huntington’s disease (HD) is a severe neurological disease leading to psychiatric symptoms, motor impairment and cognitive decline. The disease is caused by a CAG expansion in the huntingtin (HTT) gene, but how this translates into the clinical phenotype of HD remains elusive. Using liquid chromatography mass spectrometry, we analyzed the metabolome of cerebrospinal fluid (CSF) from premanifest and manifest HD subjects as well as control subjects. Inter-group differences revealed that the tyrosine metabolism, including tyrosine, thyroxine, L-DOPA and dopamine, was significantly altered in manifest compared with premanifest HD. These metabolites demonstrated moderate to strong associations to measures of disease severity and symptoms. Thyroxine and dopamine also correlated with the five year risk of onset in premanifest HD subjects. The phenylalanine and the purine metabolisms were also significantly altered, but associated less to disease severity. Decreased levels of taurine were commonly found in mutated HTT carriers and the levels correlated with the five year risk of disease onset in premanifest carriers. These biochemical findings demonstrates that the CSF metabolome can be used to characterize molecular pathogenesis occurring in HD, which may be essential for future development of novel HD therapies.

Homo sapiens

PUBLICATIONS

Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling i...

Herman Stephanie, Niemelä Valter, Emami Khoonsari Payam, Sun...
MTBLS749: Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington’s disease subjects

Kim Kurima, Stephanie Hermann

Huntington’s disease (HD) is a severe neurological disease leading to psychiatric symptoms, motor impairment and cognitive decline. The disease is caused by a CAG expansion in the huntingtin (HTT) gene, but how this translates into the clinical phenotype of HD remains elusive. Using liquid chromatography mass spectrometry, we analysed the metabolome of cerebrospinal fluid (CSF) from premanifest and manifest HD subjects as well as control subjects. Inter-group differences revealed that the tyrosine metabolism, including tyrosine, tyramine, L-DOPA and dopamine, was significantly altered in manifest compared with premanifest HD. These metabolites demonstrated moderate to strong associations to measures of disease severity and symptoms. Tyrosine and dopamine also correlated with the five-year risk of onset in premanifest HD subjects. The phenylalanine and the purine metabolisms were also significantly altered, but associated less to disease severity. Decreased levels of lumichrome were commonly found in mutated HTT carriers and the levels correlated with the five-year risk of disease onset in premanifest carriers. These biochemical findings demonstrates that the CSF metabolome can be used to characterize molecular pathogenesis occurring in HD, which may be essential for future development of novel HD therapies.
MTBLS749: Alterations in the tyrosine and phenylalanine pathways revealed by biochemical profiling in cerebrospinal fluid of Huntington’s disease subjects

Kim Kullima, Stephanie Herman

Huntington’s disease (HD) is a severe neurological disease leading to psychiatric symptoms, motor impairment and cognitive decline. The disease is caused by a CAG expansion in the huntingtin (HTT) gene, but how this translates into the clinical phenotype of HD remains elusive. Using liquid chromatography mass spectrometry, we analysed the metabolome of cerebrospinal fluid (CSF) from premanifest and manifest HD subjects as well as control subjects. Inter-group differences revealed that the tyrosine metabolism, including tyrosine, tryptophan, L-DOPA and dopamine, was significantly altered in manifest compared with premanifest HD. These metabolites demonstrated moderate to strong associations to measures of disease severity and symptoms. Tyrosine and dopamine also correlated with the five year risk of onset in premanifest HD subjects. The phenylalanine and the purine metabolisms were also significantly altered, but associated less to disease severity. Decreased levels of lumichrome were commonly found in mutated HTT carriers and the levels correlated with the five year risk of disease onset in premanifest carriers. These biochemical findings demonstrate that the CSF metabolome can be used to characterize molecular pathogenesis occurring in HD, which may be essential for future development of novel HD therapies.
Sample collection

The lumbar puncture was performed through the L3/4 or L4/5 interspace, and CSF was collected in a polypropylene tube that was centrifuged for 5 min at 260 x g at room temperature. The supernatant was transferred into fresh polypropylene tubes in aliquots of 240 µL and stored at -80 °C until analysis.

Extraction

CSF samples were thawed on ice, and 100 µL was transferred and mixed with 410 µL ice-cold methanol (MeOH) supplemented with an internal standard cocktail (D4-iketas-prostaglandin-F1-alpha, D4-thromboxane-B2, D4-prostaglandin-F2-alpha, D4-prostaglandin-E2, D4-prostaglandin-D2, D4-15-deoxy-delta12, 14-prostaglandin-J2, D4-cortisol and levonorgestrel) at a final concentration of 0.25 µM. The samples were further vortexed for 15 s and incubated at -20 °C for 30 min, followed by centrifugation at 2040 x g.

Chromatography

The liquid chromatography-mass spectrometry analyses were performed using a Thermo Ultimate 3000 HPLC and Thermo Q Exactive Orbitrap mass spectrometer. 10 µL sample was injected to a Thermo Accucore AQ RP C18 column (100 x 2.1 mm, 2.6 µm particle size). The analytical gradient was initiated with an isocratic flow for 3 min (0% B), followed by a 2.6 min gradient (0-10% B), 8.3 min (10-100% B) and 3 min (100% B), followed, finally, by re-equilibration and washing of the column for 3 min (0% B).

Mass spectrometry

Mass spectrometry data were acquired in profile mode (in positive and negative ionization mode) on a Thermo Q Exactive Orbitrap mass spectrometer, using a mass range of 70-900 m/z during the first 5 min and 140-900 m/z in the following 15 min (to avoid low mass contaminants) in the positive ionization mode and 70-900 m/z throughout in negative ionization mode. To improve the identification of metabolites, tandem mass spectrometry analyses in positive and negative ionization mode were performed.

Data transformation

The acquired raw data was converted to an open source format (mzML). Peaks were centroided by msconvert from ProseWizard [1] and preprocessed using the following pipeline within the KNIME platform [2]: the peak-picked data was quantified by FeatureFinderMetabo [3] and the resulting features were linked across the samples using featurelinkerUnsabeledQ [4], allowing 10 s retention time tolerance and 5 ppm mass deviation (performed irrespective of charge state across the samples). The non-default.

Metabolite identification

All metabolic features with a 75% coverage across samples were matched against an in-house library of characterized metabolites using a 15 ppm mass tolerance and a 20 s time window. Metabolites (identified metabolic features) of interest were manually curated on MZmine level when available.
Sample collection

The lumbar puncture was performed through the L3/L4 or L4/L5 interspace, and CSF was collected in a polypropylene tube that was centrifuged for 5 min at 260 x g at room temperature. The supernatant was transferred into fresh polypropylene tubes in aliquots of 240 μL and stored at -80 °C until analysis.

Extraction

CSF samples were thawed from -80°C, cocktai (0.4g/mL glycine/HCl buffer, 1N HCl, 0.1M MgCl2, 14-propanol, chloroform), D4-NSP (n-stearoyl phosphatidylcholine, D4-NSP, and D4-PPC (1-palmitoyl-2-oleoyl phosphatidylcholine, D4-PPC), at -20°C for 10 min, followed by vortexing.

Chromatography

The liquid chromatography–mass spectrometry analyses were performed using a Thermo Ultimate 3000 HPLC and Thermo Q-Exactive Orbitrap mass spectrometer. 10 μl sample was injected to a Thermo Accucore aQ RP C18 column (100 x 2.1 mm, 2.6 μm particle size). The analytical gradient was initiated with an isocratic flow for 3 min (0% B), followed by a 2.6 min gradient (0–10% B), 8.3 min (10–100% B) and 3 min (100% B), followed, finally, by re-equilibration and washing of the column for 3 min (0% B), with negative ionization mode.

Mass spectrometry

Mass spectrometry data were acquired in profile mode (in positive and negative ionization mode) on a Thermo Q-Exactive Orbitrap mass spectrometer, using a mass range of 70–900 m/z during the first 5 min and 148–900 m/z in the following 15 min (to avoid low mass contaminants) in the positive ionization mode and 70–900 m/z throughout in negative ionization mode. To improve the identification of metabolites, tandem mass spectrometry analyses in positive and negative ionization mode were performed.

Metabolite identification

All metabolic features with a 75% coverage across samples were matched against an in-house library of characterized metabolites using a 15 ppm mass tolerance and a 20 s time window. Metabolites (identified metabolic features) of interest were manually curated on MS/MS level when available.
## Study information

### Filter

<table>
<thead>
<tr>
<th>Protocol REF</th>
<th>Sample Name</th>
<th>Organism</th>
<th>Organism part</th>
<th>Phenotype</th>
<th>Replicate</th>
<th>MS format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample collection</td>
<td>1_Rep1_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>1_Rep1_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>1_Rep2_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>1_Rep2_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>10_Rep1_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>10_Rep1_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>10_Rep2_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>10_Rep2_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>11_Rep1_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>11_Rep1_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>1</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>11_Rep2_NEG</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
<tr>
<td>Sample collection</td>
<td>11_Rep2_POS</td>
<td>Homo sapiens</td>
<td>cerebrospinal fluid</td>
<td>control</td>
<td>2</td>
<td>MS1 format</td>
</tr>
</tbody>
</table>
MetaboLights – Study information

<table>
<thead>
<tr>
<th>Protocol REF</th>
<th>Parameter Value - Chromatography Instrument</th>
<th>Parameter Value - Autosampler model</th>
<th>Parameter Value - Column model</th>
<th>Parameter Value - Guard column</th>
<th>Labeled Extract Name</th>
<th>Label</th>
<th>Protocol REF</th>
<th>Parameter Value - Scan polarity</th>
<th>Parameter Value - Scan m/z range</th>
<th>Parameter Value - Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromatography</td>
<td>Waters ACQUITY UPLC H-Class System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromatography</td>
<td>Waters ACQUITY UPLC H-Class System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chromatography</td>
<td>Waters ACQUITY UPLC H-Class System</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### MetaboLights – Study information

<table>
<thead>
<tr>
<th>Assay Sheet 1</th>
<th>Assay Sheet 2</th>
<th>Assay Sheet 3</th>
<th>Assay Sheet 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC_HILICpen01</td>
<td>QC_HILICpen02</td>
<td>QC_HILICpen03</td>
<td></td>
</tr>
</tbody>
</table>

**File:** a_MTBL81987_LC-MS_negative_reverse_phase_metabolite_profiling.txt

**Filter:**

<table>
<thead>
<tr>
<th>MS Assay Name</th>
<th>Raw Spectral Data File</th>
<th>Protocol REF</th>
<th>Normalization Name</th>
<th>Derived Spectral Data File</th>
<th>Protocol R</th>
</tr>
</thead>
<tbody>
<tr>
<td>QC_RPLCneg01</td>
<td>RPLCneg/QC_RPLCneg01.raw</td>
<td>Data transformation</td>
<td>RPLCneg/RPLCnegative_MSMS_20eV.mgf</td>
<td>Metabolite identifiable</td>
<td></td>
</tr>
<tr>
<td>QC_RPLCneg02</td>
<td>RPLCneg/QC_RPLCneg02.raw</td>
<td>Data transformation</td>
<td>RPLCneg/RPLCnegative_MSMS_20eV.mgf</td>
<td>Metabolite identifiable</td>
<td></td>
</tr>
<tr>
<td>QC_RPLCneg03</td>
<td>RPLCneg/QC_RPLCneg03.raw</td>
<td>Data transformation</td>
<td>RPLCneg/RPLCnegative_MSMS_20eV.mgf</td>
<td>Metabolite identifiable</td>
<td></td>
</tr>
<tr>
<td>Structure</td>
<td>Database identifier</td>
<td>Chemical formula</td>
<td>SMILES</td>
<td>InChI</td>
<td>Metabolite identification</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------</td>
<td>------------------</td>
<td>---------</td>
<td>-------</td>
<td>---------------------------</td>
</tr>
<tr>
<td><img src="image" alt="Proline structure" /></td>
<td>CHEBI:26571</td>
<td>OCN(=O)C1(CCN1)</td>
<td>InChI=1S/C5H9NO2/c7-5(6)h4-2-1-3-4-6/h4H,1-3H2,1-7H3</td>
<td>Proline</td>
<td>116.07029</td>
</tr>
<tr>
<td><img src="image" alt="Betaine structure" /></td>
<td>CHEBI:17750</td>
<td>C[N+]C(C)[O][O]O-3=O</td>
<td>InChI=1S/C5H11N03O/c1-6q,3H4-5(7)8H3+4H2-1,3H3</td>
<td>Betaine</td>
<td>118.08747</td>
</tr>
<tr>
<td><img src="image" alt="Glutamine structure" /></td>
<td>CHEBI:28300</td>
<td>NQCCO[N:]=O[C][O]:O</td>
<td>InChI=1S/C5H11NO2C/c5-3(9)h1,2-4-7h8,c3H4+2H2,1H2,7,8(9,10)</td>
<td>Glutamine</td>
<td>147.07607</td>
</tr>
<tr>
<td><img src="image" alt="Methionine structure" /></td>
<td>CHEBI:16811</td>
<td>CSC(CCN[O]C)=O</td>
<td>InChI=1S/C5H11NO2S/c5-3-2-4-8(9)8H4H2,1H2,3,7,8(9,10)</td>
<td>Methionine</td>
<td>150.05818</td>
</tr>
</tbody>
</table>
**MetaboLights – Study information**

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Protocols</th>
<th>Samples</th>
<th>Assays</th>
<th>Metabolites</th>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FTP Download</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="a_MTBLS1987 LC-MS_negative_hilic_metabolite_profiling.txt" /> September 08 2020 08:33:26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="a_MTBLS1987 LC-MS_negative_reverse-phase_metabolite_profiling.txt" /> September 08 2020 08:22:02</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="a_MTBLS1987 LC-MS_positive_hilic_metabolite_profiling.txt" /> September 08 2020 08:33:47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="a_MTBLS1987 LC-MS_positive_reverse-phase_metabolite_profiling.txt" /> September 08 2020 08:57:54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="i_Investigation.txt" /> September 08 2020 14:10:11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="m_MTBLS1987 LC-MS_negative_hilic_metabolite_profiling_v2.maf.txt" /> December 06 2020 12:20:17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="m_MTBLS1987 LC-MS_negative_reverse-phase_metabolite_profiling_v2.maf.txt" /> September 08 2020 08:33:13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="m_MTBLS1987 LC-MS_positive_hilic_metabolite_profiling_v2.maf.txt" /> September 08 2020 08:50:33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="m_MTBLS1987 LC-MS_positive_reverse-phase_metabolite_profiling_v2.maf.txt" /> September 08 2020 09:01:21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td><img src="MTBLS1987" alt="s_MTBLS1987.txt" /> December 06 2020 12:45:46</td>
</tr>
</tbody>
</table>

**RAW / DERIVED FILES**

<table>
<thead>
<tr>
<th>RAW / DERIVED FILES</th>
<th>File Name</th>
<th>Last modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>search raw Files</td>
<td>HILICneg September 07 2020 15:54:36</td>
<td><img src="MTBLS1987" alt="Download" /></td>
</tr>
<tr>
<td></td>
<td>HILICpos September 07 2020 15:53:13</td>
<td><img src="MTBLS1987" alt="Download" /></td>
</tr>
<tr>
<td></td>
<td>RPLCneg September 16 2020 22:14:29</td>
<td><img src="MTBLS1987" alt="Download" /></td>
</tr>
<tr>
<td></td>
<td>RPLCpos September 07 2020 16:33:31</td>
<td><img src="MTBLS1987" alt="Download" /></td>
</tr>
</tbody>
</table>
Acknowledgements

EMBL-EBI Metabolomics team
• Claire O’Donovan (PI)
• Pamela Pruski
• Mark Williams
• Felix Amaladoss
• Callum Martin
• Thomas Payne

The Metabolomics team’s activities were supported in the past year by the above funding bodies
Questions:

• Do you submit your data to public repositories? MetaboLights/Metabolomics Workbench/MetaboBank (and PRIDE)

• Do you reuse public mass spectrometry studies?
  a) Metadata
  b) Raw data
  c) Both
Questions:

• What data do you think we are missing?

• What tools do you use when manipulating/viewing the data? – either your internal data or the public data

• What tools would you like available with the repositories?