
METUMPX v2.1

User Guide



Copyright (C) 2019, Team MetumpX

Team MetumpX, 2017-2019

Bilal Wajid
Hasan Iqbal
Momina Jamil

MetumpX 2.1 User Guide

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see:
<https://www.gnu.org/licenses/>

NECESSARY SERVICING, REPAIR OR CORRECTION.

Limitation of Liability:

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Disclaimer of Warranty:

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL

MetumpX v2.1

Contents

About this Guide	1
Abstract	1
Purpose	1
Intended Audience	1
Required Background	1
How This Guide is Organized	2
Contact	2
MetumpX Package	3
Installing MetumpX software	3
Software Installation Process	9
Installing MetumpX software	9
Microsoft Windows Support	19
Installing MetumpX software on Microsoft Windows	19
Introduction to Metabolomics	29
In this Chapter	29
Introduction	29
Interactome	30
Metabolomic Study Strategies	31
Metabolomics Workflow	33
In this Chapter	33
Introduction	33
Untargeted Metabolomics Workflow	34
Sample Acquisition	35
Sample Preparation	35
Data Mining and Extraction	39

Metabolite Identification	45
Data Analysis	46
Data Visualization	49

About this Guide

Abstract

MetumpX is a free, easily distributable software installation package for freeware tools related to Mass Spectrometry (MS) based Untargeted Metabolomics.

MetumpX is a Ubuntu based software package that facilitate easy download and installation of tools related to Metabolomics software pipeline as per user requirement. It also downloads the dependencies prior to the tools for which they are required. MetumpX provides an interactive graphical user interface which is easily understandable for users unaware of UNIX-shell language.

Purpose

The basic purpose of this document is to provide the complete description about the installation of MetumpX shell file. This document also explains the complete pipeline of Metabolomics; also shows how this pipeline is implemented in the installation package. Running MetumpX on Microsoft Windows through Virtualization is also explained.

Intended Audience

This document is intended for researchers, life scientists and users who wish to install the tools related to Metabolomics.

Required Background

Team MetumpX has made every attempt to make this a step by step guide. However, some familiarity with Linux operating system as well as software and hardware requirements of MetumpX are assumed.

How This Guide is Organized

This guide is organized into sections grouped according to the intended use by the user:

- About This Guide (Chapter 1) describes this document's purpose and intended audience.
- Software Installation Process (Chapter 2) describes how to install aXonica.
- Microsoft Windows Support (Chapter 3) describes a step by step procedure to configure a virtual OS for aXonica on Windows.
- Data Acquisition and Software Pipeline (Chapter 6) describes the acquisition of raw data image of MRI machine and implementation of software pipeline of Biological image development.
- Tutorial Datasets (Chapter 7) describes how to download tutorial datasets.
- Image Preprocessing (Chapter 8) describes the step by step solution to recommended software.
- Image Processing (Chapter 9) describes the step by step solution to recommended software.
- Structural Analysis (Chapter 10) describes the step by step solution to recommended software.
- Data Management & Annotation (Chapter 11) describes the step by step solution to recommended software.
- Post Installation (Chapter 12) describes the working of each software included in the package.
- Software Uninstallation Process (Chapter 13) describes how to install aXonica or any specific tools.

Contact

For any further queries and suggestions, contact us at: hasaniqbal777@gmail.com or mominaj05@gmail.com

MetumpX Package

Installing MetumpX software

List of software tools and plugins included in MetumpX package are shown in the following table. Software which are recommended are also mentioned in the table. Download size, version and latest update of each software is mentioned for user convenience:

Sr. No.	Software Name	Size (MB)	Version	Latest Update	Recommended
Data Pre-processing					
Noise Filtering					
1	C.O.Peaks[1]	4.8	1.2.17	2019	✓
2	specmine[2]	15.2	2.0.3	2018	
3	intCor[3]	3.0	1.03.0	2014	
Chromatogram Alignment					
4	batchCorr[4]	17.2	0.2.1	2018	✓
Peak Alignment					
5	mSPA[5]	0.1	1.0.0	2011	✓
6	AMDORAP[6]	40.9	1.0.6	2012	
7	MI-Pack[7]	28.9	1.0.0	2015	
Spectral Deconvolution					
8	Metab[8]	3.2	1.18.0	2019	✓
9	decoMS2[9]	5.2	0.1.0	2013	
10	GAGdecon[10]	0.1	1.0.0	2018	
11	msXpertSuite[11]	0.7	4.1.0	2019	
12	ADAP-GC[12]	3.2	3.0.0	2017	
13	MaxEnt[13]	12.4	3.4.1	2017	

Retention Time Correction					
14	HCor[14]	0.1.0	1.01.0	2014	✓
Data Processing					
Peak Detection					
15	MetMSLine[15]	1.6	1.2.1	2017	
16	X13CMS[16]	0.1	1.4.0	2014	✓
17	proFIA[17]	2.0	1.10.0	2019	
18	cosmiq[18]	17.5	1.18.0	2019	
19	mzMatch-ISO[19]	0.1	1.0.0	2019	
20	PyMS[20]	0.45	1.0.0	2012	
21	TargetSearch[21]	0.69	1.40.3	2019	
22	msPeak[22]	32.3	1.0.0	2013	
Data Normalization					
23	Metabox[23]	53.0	1.2.0	2016	
24	MetNorm[24]	0.5	0.1.0	2015	✓
25	crmn[25]	2.4	0.0.20	2014	
Statistical Analysis					
26	KMMDA[26]	0.5	1.0.0	2018	✓
27	MSPrep[27]	1.0	0.0.2	2018	
28	flagme[28]	22.0	1.40.0	2019	
Quality Control					
29	SECIMTools[29]	0.6	1.0.0	2018	
30	QCScreen[30]	12.1	1.0.0	2018	✓
Metabolite Quantification					
31	MetTailor[31]	0.9	2.0.0	2015	✓
32	MetaQuant[32]	12.0	1.0.0	2010	
33	apLCMS[33]	14.1	6.6.3	2019	
Data Imputation					
34	MINMA[34]	2.6	0.1.0	2017	✓

In-silico Fragmentation					
35	SIRIUS[35]	37.0	4.0.1	2019	✓
36	iontree[36]	0.9	1.23.1	2018	
Metabolite Identification					
37	MetShot[37]	0.9	0.3.2	2018	
38	Molfind[38]	36.2	1.9.0	2013	
39	MIA[39]	2.7	1.0.0	2017	
40	MetaMS[40]	3.7	1.20.0	2019	
41	MSeasy[41]	5.8	5.3.3	2013	
42	RAMClustR[42]	51.6	0.4.1	2019	
43	MetaboSearch[43]	45.7	1.0.0	2012	
44	EI-Maven[44]	91.4	9.0.0	2019	
45	geoRge[45]	13.7	1.0.0	2017	✓
46	eRAH[46]	3.7	1.1.0	2018	
47	MetaboList[47]	0.3	1.4.0	2019	
48	I.MSSpect.[48]	0.2	1.2.0	2018	
49	AssayR[49]	71.6	0.0.9	2017	
50	MS2Analyzer[50]	3.0	2.1.0	2016	
51	Nontarget[51]	3.5	1.9.0	2019	
52	MetMask[52]	4.43	0.5.3	2017	
53	peakANOVA[53]	0.5	1.0.0	2015	
54	PUTMEDID[54]	191.2	1.0.0	2011	
Data Clustering Analysis					
55	SimExTargid[55]	49.8	0.2.1	2017	
56	MetaboAnalyst[56]	49.3	4.0.0	2019	
57	CAMERA[57]	2.1	1.40.0	2019	
58	KPIC2[58]	15.7	2.4.0	2019	✓

Data Network Analysis					
mGWAS					
59	MWASTools[59]	56.3	1.8.0	2019	✓
Metabolite Mapping					
60	InCroMAP[60]	54.6	1.5.0	2012	✓
61	PathVisio[61]	17.6	3.3.0	2019	
62	Mapping Tool[62]	15.0	1.3.0	2013	
63	ChemDistiller[63]	133.2	0.1.0	2018	
Metabolic Network Analysis					
64	PySCeSToolbox [64]	5.1	0.9.6	2018	
65	MetaDiff[65]	6.0	1.0.0	2016	
66	ReactomePA[66]	11.6	1.28.0	2019	
67	MEBS[67]	73.2	1.0.0	2017	
68	RxnSim[68]	21.6	1.0.3	2013	
69	phraSED-ML[69]	1.8	1.0.3	2018	
70	ScrumPy[70]	1093.9	1.0.0	2018	
71	Subpathway[71]	1.7	3.0.0	2013	
72	Kamneva 2016[72]	101.8	1.0.0	2016	
73	MetaboSignal[73]	169.1	1.14.0	2019	
74	ReactPRED[74]	52.4	1.0.0	2016	
75	Mminte[75]	52.2	1.0.0	2017	
76	PAPi[76]	0.6	1.24.0	2019	✓
77	MetaNetSam[77]	3.8	1.1.0	2015	
78	Fbar[78]	2.1	0.5.2	2018	
79	SED-ED[79]	6.4	2.2.3	2016	
80	QSSPN[80]	2.1	1.0.0	2015	
81	JMassBalance[81]	3.7	1.0.0	2013	
82	Pyabolism[82]	0.4	1.0.0	2017	

83	Pybrn[83]	0.6	0.4.3	2016	
84	MonaLisa[84]	17.8	5.1.0	2016	
85	MoDentify[85]	0.6	0.99.0	2019	
86	C.Calculator[86]	16.4	1.0.0	2010	
87	KEGGREST[87]	13.67	1.24.0	2019	
Data Integration Analysis					
88	integrOmics[88]	0.1	2.55.0	2012	✓
89	MetScape[89]	17.2	3.1.3	2017	
90	cPath[90]	1.32	2.0.0	2019	
91	MetDisease[91]	15.2	1.1.0	2014	
Data Enrichment Analysis					
92	zeroSum[92]	1.0	2.0.0	2019	✓
93	FELLA[93]	3.1	1.4.1	2019	
94	M.IDConvertor[94]	0.2	1.0.0	2010	
Data Visualization					
95	MZMine[95]	148.7	2.0.0	2019	
96	XCMS[96]	3.5	3.7.1	2018	
97	yamss[97]	15.5	1.9.1	2018	
98	R2DGC[98]	0.6	1.0.3	2017	
99	BatMass[99]	0.1	0.3.0	2018	✓
100	MAIT[100]	36.2	1.18.0	2019	
101	ProbMetab[101]	0.2	1.0.0	2013	
102	MetCirc[102]	5.0	1.14.0	2017	
103	xMSAnalyzer[103]	38.9	2.0.6	2019	

Software Installation Process

Installing MetumpX software

For downloading of MetumpX, visit its website:

<https://github.com/hasaniqbal777/MetumpX-bin>

1. MetumpX can also be downloaded from the following command through git:

```
$ git clone https://github.com/hasaniqbal777/MetumpX-bin
```

2. Now run the following command on terminal:

```
$ chmod +x MetumpX_setup_enUS  
$ sudo ./MetumpX_setup_enUS
```

3. Installation wizard of MetumpX will start

NOTE *Installation of aXonica require a proper internet connection to proceed, otherwise the installation terminates.*

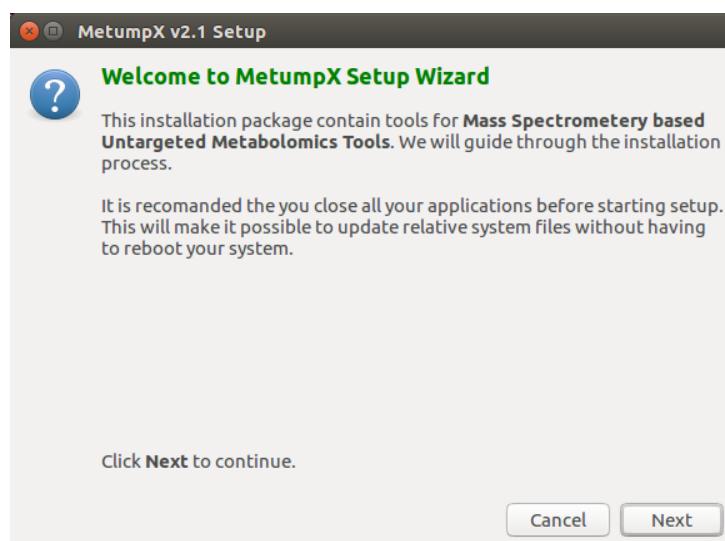


Figure 1: Installation Welcome Screen

4. Click **Next** to proceed and confirm the internet connection.

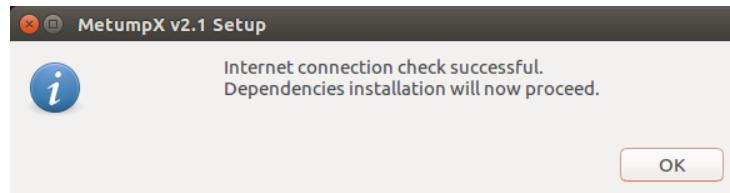


Figure 2: Internet Check Successful Screen

5. Installation is **terminated** if there is no internet.

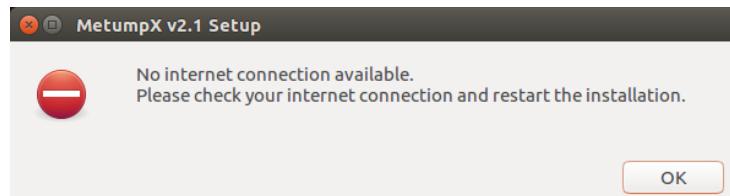


Figure 3: Internet Check failed screen

6. **Dependencies** related to software will start installing.

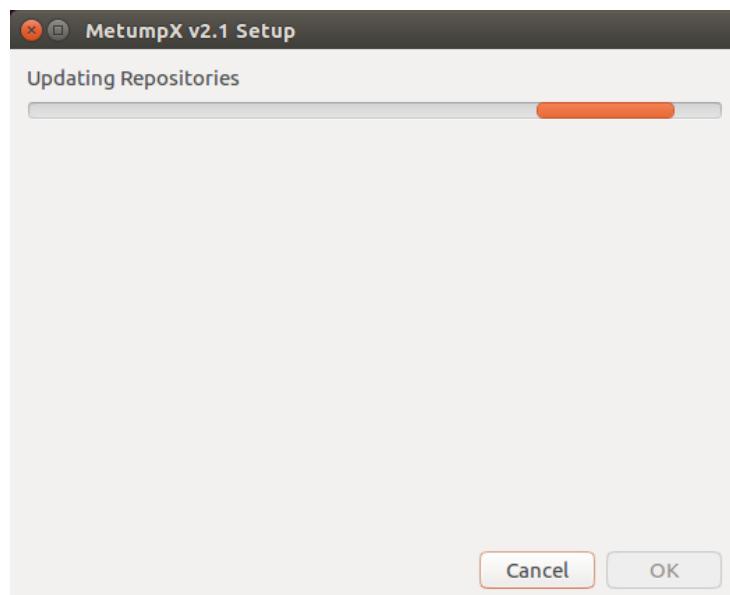


Figure 4: Installation Screen

7. All the dependencies which are installed are displayed at the end of the installation. Click **Ok**.

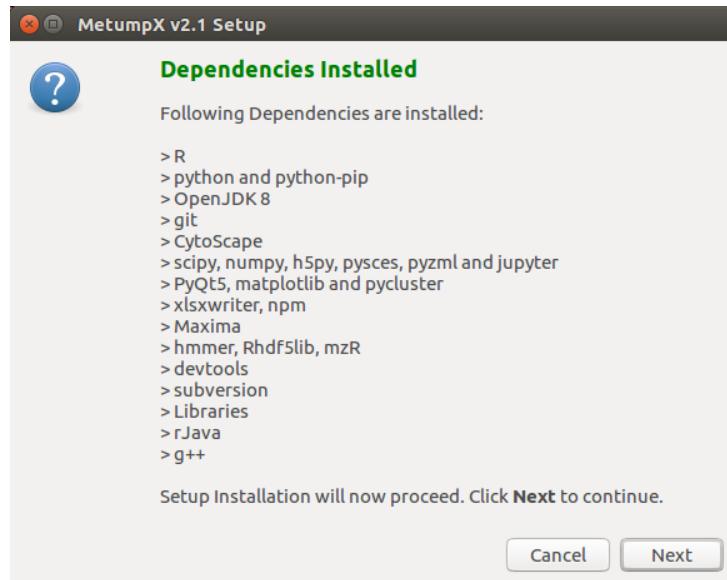


Figure 5: Dependencies installation finished screen

8. MetumpX pipeline detail screen is displayed. It has the information about which software you want to install. Click Next to Proceed.

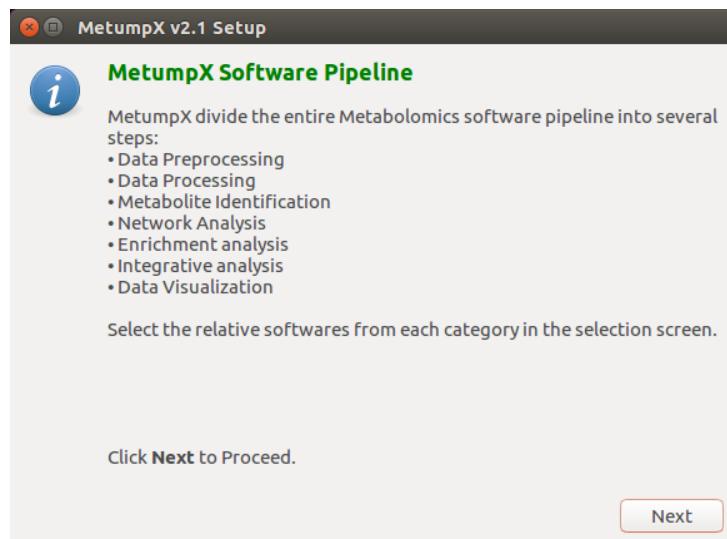


Figure 6: Pipeline information screen

9. Selection screen for Pre-processing Tools is displayed. Select the required tools and Click **Next**.

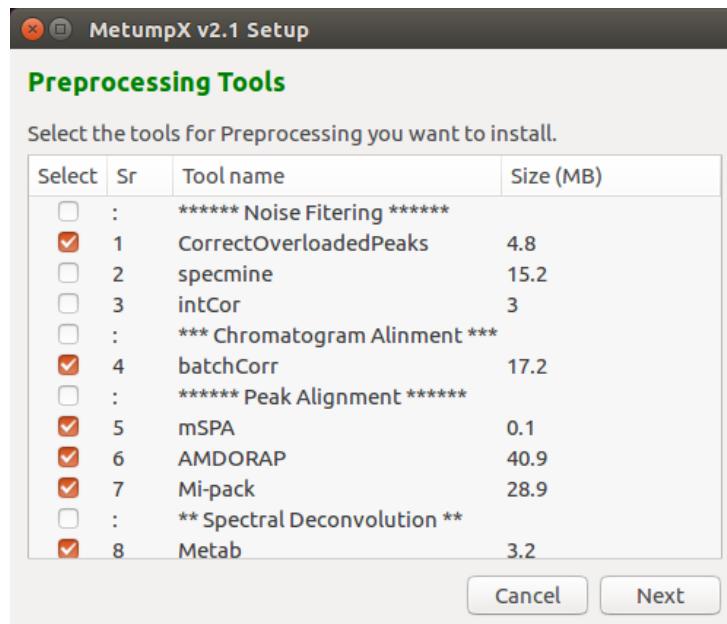


Figure 7: Preprocessing tools selection screen

10. Selection screen for Processing Tools is displayed. Select the required tools and Click **Next**.

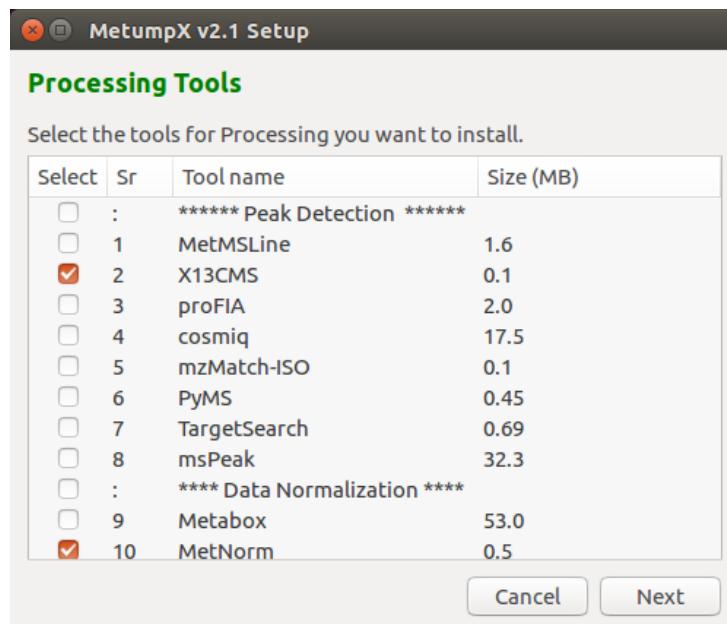


Figure 8: Processing tools selection screen

11. Selection screen for Metabolite Identification tools is displayed. Select the required tools and Click **Next**.

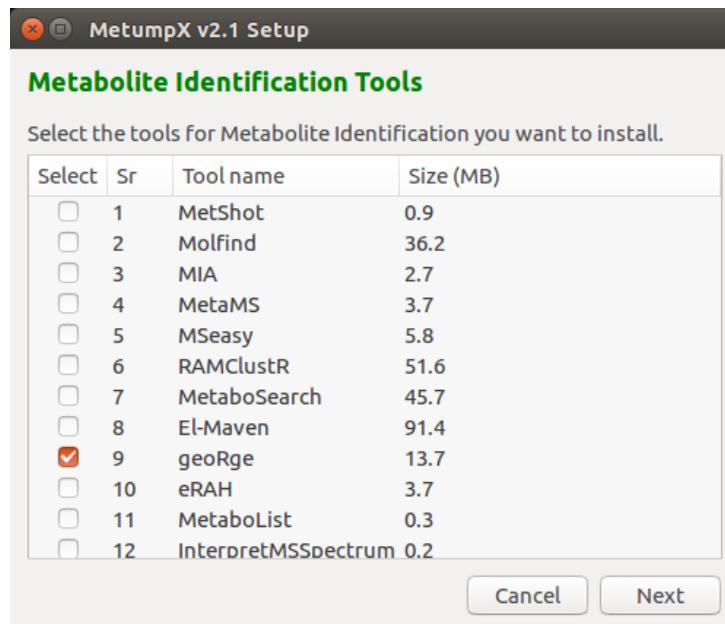


Figure 9: Metabolite identification tools selection screen

12. Selection screen for Clustering Analysis tools is displayed. Select the required tools and Click **Next**.

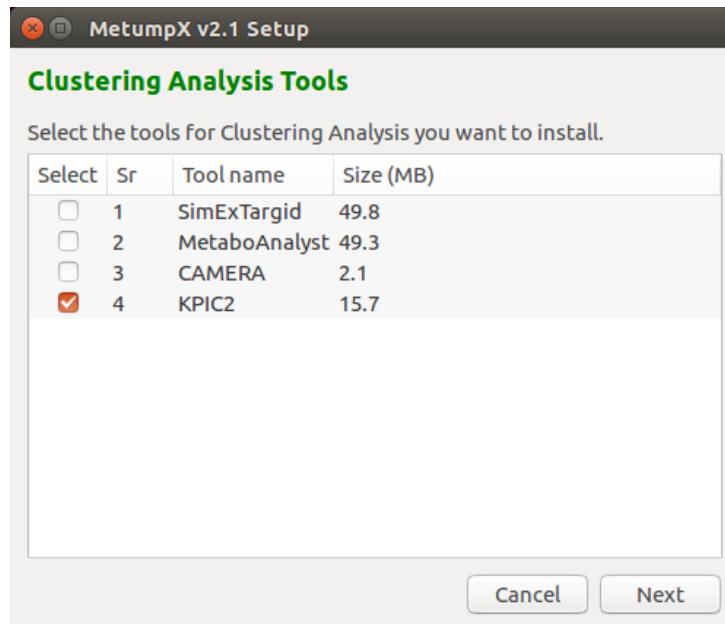


Figure 10: Clustering analysis tools selection screen

13. Selection screen for Data Network Analysis tools is displayed. Select the required tools and Click **Next**.

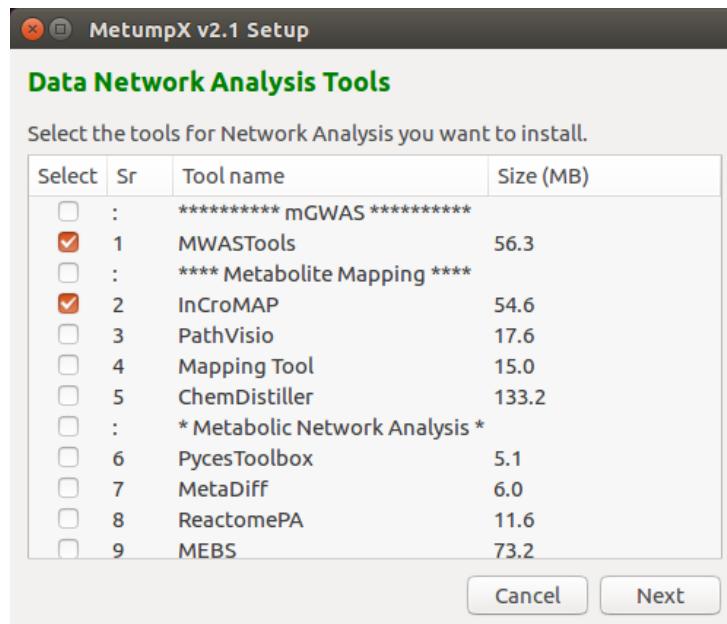


Figure 11: Data network analysis tools selection screen

14. Selection screen for Enrichment analysis tools is displayed. Select the required tools and Click **Next**.

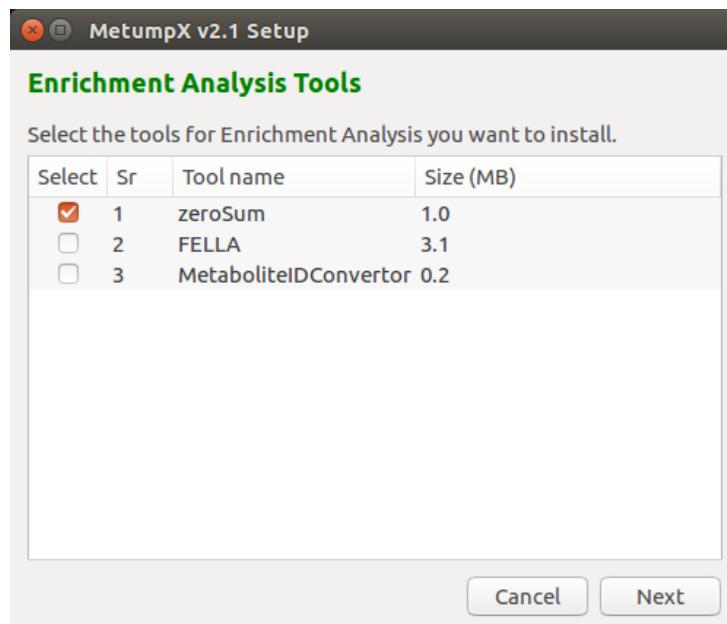


Figure 12: Enrichment analysis tools selection screen

15. Selection screen for Integrative analysis tools is displayed. Select the required tools and Click **Next**.

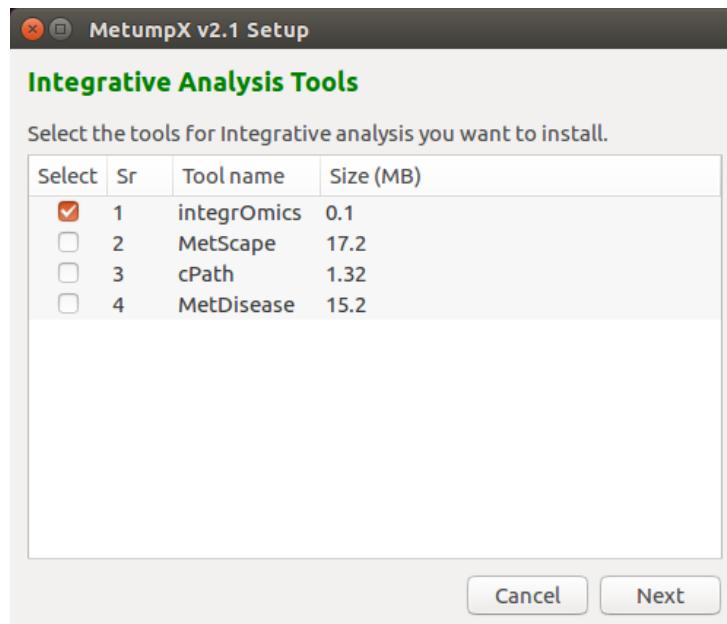


Figure 13: Integrative analysis tools selection screen

16. Selection screen for Data visualization tools is displayed. Select the required tools and Click **Next**.

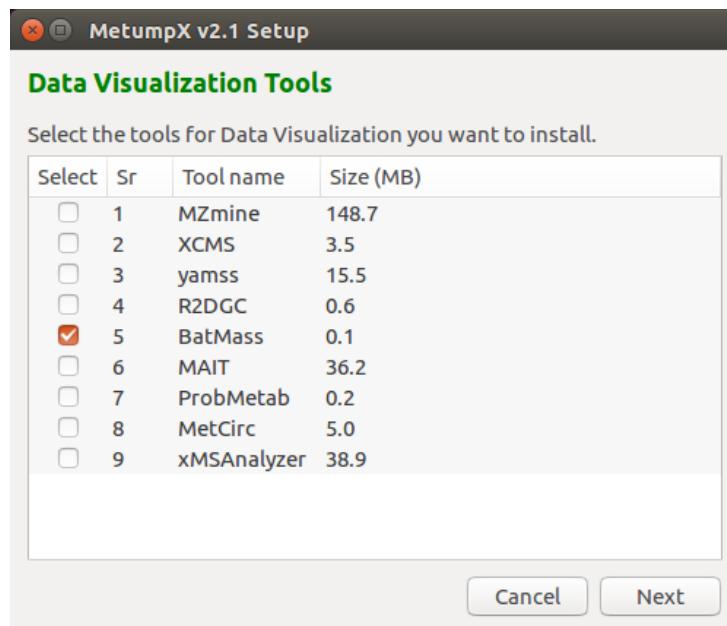


Figure 14: Data visualization tools selection screen

17. Click **Next** to proceed with the installation.

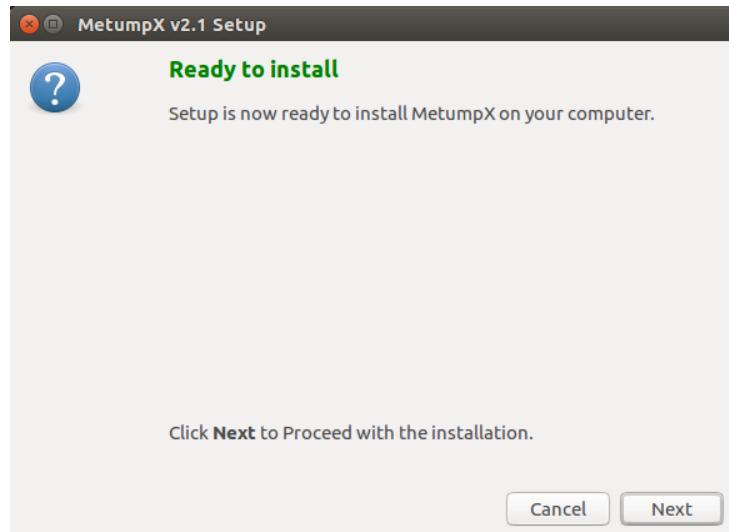


Figure 15: Ready to Install Screen

18. Tool installation will **continue**. Some tools **install** as standalone installation and will be called automatically.

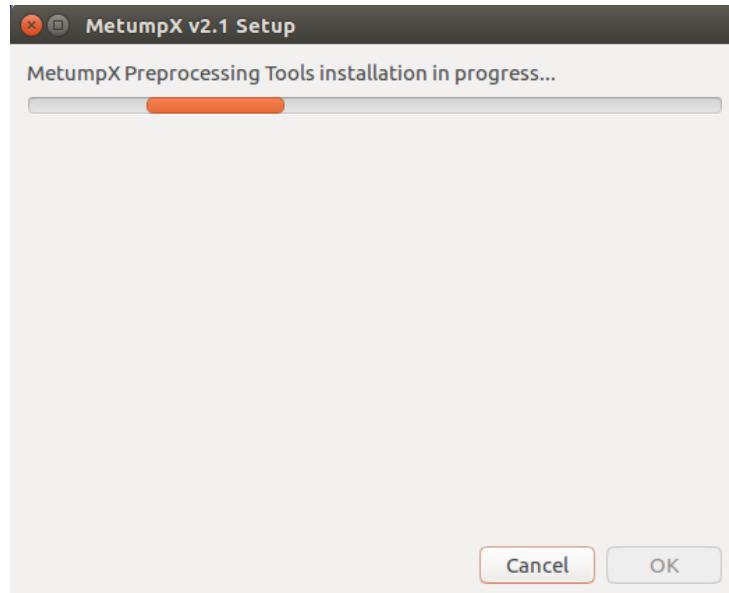


Figure 16: Tools installation progress screen

19. Installation of MetumpX is now finished. Click **Finish** to use the tools.

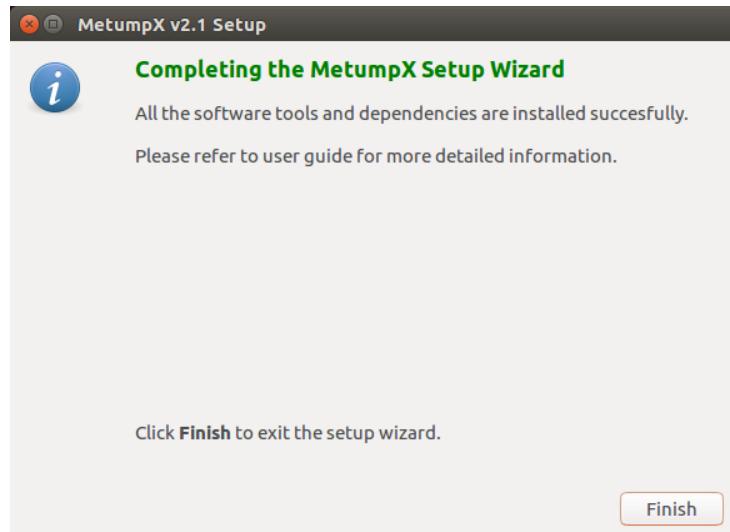


Figure 17: Finishing Installation Screen

Microsoft Windows Support

Installing MetumpX software on Microsoft Windows

MetumpX is software package for Linux, specifically Ubuntu users, but Windows users can use aXonica using a Virtual Machine. Follow these steps to initialize a Virtual Machine on your Windows host.

NOTE MetumpX support a 64bit Windows host. Microsoft Windows XP and Vista support is discontinued.

1. First you have to **download** the image file of the required OS (Ubuntu) from its website or use the following link:

<http://repo.isra.edu.pk/ubuntu-release/18.04.2/ubuntu-18.04.2-desktop-amd64.iso>

2. **Download** the virtualization software (Oracle VM VirtualBox) from the following link:

<https://download.virtualbox.org/virtualbox/6.0.10/VirtualBox-6.0.10-132072-Win.exe>

3. **Install** this software in Windows OS host.
4. **Start** Virtual Box, and click on the **New** symbol.

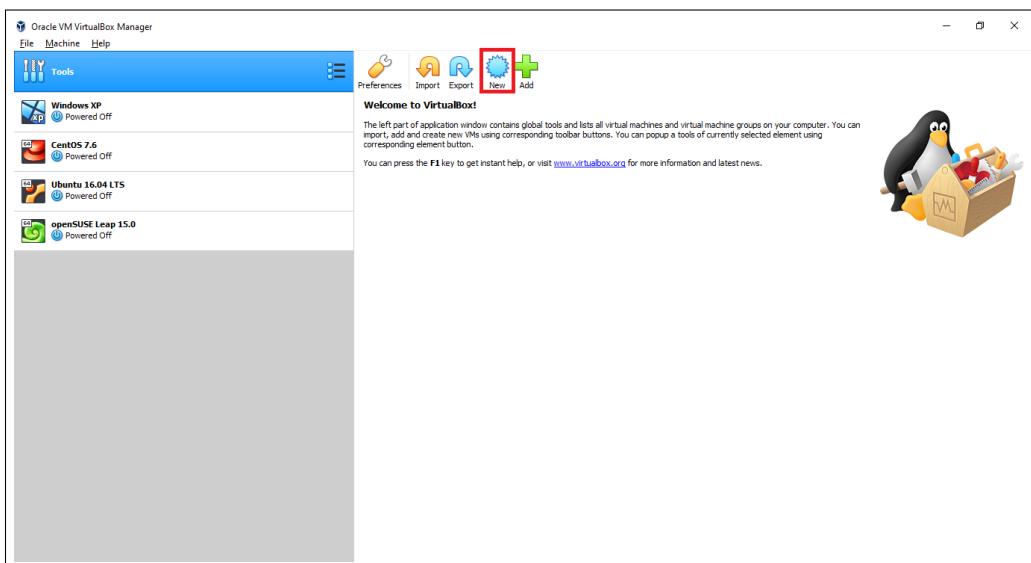


Figure 1: New Virtual OS

5. Give the virtual OS a relevant **Name**. Select the **Type** (Linux) and **Version** (Ubuntu 64-bit) and Click **Next**.

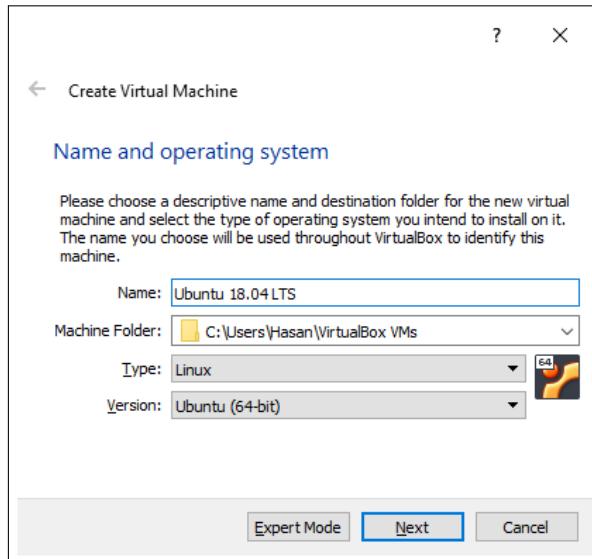


Figure 2: Assigning Name

6. Allocate RAM to the virtual OS. Following system has 8GB of RAM so 2GB of RAM is allocated. You can use more RAM if your system has enough extra RAM.

NOTE Allocate about half of the RAM to the virtual OS. Click Next.

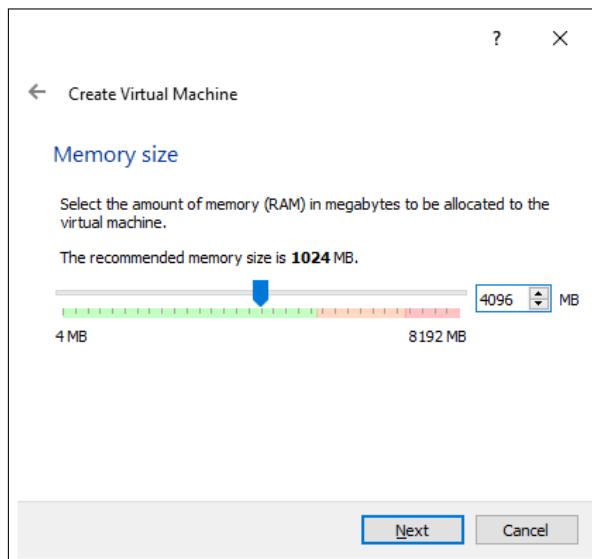


Figure 3: Memory Allocation

7. **Create** a virtual disk. This works as the hard disk of the virtual Linux system. This is where the virtual system will store its files. Click **Create**.

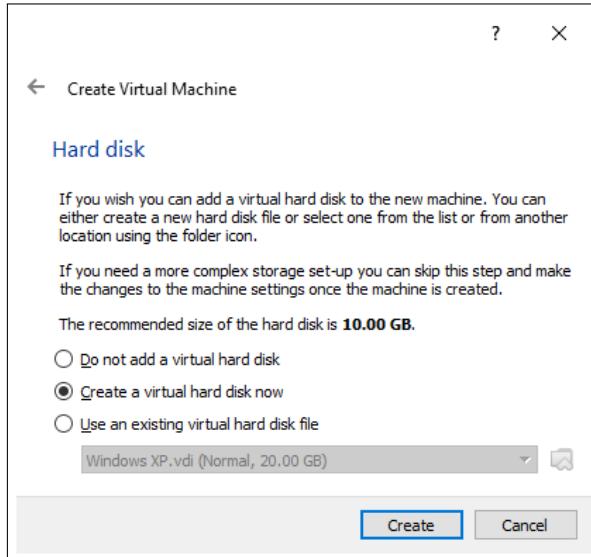


Figure 4: Creating a virtual hard disk

8. Select **VDI file type** here (recommended). Click **Next**.

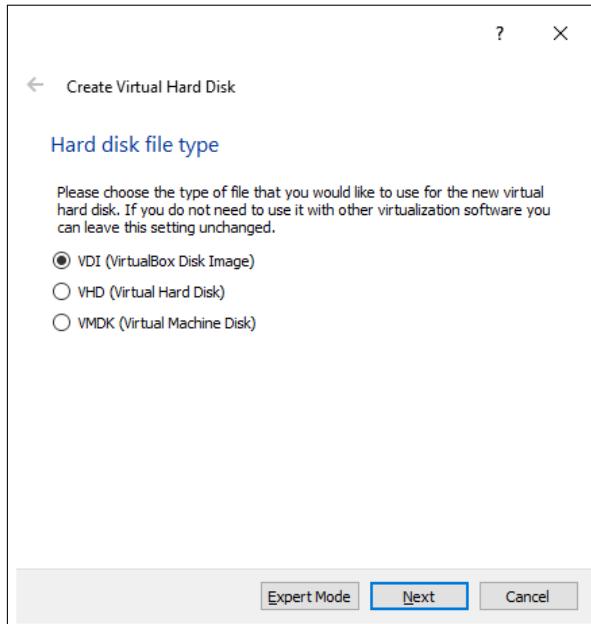


Figure 5: Hard disk file type

9. You can choose either of Dynamically allocated or Fixed size option for creating the virtual hard disk. Choose **Dynamically allocated**. (recommended). Click **Next**.

NOTE *Dynamic allocation is allocated as time passes and data is increased whereas fixed is allocated instantly.*

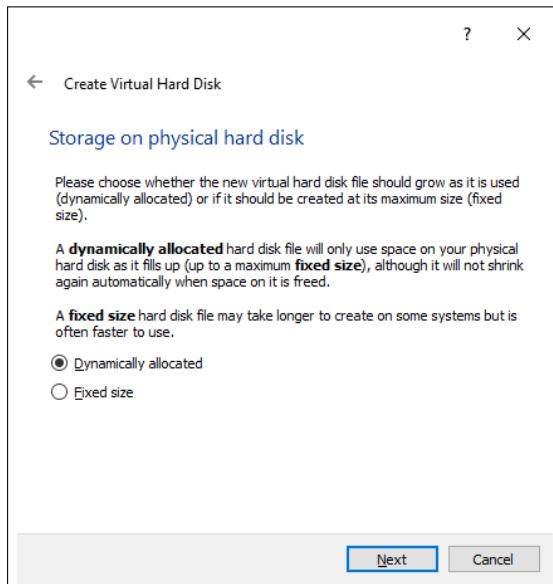


Figure 6: Storage type

10. Select **VDI file type** here (recommended) and Select the **Hard Disk size**. (recommended size: 100 GB). Click **Create**.

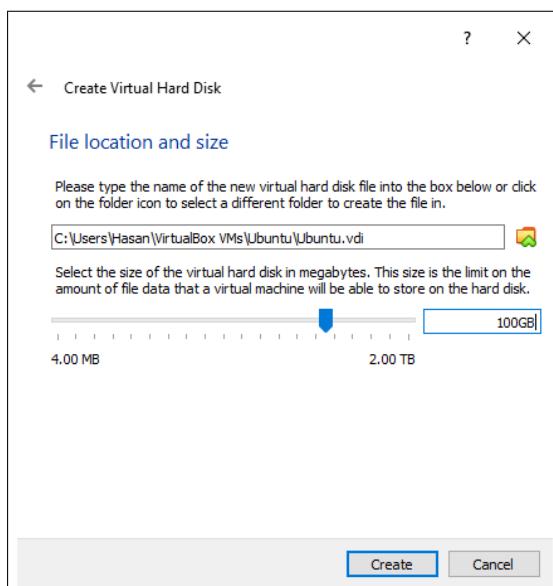


Figure 7: Hard disk size

11. Click **Next**. Now, Select **Settings** to assign the image file of respective OS to VB.

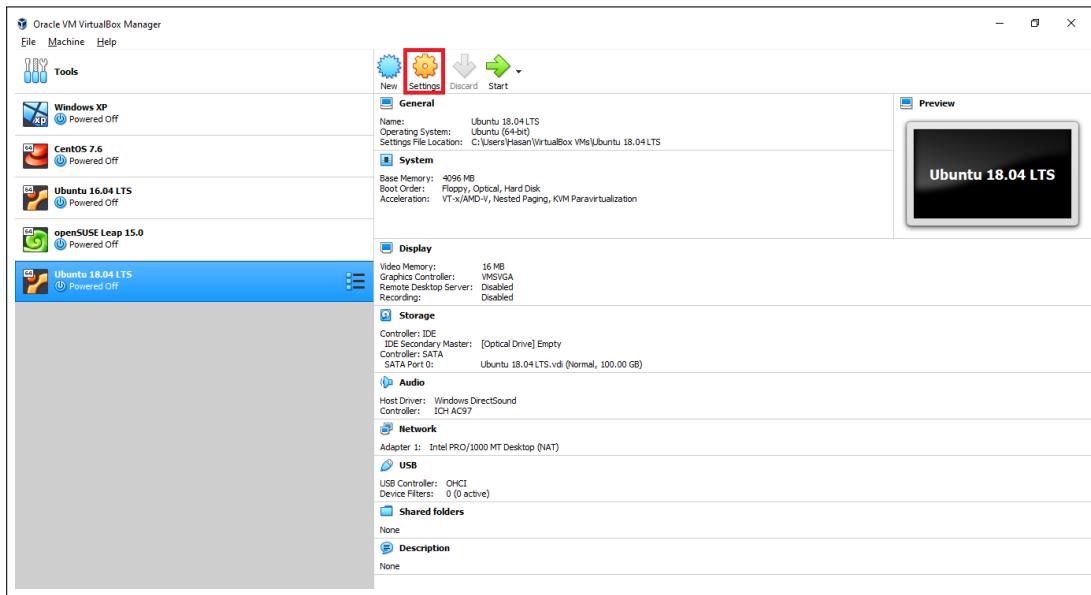


Figure 8: Select Settings

12. Select **General → Advanced**. Now, select the **Shared Clipboard** and Drag'n'Drop option to **Bidirectional**.

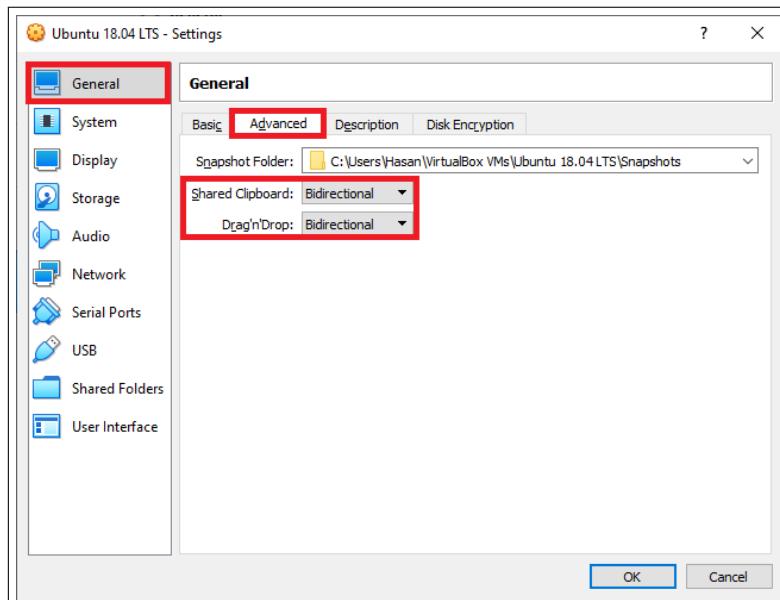


Figure 9: Advanced Settings

13. Select **Settings** to assign the image file of respective OS to VB. Select **Storage** → **Controller : IDE** → **Empty**. Now, in the **Attributes** tab, click on **New Disk** and provide the path of downloaded image file of Ubuntu OS. Click **OK**.

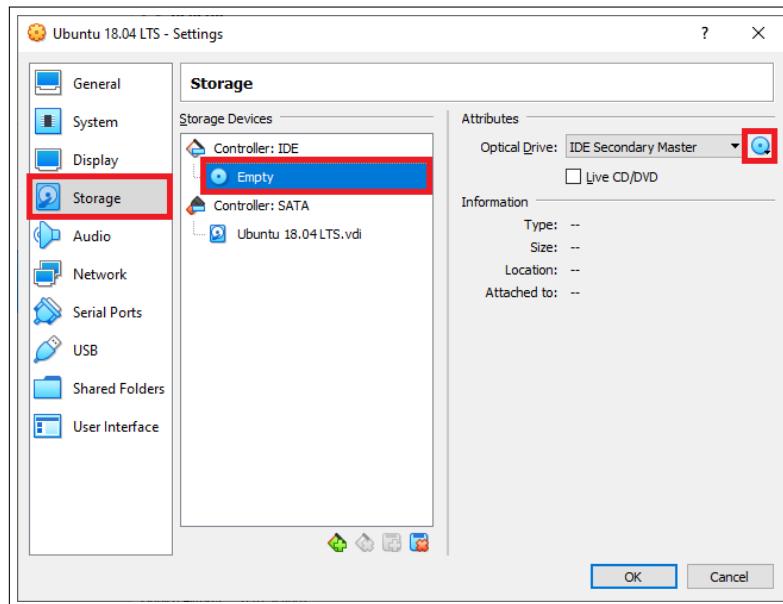


Figure 10: Providing Image file

14. Once everything is in place, it's time to boot that ISO and install Linux as a virtual operating system. Click Start.

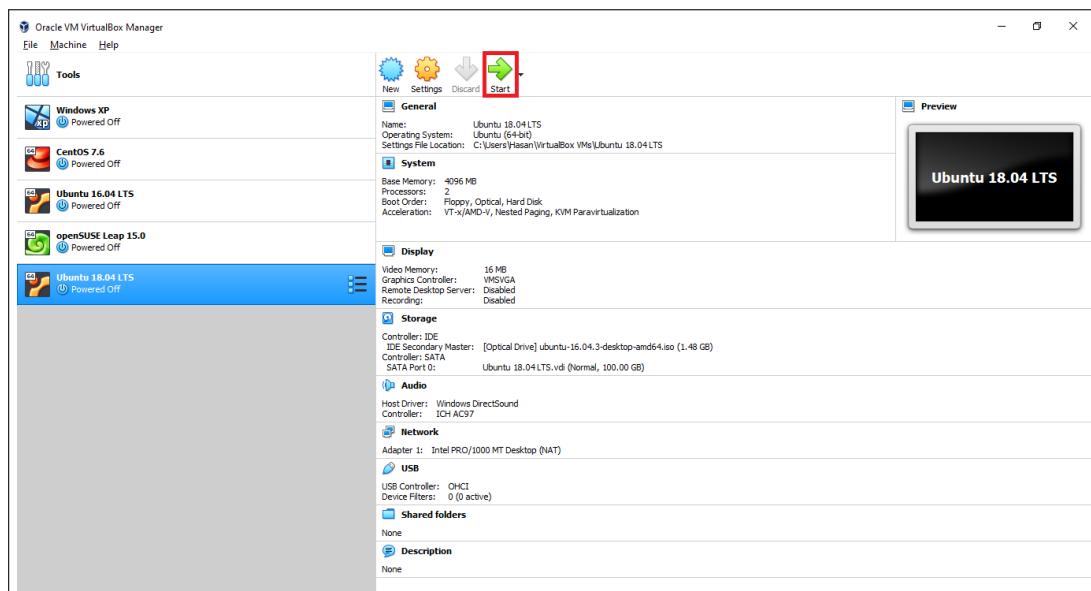


Figure 11: Starting Virtual OS

15. Virtual OS will boot into Linux Installation process. You should be presented with the option to install it. Click **Install Ubuntu**.

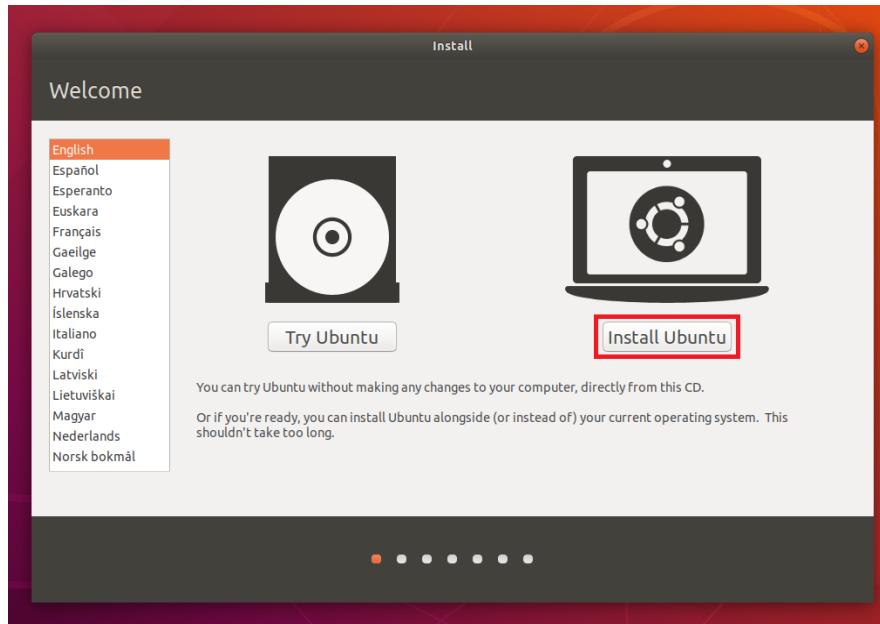


Figure 12: Installing Ubuntu

16. Continue with Normal Installation.

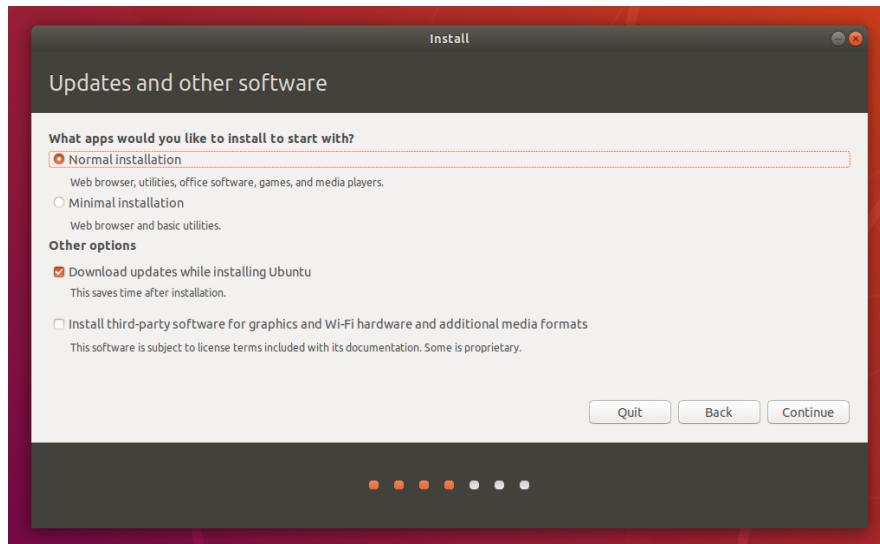


Figure 13: Update Screen

17. In Installation type screen, select **Erase disk** and **Install Now** option.

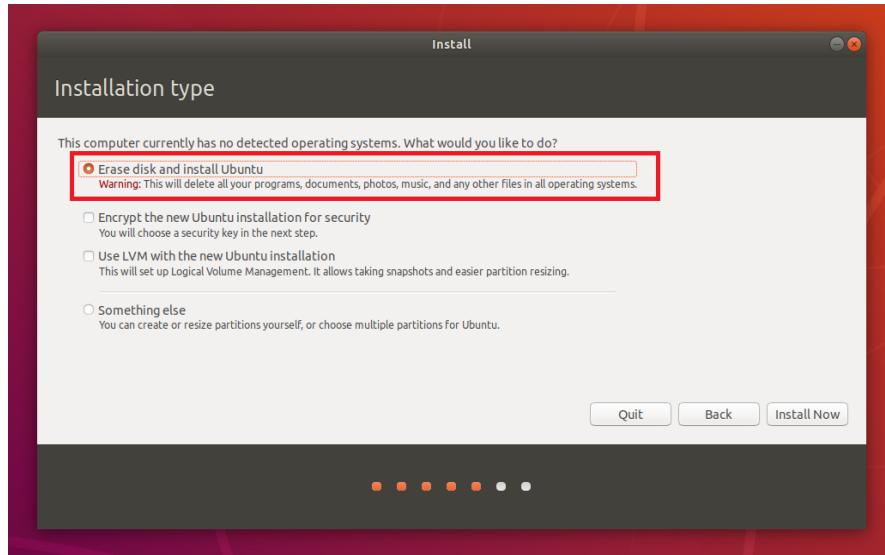


Figure 14: Installation type Screen

18. Select Continue.

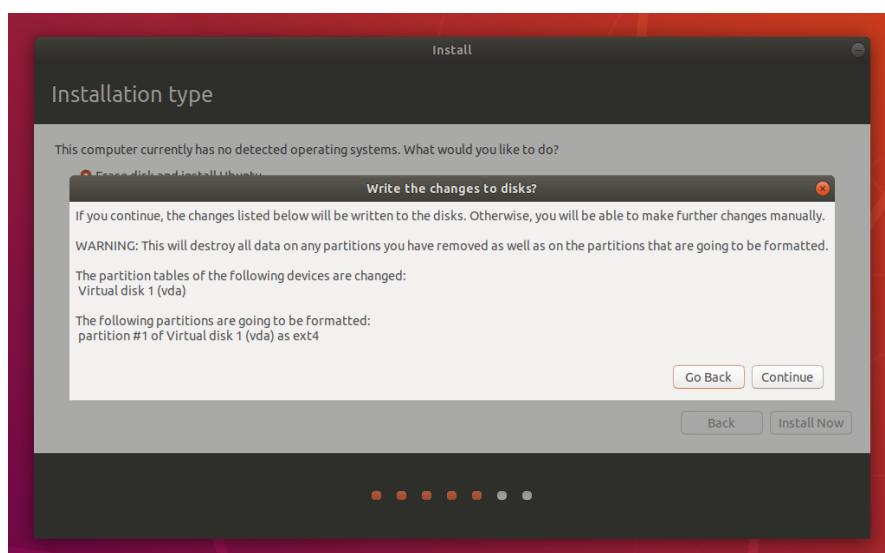


Figure 15: Confirmation Screen

19. Select your **Current Location** and Continue. Select **Continue**.

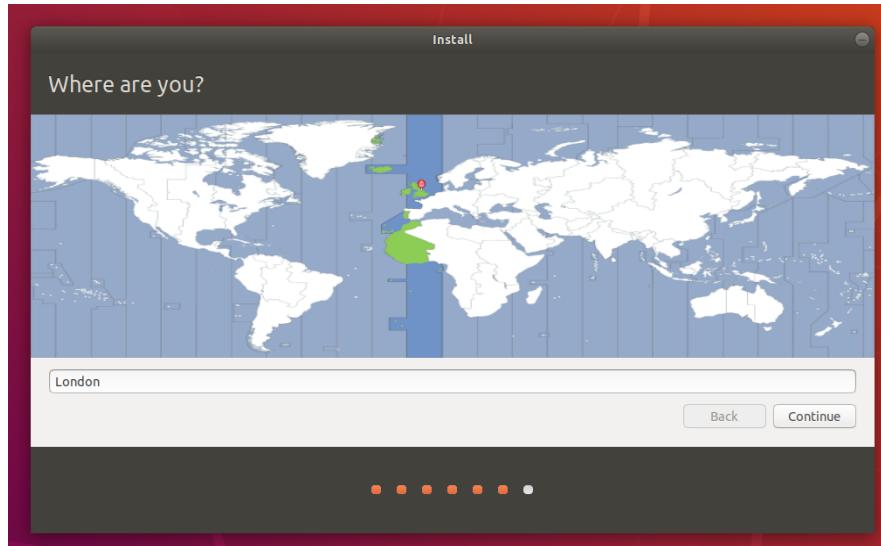


Figure 16: Location selection screen

20. Fill your Info and click **Continue**.

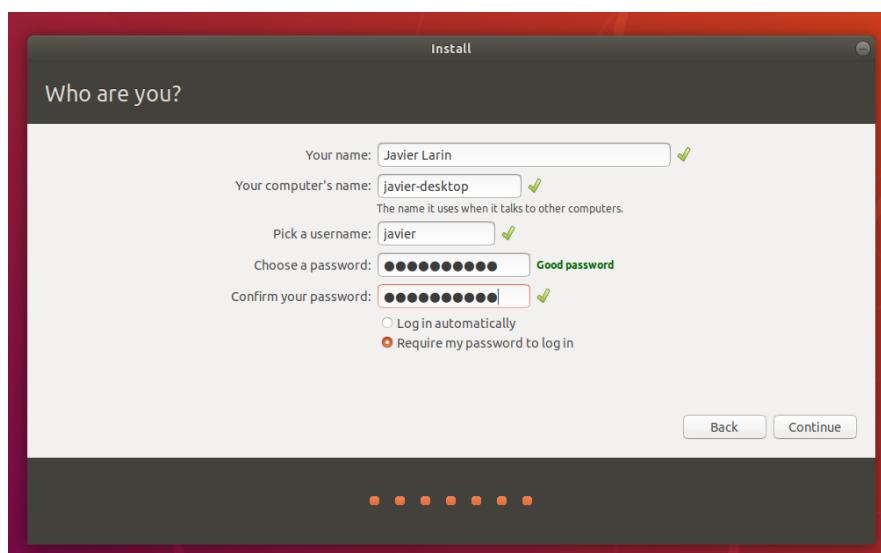


Figure 17: Intro Screen

21. Installation will Continue.

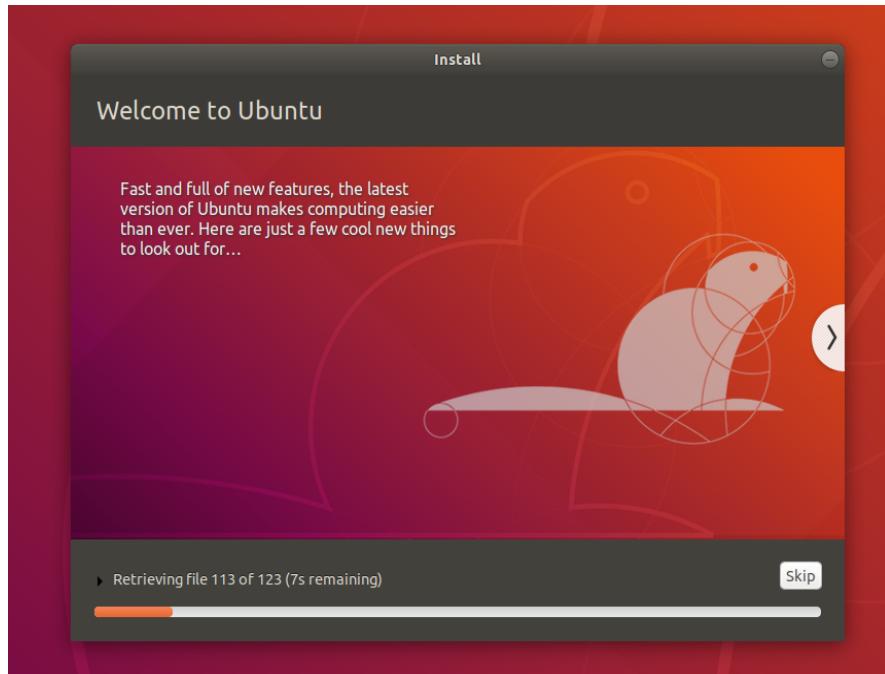


Figure 18: Installation Screen

22. Installation is Complete. Click **Restart Now**.

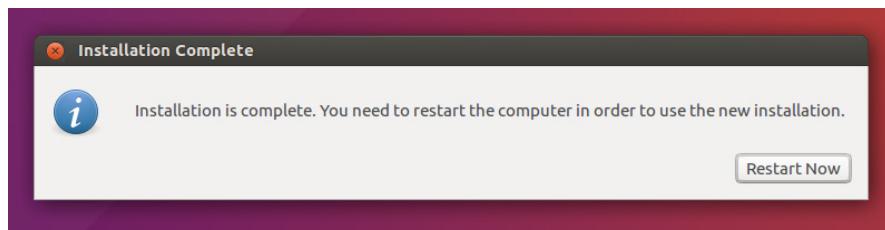


Figure 19: Complete Installation Screen

MetumpX can be installed in the Virtual Ubuntu OS normally as described in the previous chapter.

Introduction to Metabolomics

In this Chapter

We will learn about:

- Basics about Metabolomics
- Types of Metabolomics

Introduction

Metabolomics is the comprehensive (qualitative and quantitative) study of all the small molecules in an organism called **metabolites**. Collectively, these small molecules and their interactions within a biological system are known as the **metabolome**.

These molecules are smaller than 1500 daltons (Da). Some important metabolites are peptides, oligonucleotides, sugars organic acids ketones, aldehydes, amines and amino acids.

Interactome

Recent genetic studies have made major contributions to understanding disease processes. However, biological processes operate through complex interactions between genes, RNA, proteins, and metabolites—the composite of this complex interaction network is defined as the interactome.[104]

Metabolomics is a powerful approach because metabolites and their concentrations, unlike other omics (shown in the following Figure 1) measures, directly reflect the underlying biochemical activity and state of cells / tissues. Thus metabolomics best represents the molecular **phenotype**.

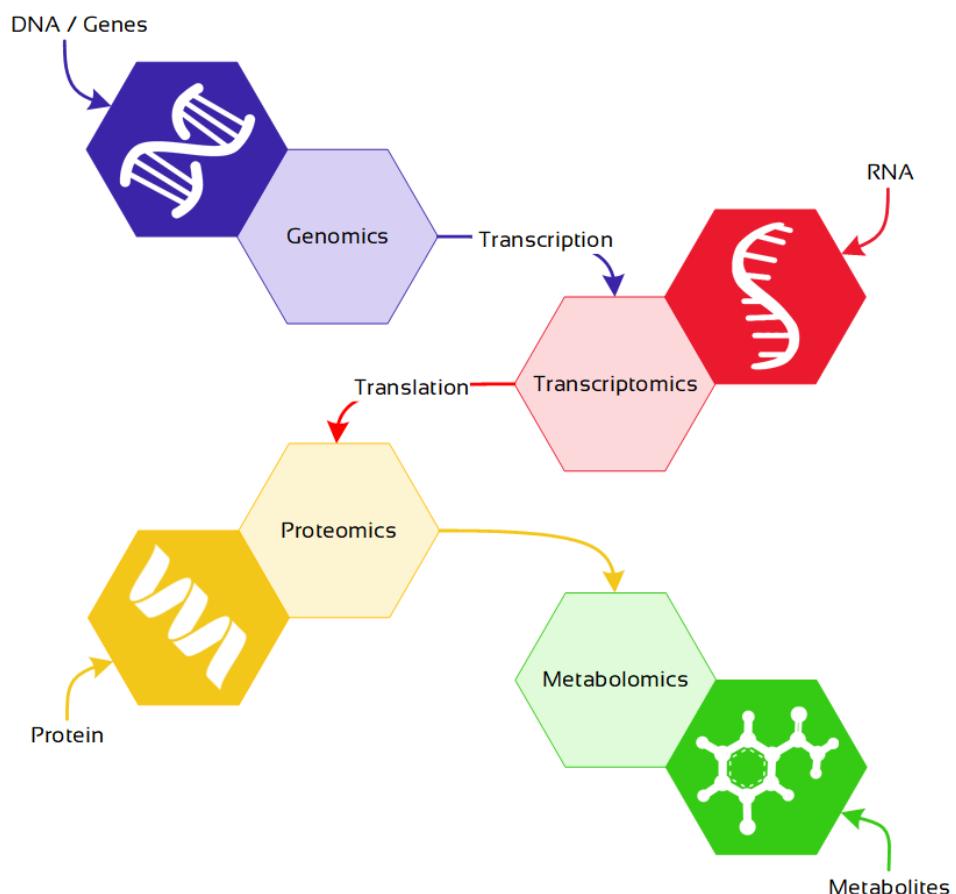


Figure 1: An overview of the four major “omics” fields, from genomics to metabolomics.

Metabolomics is used as a **complementary approach** to genomics, transcriptomics, and proteomics, and is the most recently developed technique.

Metabolomic Study Strategies

A range of analytical techniques are used to analyze metabolites in different organisms, tissues or fluids. There are **two** analytical chemistry strategies applied in metabolomics to analyze metabolites.

1. Untargeted Metabolomics
2. Targeted Metabolomics

The objectives of a study define which of the three analytical strategies are applied. **Untargeted** approach is applied in hypothesis-generating studies i.e. **discovery**, while Targeted approach is usually applied to **validate and translate** the novel discoveries of a hypothesis-generating study.

The major differences between untargeted and targeted studies are:

- the level of sample preparation required,
- the number of metabolites detected, and
- the level of quantification of the metabolites.

Both the strategies are briefly explained below.

Untargeted Metabolomics

It involves measurement of as many metabolites as possible from a biological sample to classify phenotypes based on metabolite pattern. It is also known as Metabolic fingerprinting.

Targeted Metabolomics

It involves measurement of metabolites of a focused group from a biological sample. It is also known as Metabolic profiling.

NOTE *MetumpX is focused on the study strategy of Untargeted Metabolomics.*

Metabolomics Workflow

In this Chapter

We will learn about:

- Raw Data Acquisition
 - Untargeted Metabolomics Workflow
 - Introduction to Software Pipeline
-

Introduction

Metabolomics is an emerging technology and with increasing demand for personalized medicines, the metabolomics market is growing along with massive investments. Honorable mentions contain Bayer CropScience, Genomatica, MDXHEALTH and Metabolon.

A standard set of procedures is followed in order to perform metabolomic analysis and this set of steps is described in this chapter.

Untargeted Metabolomics Workflow

Metabolomics Workflow can be divided into many steps as shown in [Figure 1](#). Details of each step is as follows:

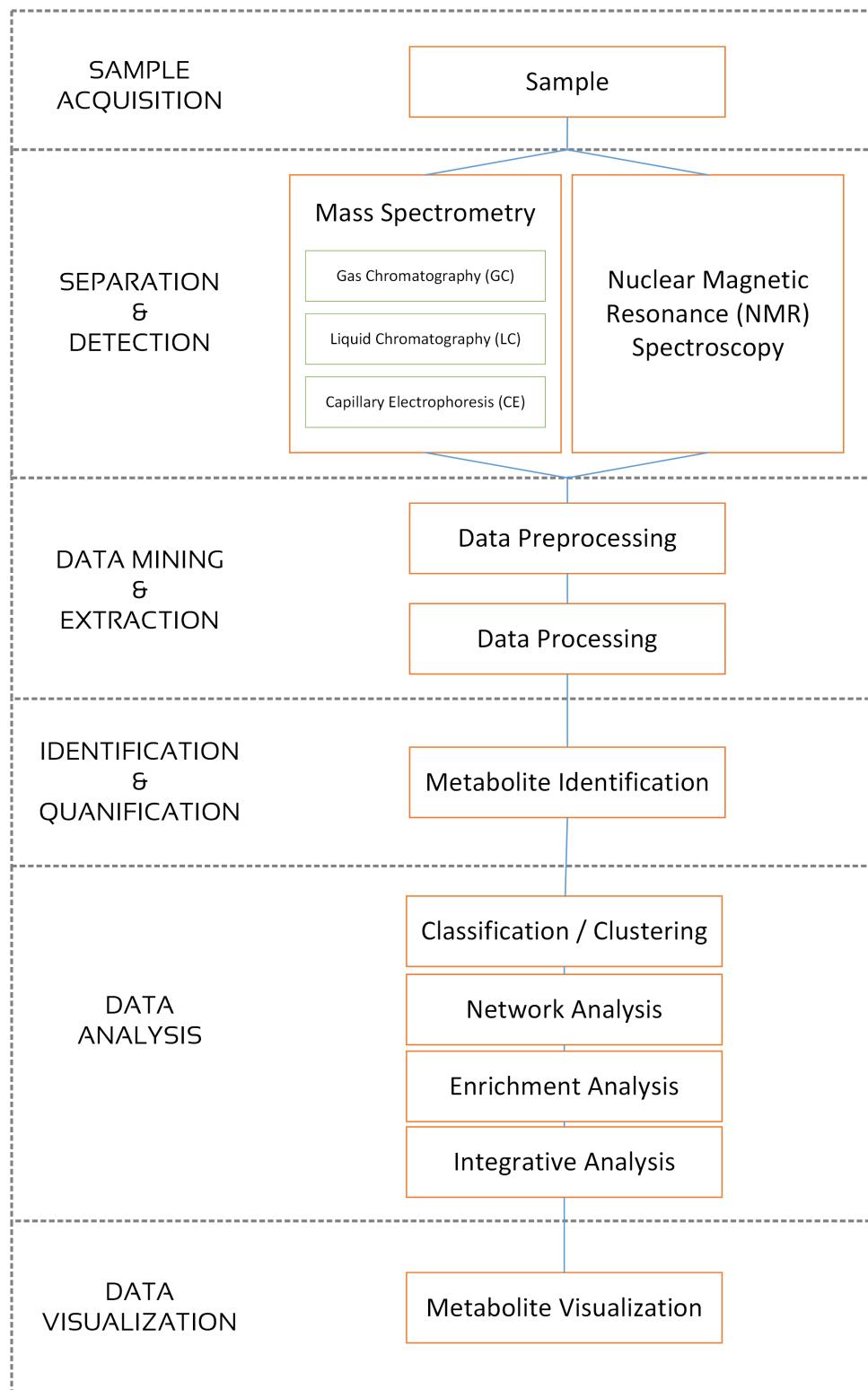


Figure 1: Metabolomics Analytical Workflow

Sample Acquisition

Metabolites are measured in different samples such as tissue, biofluids (blood, urine, feces, seminal fluid, saliva, bile, cerebrospinal fluid) and cell cultures.

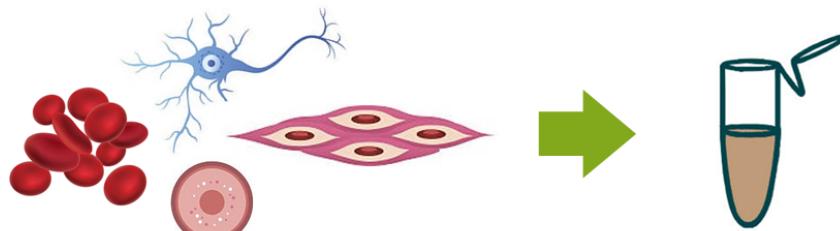


Figure 2: Sample Acquisition from Cells, Tissues and Fluids

Sample Preparation

Sample Preparation usually involve the following steps shown in detail in [Figure 3](#):

- storage
- extraction
- separation
- detection

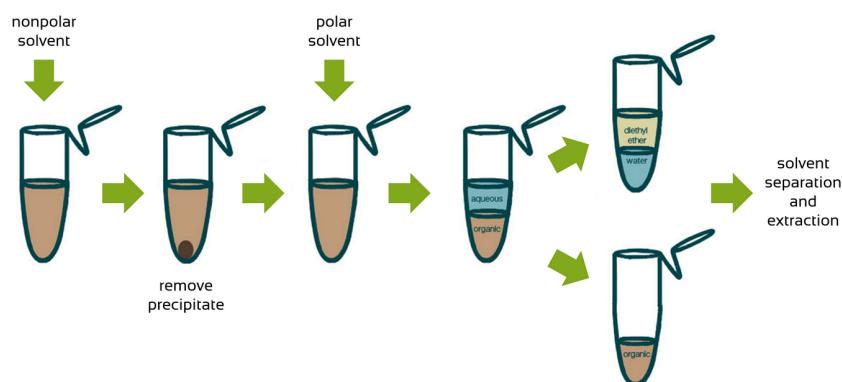


Figure 3: Main Steps involved in sample preparation

The two main analytical methods for separation and detection are:

- Nuclear Magnetic Resonance (NMR) Spectroscopy
- Mass Spectrometry (MS)

NOTE MetumpX is focused on the **Mass Spectrometry** based separation and detection.

Nuclear Magnetic Resonance (NMR) Spectroscopy

In NMR spectroscopy, biological sample is placed in a magnetic field. Isotopes in the sample absorb the radiation and resonate at frequencies relative to the size of the molecules. It is mainly used for a mixture or organic and inorganic compounds.

The resultant spectrum shown in [Figure 4](#) is a collection of peaks at different positions and intensities and each sample has a unique pattern.

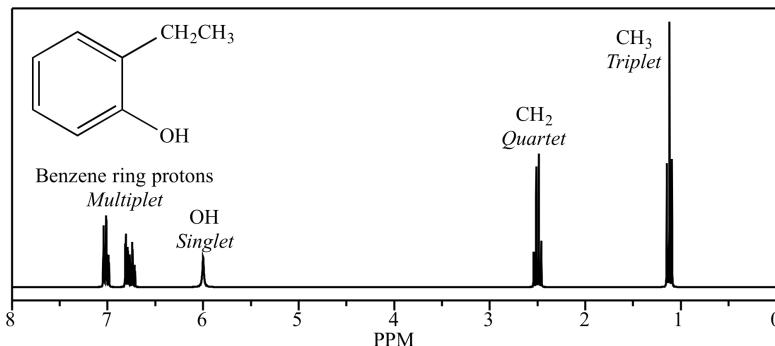


Figure 4: NMR spectrum for 2-ethyl-phenol

Mass Spectrometry (MS)

Mass Spectrometry can be used to analyze biological samples by ionizing the sample and then sorting the ions according to their mass-to-charge (m/z) ratio. Mass Spectrometer function is shown in [Figure 5](#).

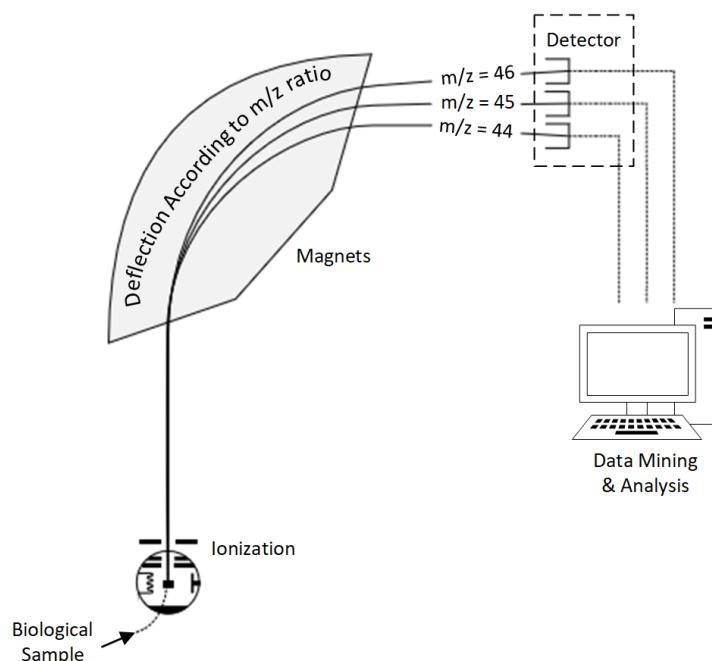


Figure 5: Standard Mass Spectrometer

Significant methods for ionization of sample contain Chemical Ionization (CI), Electron Impact Ionization (EI) and Electrospray Ionization (ESI). A resulting MS Spectrum is shown in Figure 6.

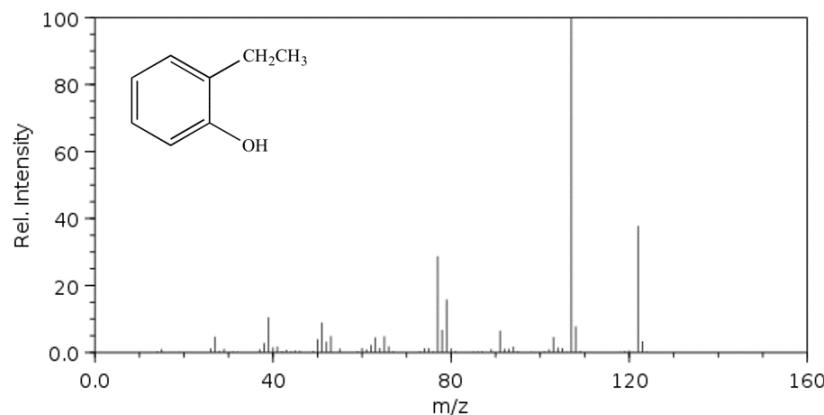


Figure 6: MS spectrum for 2-ethyl-phenol

An important enhancement to the mass resolving and mass determining capabilities of mass spectrometry is using it in tandem with chromatographic and other separation techniques. Some major separation techniques are:

- Gas Chromatography (GC)
- Liquid Chromatography (LC)
- Capillary Electrophoresis (CE)

These techniques are coupled with Mass Spectrometry and generally referred as GC-MS, LC-MS and CE-MS respectively. A brief explanation is described ahead.

Gas Chromatography Mass Spectrometry (GC-MS)

GC-MS is commonly used in metabolomics for measurement of volatile compounds such as fatty acids and organic acids. Sample need to be volatile and thermally stable as separation in GC occurs in an oven at high temperatures. Metabolite loss is one of the major drawbacks of GC/MS. Characteristic spectral patterns and extensive libraries are available online for GC-MS. A Gas Chromatograph is shown in Figure 7.

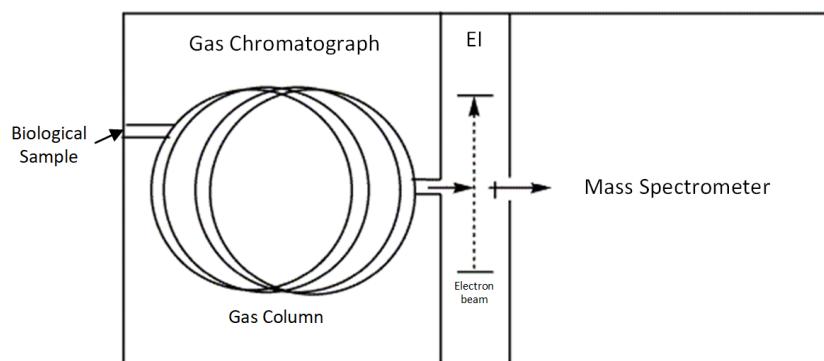


Figure 7: Gas Chromatograph Mass Spectrometer

Liquid Chromatography Mass Spectrometry (LC-MS)

Liquid chromatography (LC) and High-Performance Liquid Chromatography (HPLC) is a technique that has high resolution and analytical flexibility. It can be used for the analysis of a specific metabolite or class of compounds. LC-MS has one advantage over GC-MS that there is no need for chemical derivatization of metabolites. A Liquid Chromatograph is shown in [Figure 8](#).

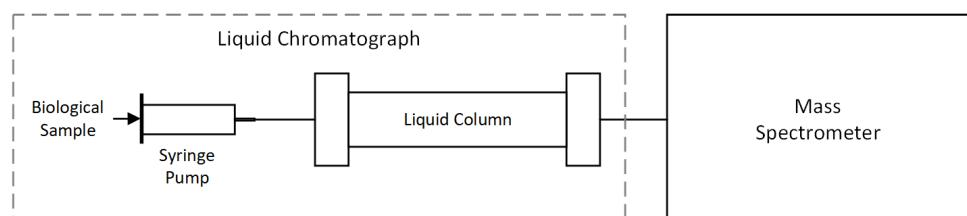


Figure 8: Liquid Chromatograph Mass Spectrometer

Capillary Electrophoresis Mass Spectrometry (CE-MS)

CE-MS provides several advantages over other separation techniques such as high resolving power, very small sample requirement and short analysis time. One of the significant advantages of the CE-MS is that it separates cations, anions and uncharged molecules in a single analytical run, and therefore CE can be used for simultaneous profiling of metabolites. Capillary Electrophoresis is shown in [Figure 9](#).

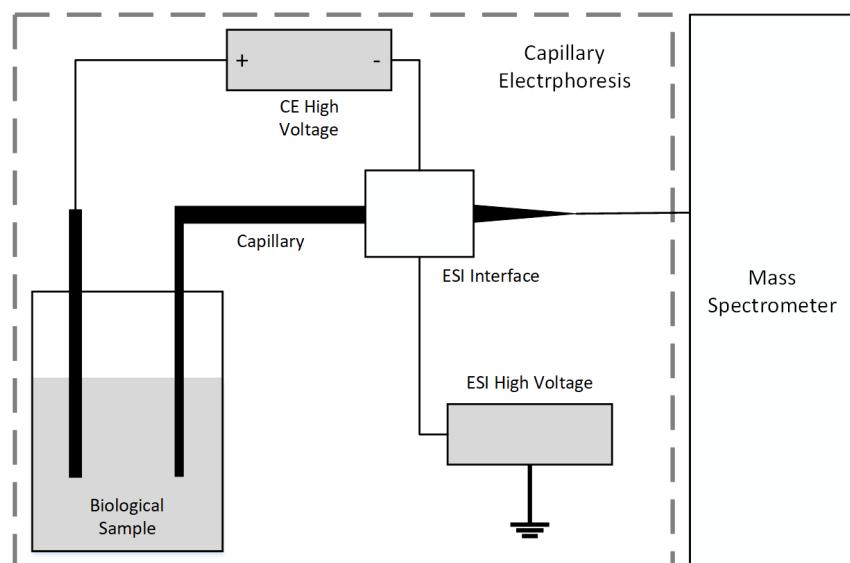


Figure 9: Capillary Electrophoresis Mass Spectrometer

Data Mining and Extraction

Untargeted Metabolomic analysis generate large amount of complex data sets that require analysis by specialized software to properly interpret the data. Data is first preprocessed and then processed at this stage. A good understanding of the steps involved is important to minimize the risk of false results.

There are several commercial and free software packages available to automate this process. A sample MS based data is shown in

row.names	85.02867535	85.04812685	86.03200066	86.060363	86.09659861	86.99289159	87.04437142	87.05551116	87.06353237
X20100920_11_AC_.30	15.506157	0.74182518	0.6921877	5.370489	1.2064942	5.491629462	9.674962	1.40180223	0.4158399
X20100920_47_AC_.54	15.157152	0.93265191	0.7593487	3.798822	0.9618451	6.503842372	6.366602	1.71769076	0.2881482
X20100917_15_NOR_30	16.375372	0.58056653	0.7405717	7.138574	1.4357149	5.425450082	12.737057	0.95402230	0.4449065
X20100920_64_NOR_49	18.477925	0.88054012	0.8726300	4.741568	1.0756890	5.868078071	8.566578	1.68399379	0.4809376
X20100917_57_NOR_49	15.762851	0.56200899	0.6930960	4.884333	0.5744624	2.503798996	8.690681	1.22498802	0.1215892
X20100917_63_AC_.40	19.536413	0.64521509	0.8932008	7.651749	0.5174646	10.398189262	12.884521	1.39416892	0.5455830
X20100920_06_CNR_25	18.593834	0.41586627	0.8442614	7.022921	0.3035749	8.819085138	12.114758	0.63710216	0.5436208
X20100917_60_RIN_49	18.351163	0.64216709	0.8988681	5.866678	1.0310843	4.412544513	10.356847	1.18170686	0.3557551
X20100921_11_AC_.20	13.240275	0.71741791	0.6292936	6.653956	2.6753299	7.835304659	12.302572	0.95315835	0.3919214
X20100917_20_AC_.52	20.810549	0.52505407	1.0030810	3.292031	0.4360519	5.517507684	5.637380	1.18339989	0.3223558
X20100921_60_CNR_40	19.361601	0.33245938	0.8999786	4.780526	0.4920936	7.832031350	8.593449	0.48339447	0.5449337
X20100920_33_CNR_20	15.766320	0.48723240	0.7254843	7.788746	2.0423227	9.097518700	13.916318	0.45799541	0.4909745
X20100917_04_AC_.53	17.228356	0.64617037	0.8350225	4.558736	0.7396387	7.094692825	7.982026	1.47121290	0.3217715
X20100921_35_AC_.15	9.099162	0.63347865	0.3985374	7.603451	1.1537385	7.715451754	13.225294	0.49742452	0.4055616
X20100917_05_RIN_53	19.110243	0.72510414	0.9771563	3.883035	1.2589184	5.789768841	6.878934	1.68578065	0.3934001
X20100920_12_RIN_30	15.079683	0.66735232	0.7217023	7.878826	2.12114839	5.099457599	14.034617	1.10819566	0.4834576
X20100917_26_CNR_54	17.370201	0.42872721	0.7851646	3.057463	0.5398271	7.240740668	5.355297	1.32541153	0.3296503
X20100920_62_RIN_49	18.537254	0.65715251	0.9260807	5.270517	1.1963665	3.614631140	9.508761	1.30148389	0.4028041

Figure 10: Example of MS based data matrix. Transpositions of the matrix are also common

NOTE MetumpX v2.1 has 103 software tools packaged for easy installation and working for the benefit of life scientists which cover the entire software pipeline.

Data Pre-processing

Data Preprocessing is divided into several sub-steps as follows:

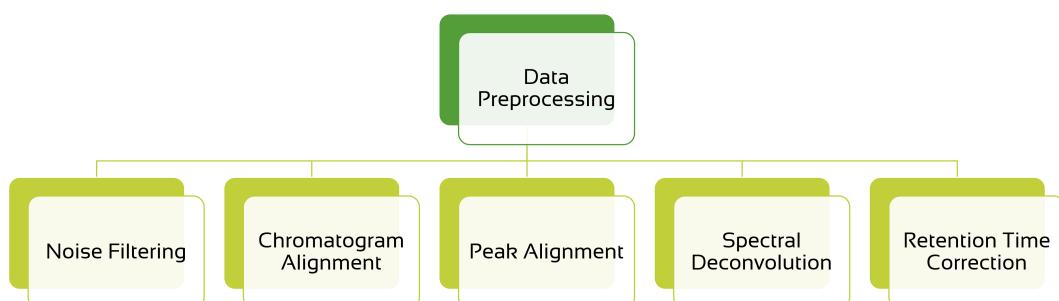


Figure 11: Data Pre-processing

STEP 1 - Noise Filtering

Noise filtering is the process of removing noise from a signal which facilitates further peak detection. It is an optional stage in data processing and can also be left out if the data is not noisy.

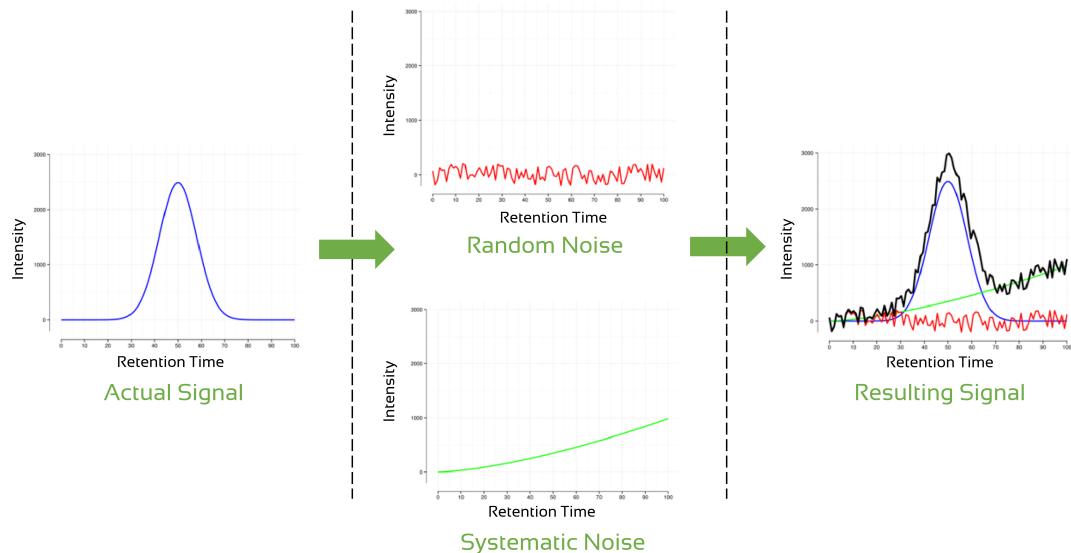


Figure 12: A summary of noise components contributing to signal distortions

- **Random noise** - this results from contaminants and general technological limitations.
- **Systematic noise** - this results from external factors that are not relevant for the study e.g Baseline drift shown in Figure 12.

STEP 2 - Chromatogram Alignment

Chromatogram alignment is the process of aligning retention time for chromatographic methods with the mass spectrometers. Figure 13 shows the unaligned chromatogram $f(x)$, target chromatogram $f(y)$ and aligned chromatogram $f(x')$ [105].

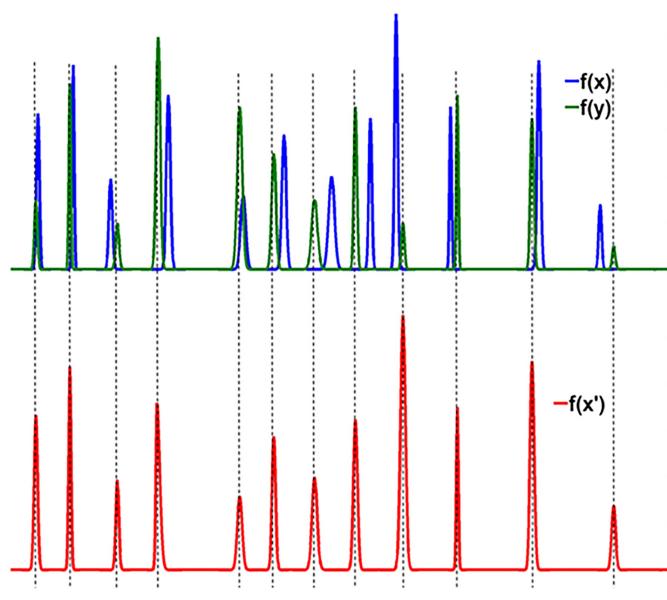


Figure 13: Chromatogram Alignment

STEP 3 - Peak Alignment

Peak alignment is a process of correction of samples to point to the same metabolite or component. It is important in metabolomic studies as there is always a difference in the samples due to machine drift.

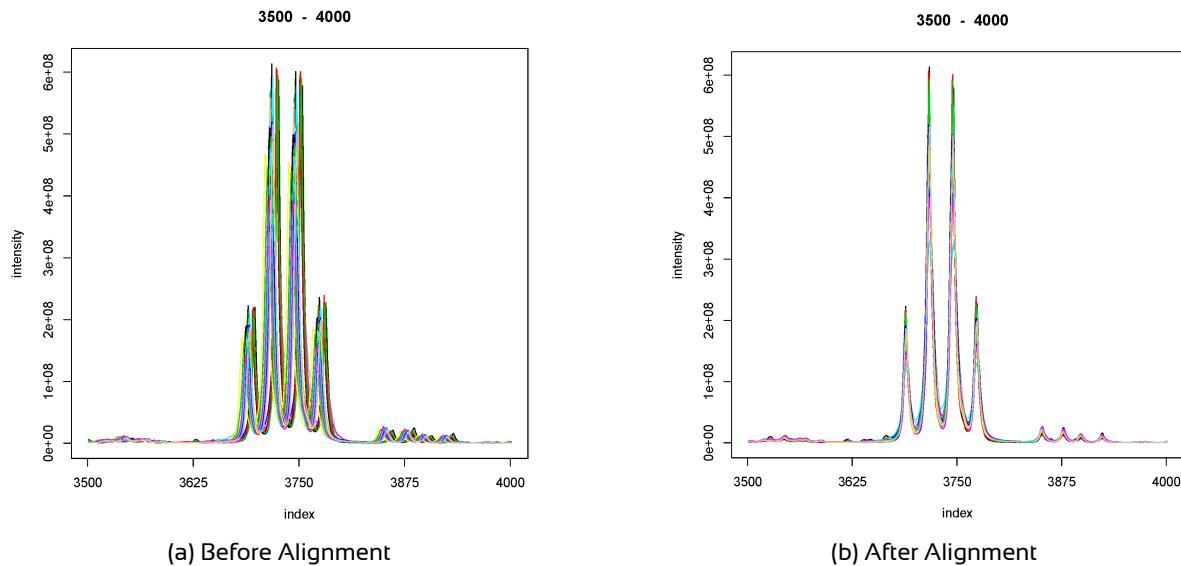


Figure 14: Peak Alignment

STEP 4 - Spectral Deconvolution

Deconvolution is the process of computationally separating co-eluting components and creating a pure spectrum for each component. Deconvolution calculates the contribution of each component from an Extracted Ion Chromatogram that results from two or more components.

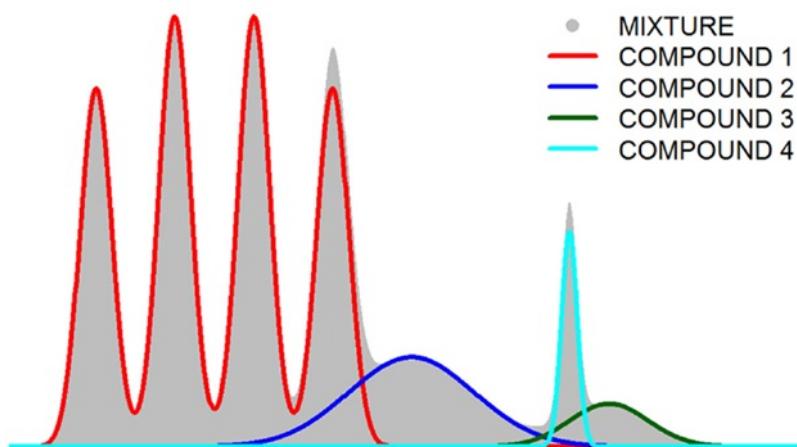


Figure 15: Spectra (gray shaded area) with decomposed (i.e., deconvoluted) multiple components corresponding to different metabolite compounds.

STEP 5 - Retention Time Correction

A process of removals of errors in the retention time due to temperature, polarity etc.

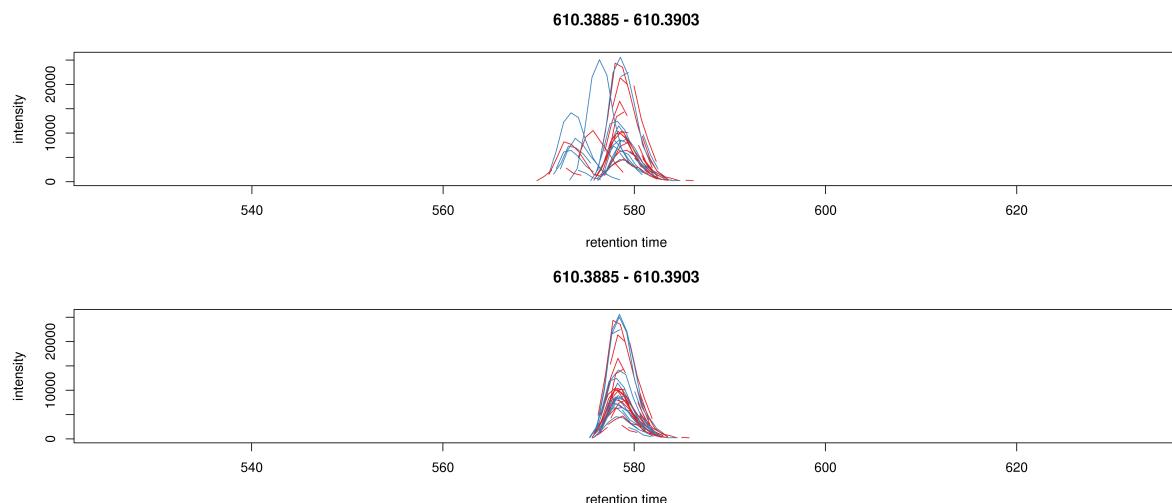


Figure 16: Retention Time Correction

Data Pre-processing

Data Processing is divided into several substeps as follows:

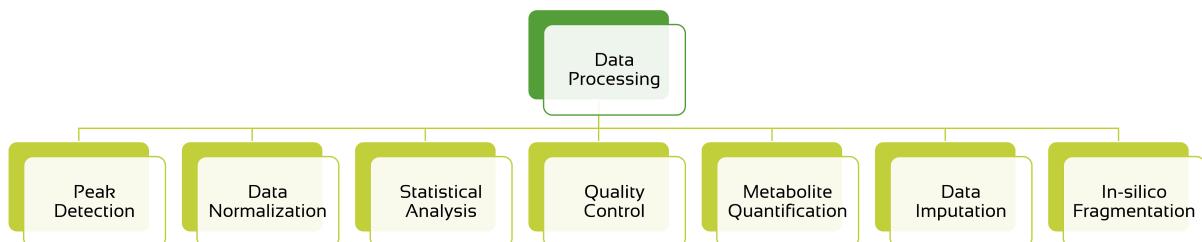


Figure 17: Data Pre-processing

STEP 1 - Peak Detection

Peak Detection is a process of identification and quantification of the features present in the spectra acquired from the spectrometer.

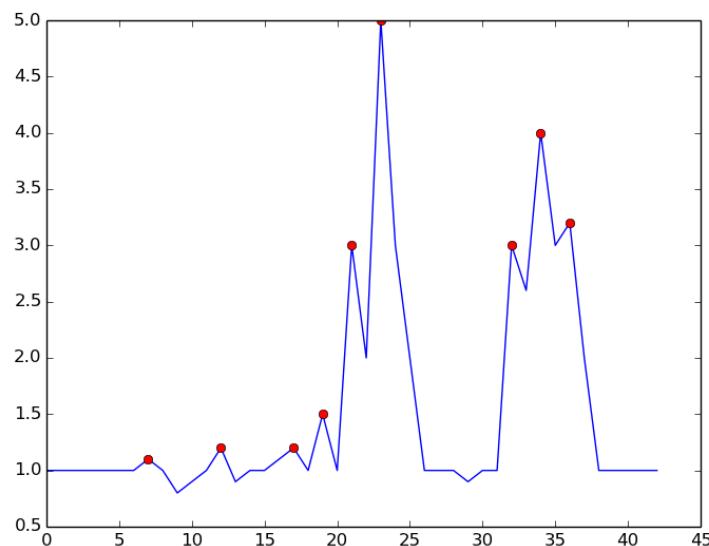


Figure 18: Peak Detection

STEP 2 - Data Normalization

The total sample amount or concentration of metabolites in metabolomic workflow can be significantly different from one sample to another in each step. Data Normalization is the reduction or elimination of the effect of this variation and align them according to some standard.

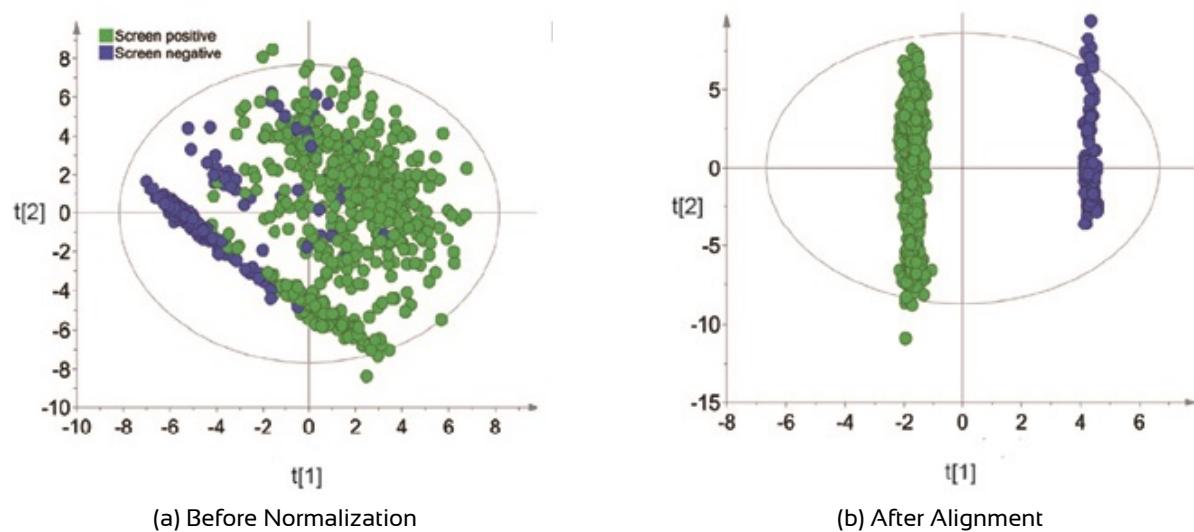


Figure 19: Data Normalization

Quality Control

Quality assurance and quality control provides a mechanism to ensure that a scientific process meets the predefined criteria. Quality control (QC) sample should qualitatively and quantitatively represent the entire collection of samples included in the study, providing an average of all of the metabolomes analysed in the study. To establish

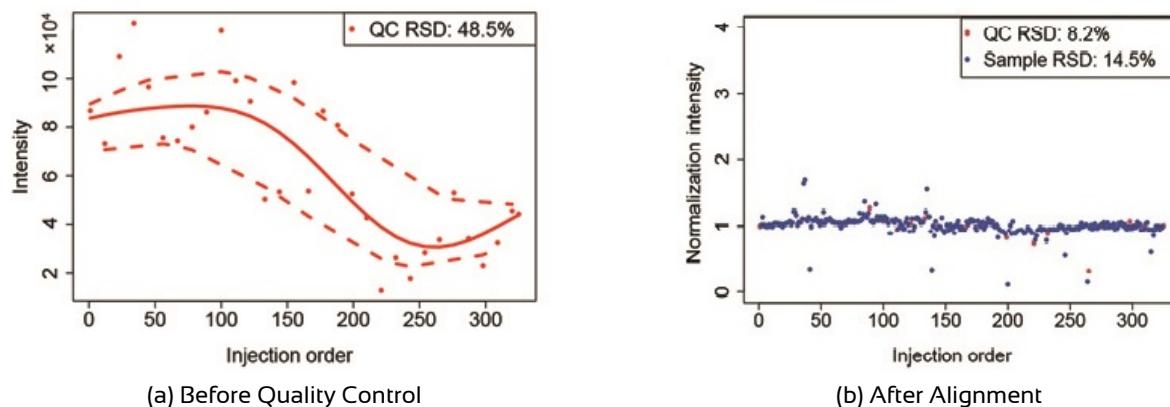


Figure 20: Data Normalization

the ability to generate precision results and determine accuracy of the method, many replicates of a standard are analyzed. The average percent recovery of the replicates and standard deviation of the analyte is calculated.

The percentage of **Relative Standard Deviation (RSD)** is calculated by dividing the standard deviation by the average. The QC-RSD is of uncontrolled and controlled specimen is shown in [Figure 20](#)

STEP 3 - Statistical Analysis

Metabolomics samples are typically complex and there are multiple interactions between metabolites in biological states. To uncover significant events, univariate and multivariate statistical analysis (chemometric methods) platforms use visualization tools to assess abundance relationships between different lipid components.

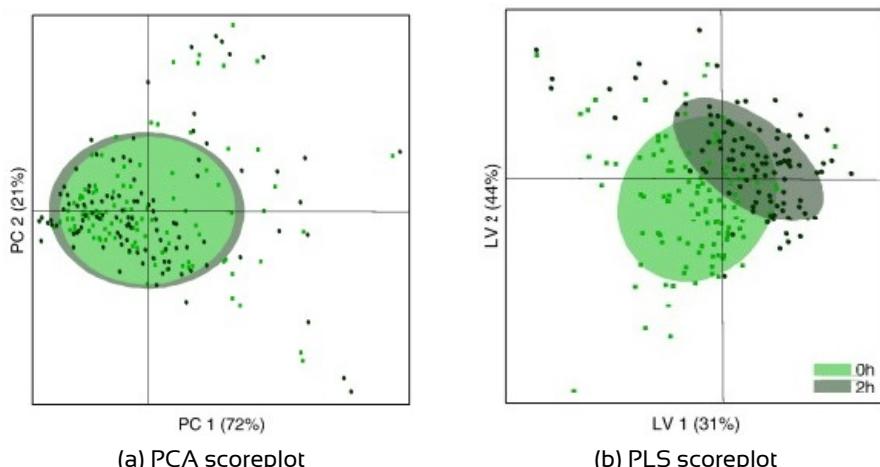


Figure 21: Statistical scoreplots

- **Univariate methods** are the most common statistical approach and analyze metabolite compound features independently. When assessing differences between two or more groups, parametric tests such as student's t-test and ANOVA (analysis of variance) are commonly used.
- **Multivariate methods** analyze metabolite or compound features simultaneously and can identify relationships between them. **Principal component analysis (PCA)** and **Partial Least Squares (PLS)** are common examples of a multivariate method approaches shown in [Figure 21](#)

STEP 4 - Metabolite Quantification

Metabolite Quantification is method which evaluate changes in metabolic activity in response to disease, treatment, environmental and genetic perturbations.

STEP 5 - Data Imputation

Data Imputation is the process of handling the missing values in mass spectrometry. Typically, there are three types of missing values, missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR).

STEP 6 - In silico fragmentation

In silico fragmentation is used to identify unknown compounds outside the database domain by comparing theoretical and experimental data.

Metabolite Identification

Metabolite Identification is the main step in which the metabolites are identified, profiled and compared with the chemical database. This is done by matching against the widely available NIST and Wiley libraries for compound identification.

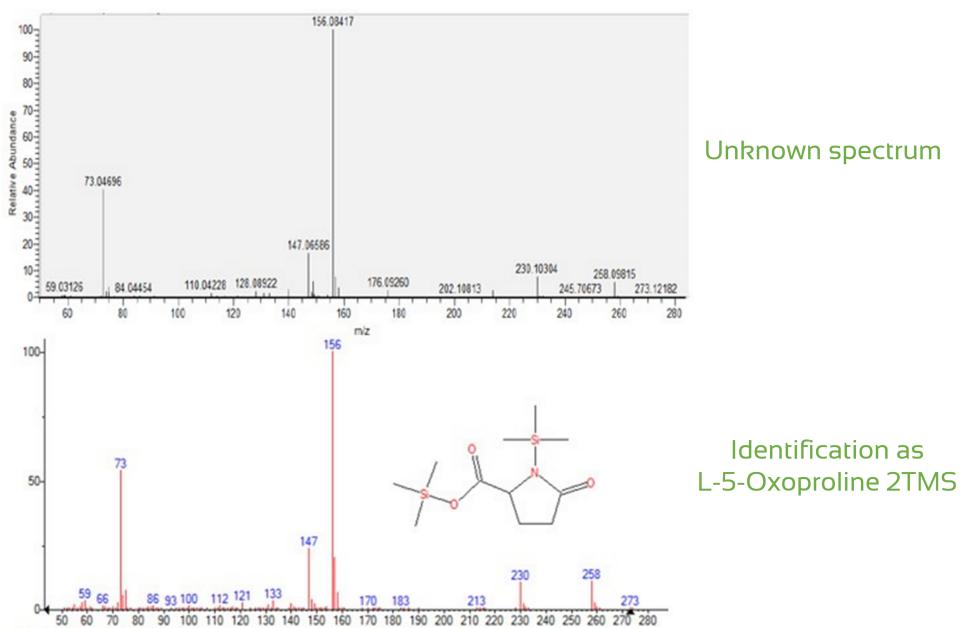


Figure 22: Metabolite Identification

Data Analysis

Several types of Data Analysis is done as follows:

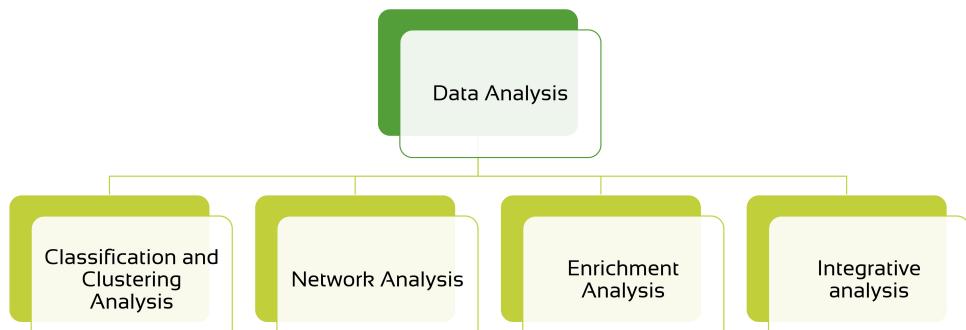


Figure 23: Data Analysis

Classification and Clustering Analysis

Clustering is a well-established technique in which samples are grouped and visualized according to intrinsic similarities in their measurements, irrespective of sample groupings.

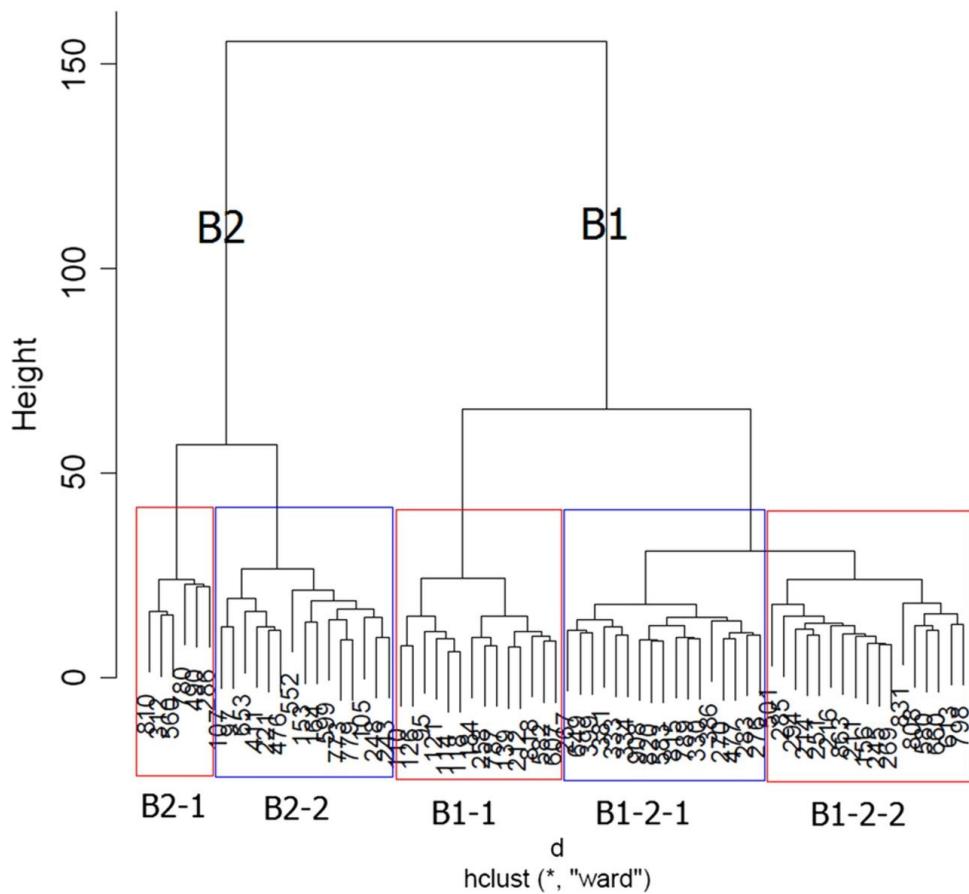


Figure 24: Classification and Clustering Analysis

Network Analysis

Network Analysis can be further divided into three steps:

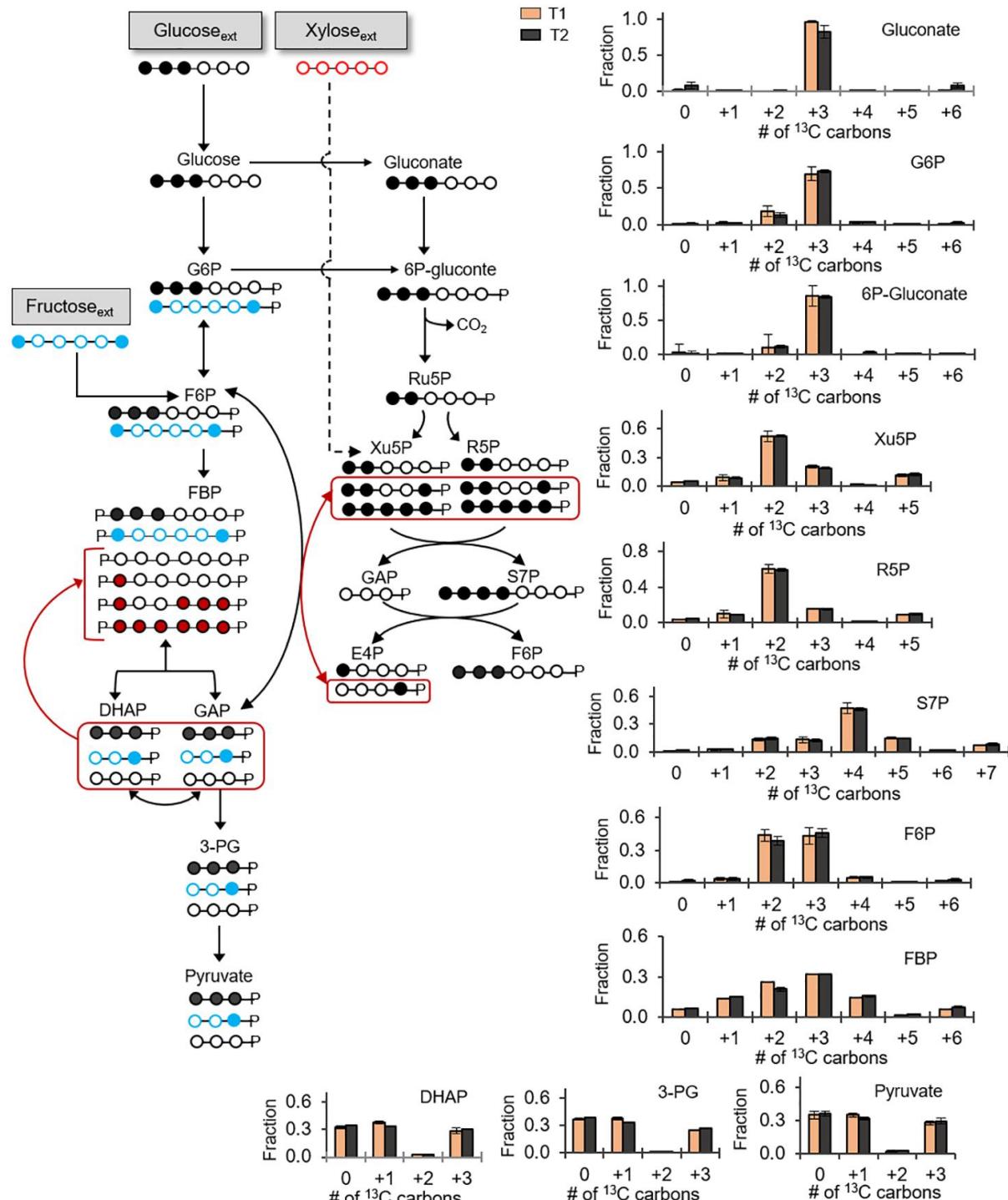


Figure 25: Carbon mapping (left) and metabolite labeling (right) on network mapping of [$1,2,3\text{-}{}^{13}\text{C}_3$]-glucose (black circles), [$1,6\text{-}{}^{13}\text{C}_2$]-fructose (blue circles), and unlabeled xylose (red circles). [106]

STEP 1 - Metabolic Network Analysis

Metabolite datasets are combined through clustering analysis and represented as connection networks between genes and metabolites. The process of analyzing these datasets in called Metabolic Network Analysis.

STEP 2 - metabolite Genome-wide Association Study (mGWAS)

Genome-wide association studies with metabolic traits (mGWAS) investigate how genetic variation effect metabolic phenotypes specially metabolism and complex disease. It maps the loci, position of a genes on a chromosome, responsible for natural variations in a target phenotype.

STEP 3 - Metabolite Mapping

Metabolite mapping is a process of Integration of biochemical pathway and chemical relationships to map all detected metabolites in network graphs.

Enrichment analysis

Metabolite Enrichment Analysis is statistical analysis of metabolite annotations and/or associated quantitative data.

Integrative analysis

Integrative Analysis is process of linking metabolite data with other types of data (e.g. transcriptomics, proteomics), and incorporating prior knowledge of pathways and molecular interactions.

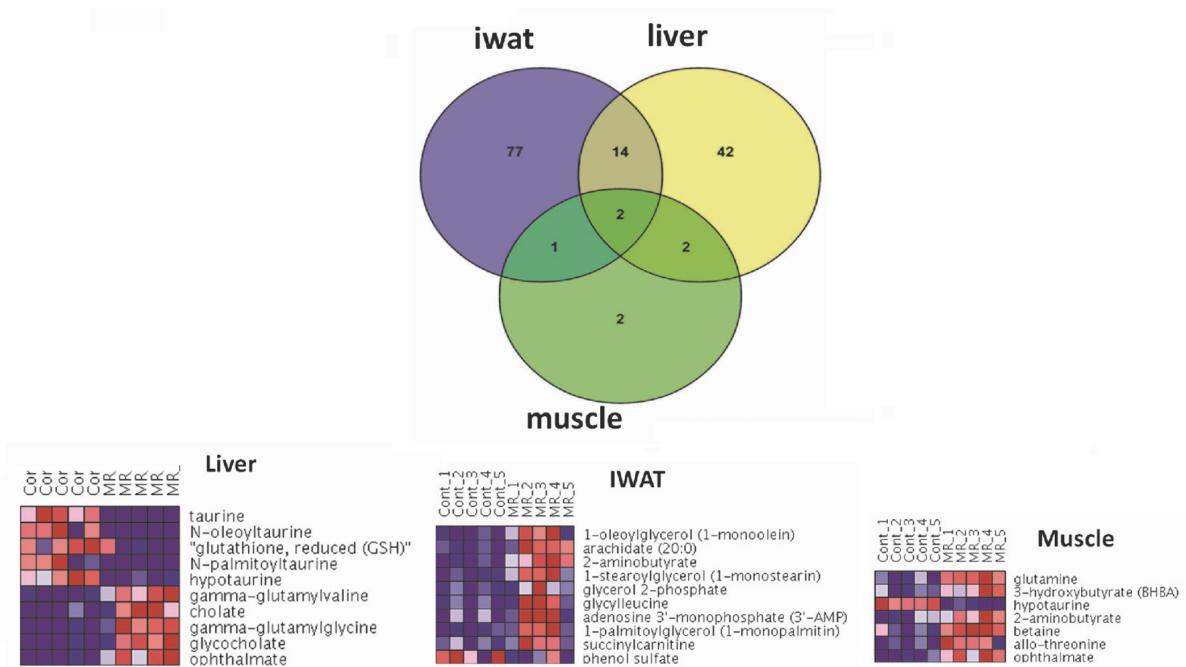


Figure 26: Overlap analysis of the differentially expressed metabolites and respective Heatmaps

Data Visualization

Visualization of the metabolic spectrum, pathways and maps. All the visualization of these things is already shown in each explanation.

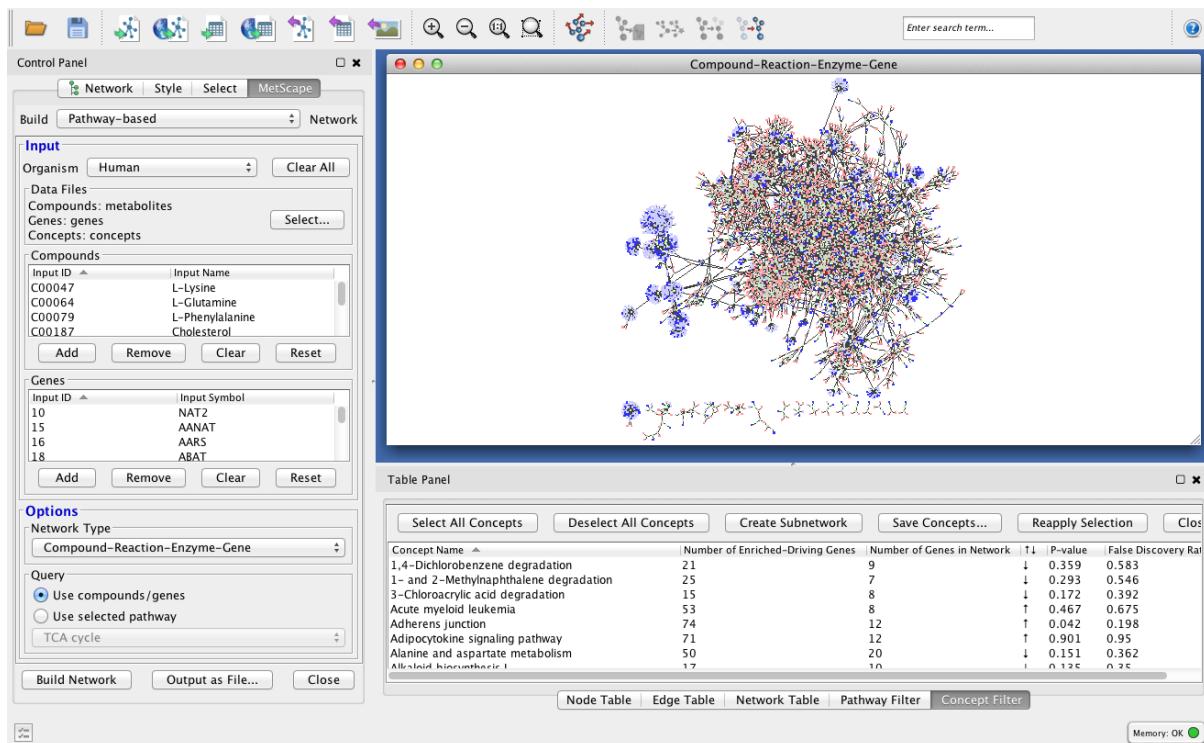


Figure 27: Visualization of pathways by using Cytoscape

Bibliography

- [1] H. Poppe, "Overloading and interaction phenomena in electrophoretic separations," *Analytical Chemistry*, vol. 64, no. 17, pp. 1908–1919, 1992.
- [2] C. Costa, M. Maraschin, and M. Rocha, "An r package for the integrated analysis of metabolomics and spectral data," *Computer methods and programs in biomedicine*, vol. 129, pp. 117–124, 2016.
- [3] D. Ghosh and H. Choi, "metaarray package for meta-analysis of microarray data," *R package*. <http://www.r-project.org>, 2006.
- [4] E. Valentini, M. Zampieri, M. Malavolta, M. G. Bacalini, R. Calabrese, T. Guastafierro, A. Reale, C. Franceschi, A. Hervonen, B. Koller *et al.*, "Analysis of the machinery and intermediates of the 5hmC-mediated dna demethylation pathway in aging on samples from the mark-age study," *Aging (Albany NY)*, vol. 8, no. 9, p. 1896, 2016.
- [5] E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J. H. Gundlach, "Reading dna at single-nucleotide resolution with a mutant mspA nanopore and phi29 dna polymerase," *Nature biotechnology*, vol. 30, no. 4, p. 349, 2012.
- [6] H. Takahashi, T. Morimoto, N. Ogasawara, and S. Kanaya, "Amdorap: Non-targeted metabolic profiling based on high-resolution lc-ms," *BMC bioinformatics*, vol. 12, no. 1, p. 259, 2011.
- [7] R. J. Weber and M. R. Viant, "Mi-pack: Increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways," *Chemometrics and Intelligent Laboratory Systems*, vol. 104, no. 1, pp. 75–82, 2010.
- [8] R. Aggio, S. G. Villas-Bôas, and K. Ruggiero, "Metab: an r package for high-throughput analysis of metabolomics data generated by gc-ms," *Bioinformatics*, vol. 27, no. 16, pp. 2316–2318, 2011.
- [9] H. Tsugawa, T. Cajka, T. Kind, Y. Ma, B. Higgins, K. Ikeda, M. Kanazawa, J. VanderGheynst, O. Fiehn, and M. Arita, "Ms-dial: data-independent ms/ms deconvolution for comprehensive metabolome analysis," *Nature methods*, vol. 12, no. 6, p. 523, 2015.

- [10] J. Zaia, K. Khatri, J. Klein, C. Shao, Y. Sheng, and R. Viner, "Complete molecular weight profiling of low-molecular weight heparins using size exclusion chromatography-ion suppressor-high-resolution mass spectrometry," *Analytical chemistry*, vol. 88, no. 21, pp. 10 654–10 660, 2016.
- [11] M. Bartholomae, T. Baumann, J. H. Nickling, D. Peterhoff, R. Wagner, N. Budisa, and O. P. Kuipers, "Expanding the genetic code of lactococcus lactis and escherichia coli to incorporate non-canonical amino acids for production of modified lantibiotics," *Frontiers in microbiology*, vol. 9, p. 657, 2018.
- [12] Y. Ni, Y. Qiu, W. Jiang, K. Suttlemyre, M. Su, W. Zhang, W. Jia, and X. Du, "Adap-gc 2.0: deconvolution of coeluting metabolites from gc/tof-ms data for metabolomics studies," *Analytical chemistry*, vol. 84, no. 15, pp. 6619–6629, 2012.
- [13] J. Elith, S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates, "A statistical explanation of maxent for ecologists," *Diversity and distributions*, vol. 17, no. 1, pp. 43–57, 2011.
- [14] D. M. Small, "The effects of glyceride structure on absorption and metabolism," *Annual review of nutrition*, vol. 11, no. 1, pp. 413–434, 1991.
- [15] W. M. Edmands, D. K. Barupal, and A. Scalbert, "Metmsline: an automated and fully integrated pipeline for rapid processing of high-resolution lc-ms metabolomic datasets," *Bioinformatics*, vol. 31, no. 5, pp. 788–790, 2014.
- [16] X. Huang, Y.-J. Chen, K. Cho, I. Nikolskiy, P. A. Crawford, and G. J. Patti, "X13cms: global tracking of isotopic labels in untargeted metabolomics," *Analytical chemistry*, vol. 86, no. 3, pp. 1632–1639, 2014.
- [17] A. Delabrière, U. M. Hohenester, B. Colsch, C. Junot, F. Fenaille, and E. A. Thévenot, "profia: a data preprocessing workflow for flow injection analysis coupled to high-resolution mass spectrometry," *Bioinformatics*, vol. 33, no. 23, pp. 3767–3775, 2017.
- [18] V. Santini, A. Sanna, A. Bosi, G. Alimena, G. Loglisci, A. Levis, F. Salvi, C. Finelli, C. Clissa, G. Specchia *et al.*, "An observational multicenter study to assess the cost of illness and quality of life in patients with myelodysplastic syndromes in italy," 2011.
- [19] A. Chokkathukalam, A. Jankevics, D. J. Creek, F. Achcar, M. P. Barrett, and R. Breitling, "mzmatch-iso: an r tool for the annotation and relative quantification of isotope-labelled mass spectrometry data," *Bioinformatics*, vol. 29, no. 2, pp. 281–283, 2012.
- [20] J. J. Boon, "Analytical pyrolysis mass spectrometry: new vistas opened by temperature-resolved in-source pyms," *International Journal of Mass Spectrometry and Ion Processes*, vol. 118, pp. 755–787, 1992.
- [21] M. C. Potter, M. M. Chun, B. S. Banks, and M. Muckenhaupt, "Two attentional deficits in serial target search: the visual attentional blink and an amodal task-switch deficit." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 24, no. 4, p. 979, 1998.

-
- [22] P. Tasiopoulou and N. Gerolymos, "Constitutive modeling for sand with emphasis on the evolution of bounding and phase transformation lines," in *Proc. of the 8th European Conference on Numerical Methods in Geotechnical Engineering, Delft, the Netherlands*, 2014, pp. 109–114.
 - [23] K. Wanichthanarak, S. Fan, D. Grapov, D. K. Barupal, and O. Fiehn, "Metabox: A toolbox for metabolomic data analysis, interpretation and integrative exploration," *PLoS One*, vol. 12, no. 1, p. e0171046, 2017.
 - [24] J. Suhy, W. Rooney, D. Goodkin, A. Capizzano, B. Soher, A. A. Maudsley, E. Waubant, P. Andersson, and M. Weiner, "1h mrsi comparison of white matter and lesions in primary progressive and relapsing-remitting ms," *Multiple Sclerosis Journal*, vol. 6, no. 3, pp. 148–155, 2000.
 - [25] N. Desmarais, F. Reuse, and S. Khanna, "Magnetic coupling in neutral and charged cr 2, mn 2, and crmn dimers," *The Journal of Chemical Physics*, vol. 112, no. 13, pp. 5576–5584, 2000.
 - [26] W. Peng and D. J. Edwards, "K-means like minimum mean distance algorithm for wireless sensor networks," in *2010 2nd International Conference on Computer Engineering and Technology*, vol. 1. IEEE, 2010, pp. V1–120.
 - [27] G. Hughes, C. Cruickshank-Quinn, R. Reisdorph, S. Lutz, I. Petrache, N. Reisdorph, R. Bowler, and K. Keckris, "Msprep—summarization, normalization and diagnostics for processing of mass spectrometry-based metabolomic data," *Bioinformatics*, vol. 30, no. 1, pp. 133–134, 2013.
 - [28] P. M. Ayton, E. H. Chen, and M. L. Cleary, "Binding to nonmethylated cpg dna is essential for target recognition, transactivation, and myeloid transformation by an mll oncoprotein," *Molecular and cellular biology*, vol. 24, no. 23, pp. 10 470–10 478, 2004.
 - [29] A. S. Kirpich, M. Ibarra, O. Moskalenko, J. M. Fear, J. Gerken, X. Mi, A. Ashrafi, A. M. Morse, and L. M. McIntyre, "Secimtools: a suite of metabolomics data analysis tools," *BMC bioinformatics*, vol. 19, no. 1, p. 151, 2018.
 - [30] A. M. Simader, B. Kluger, N. K. N. Neumann, C. Bueschl, M. Lemmens, G. Lirk, R. Krska, and R. Schuhmacher, "Qcscreen: a software tool for data quality control in lc-hrms based metabolomics," *BMC bioinformatics*, vol. 16, no. 1, p. 341, 2015.
 - [31] B. B. Misra, J. F. Fahrmann, and D. Grapov, "Review of emerging metabolomic tools and resources: 2015–2016," *Electrophoresis*, vol. 38, no. 18, pp. 2257–2274, 2017.
 - [32] B. Bunk, M. Kucklick, R. Jonas, R. Münch, M. Schobert, D. Jahn, and K. Hiller, "Metaquant: a tool for the automatic quantification of gc/ms-based metabolome data," *Bioinformatics*, vol. 22, no. 23, pp. 2962–2965, 2006.
 - [33] T. Yu, Y. Park, J. M. Johnson, and D. P. Jones, "aplcms—adaptive processing of high-resolution lc/ms data," *Bioinformatics*, vol. 25, no. 15, pp. 1930–1936, 2009.
-

- [34] Z. Jin, J. Kang, and T. Yu, "Missing value imputation for lc-ms metabolomics data by incorporating metabolic network and adduct ion relations," *Bioinformatics*, vol. 34, no. 9, pp. 1555–1561, 2017.
 - [35] A. A. Gonzalez Ibarra, K. Wrobel, E. Yanez Barrientos, A. R. Corrales Escobosa, J. F. Gutierrez Corona, I. Enciso Donis, and K. Wrobel, "Changes of metabolomic profile in helianthus annuus under exposure to chromium (vi) studied by caphplc-esi-qtof-ms and ms/ms," *Journal of analytical methods in chemistry*, vol. 2017, 2017.
 - [36] M. Cao, K. Fraser, and S. Rasmussen, "Computational analyses of spectral trees from electrospray multi-stage mass spectrometry to aid metabolite identification," *Metabolites*, vol. 3, no. 4, pp. 1036–1050, 2013.
 - [37] S. Neumann, A. Thum, and C. Böttcher, "Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data," *Metabolomics*, vol. 9, no. 1, pp. 84–91, 2013.
 - [38] L. C. Menikarachchi, S. Cawley, D. W. Hill, L. M. Hall, L. Hall, S. Lai, J. Wilder, and D. F. Grant, "Molfind: a software package enabling hplc/ms-based identification of unknown chemical structures," *Analytical chemistry*, vol. 84, no. 21, pp. 9388–9394, 2012.
 - [39] J. B. Jackson, M. X. Kirby, W. H. Berger, K. A. Bjorndal, L. W. Botsford, B. J. Bourque, R. H. Bradbury, R. Cooke, J. Erlandson, J. A. Estes *et al.*, "Historical overfishing and the recent collapse of coastal ecosystems," *science*, vol. 293, no. 5530, pp. 629–637, 2001.
 - [40] R. Wehrens, G. Weingart, and F. Mattivi, "metams: An open-source pipeline for gc-ms-based untargeted metabolomics," *Journal of Chromatography B*, vol. 966, pp. 109–116, 2014.
 - [41] F. Nicolè, Y. Guittot, E. A. Courtois, S. Moja, L. Legendre, and M. Hossaert-McKey, "Mseasy: unsupervised and untargeted gc-ms data processing," *Bioinformatics*, vol. 28, no. 17, pp. 2278–2280, 2012.
 - [42] C. D. Broeckling, F. Afsar, S. Neumann, A. Ben-Hur, and J. Prenni, "Ramclust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data," *Analytical chemistry*, vol. 86, no. 14, pp. 6812–6817, 2014.
 - [43] B. Zhou, J. Wang, and H. W. Ressom, "Metabosearch: tool for mass-based metabolite identification using multiple databases," *PLoS One*, vol. 7, no. 6, p. e40096, 2012.
 - [44] M. F. Clasquin, E. Melamud, and J. D. Rabinowitz, "Lc-ms data processing with maven: a metabolomic analysis and visualization engine," *Current protocols in bioinformatics*, vol. 37, no. 1, pp. 14–11, 2012.
 - [45] J. Capellades, M. Navarro, S. Samino, M. Garcia-Ramirez, C. Hernandez, R. Simo, M. Vinaixa, and O. Yanes, "george: a computational tool to detect the presence of stable isotope labeling in lc/ms-based untargeted metabolomics," *Analytical chemistry*, vol. 88, no. 1, pp. 621–628, 2015.
-

- [46] X. Domingo-Almenara, J. Brezmes, M. Vinaixa, S. Samino, N. Ramirez, M. Ramon-Krauel, C. Lerin, M. Diaz, L. Ibanez, X. Correig *et al.*, "erah: a computational tool integrating spectral deconvolution and alignment with quantification and identification of metabolites in gc/ms-based metabolomics," *Analytical chemistry*, vol. 88, no. 19, pp. 9821–9829, 2016.
- [47] E. Sentandreu, M. D. Peris-Diaz, S. R. Sweeney, J. Chiou, N. Munoz, and S. Tiziani, "A survey of orbitrap all ion fragmentation analysis assessed by an r metabolist package to study small-molecule metabolites," *Chromatographia*, vol. 81, no. 7, pp. 981–994, 2018.
- [48] C. Jaeger, F. Hoffmann, C. A. Schmitt, and J. Lisec, "Automated annotation and evaluation of in-source mass spectra in gc/atmospheric pressure chemical ionization-ms-based metabolomics," *Analytical chemistry*, vol. 88, no. 19, pp. 9386–9390, 2016.
- [49] J. Wills, J. Edwards-Hicks, and A. J. Finch, "Assayr: A simple mass spectrometry software tool for targeted metabolic and stable isotope tracer analyses," *Analytical chemistry*, vol. 89, no. 18, pp. 9616–9619, 2017.
- [50] Y. Ma, T. Kind, D. Yang, C. Leon, and O. Fiehn, "Ms2analyzer: a software for small molecule substructure annotations from accurate tandem mass spectra," *Analytical chemistry*, vol. 86, no. 21, pp. 10 724–10 731, 2014.
- [51] H. Mizuno, K. Ueda, Y. Kobayashi, N. Tsuyama, K. Todoroki, J. Z. Min, and T. Toyo'oka, "The great importance of normalization of lc-ms data for highly-accurate non-targeted metabolomics," *Biomedical Chromatography*, vol. 31, no. 1, p. e3864, 2017.
- [52] H. Redestig, M. Kusano, A. Fukushima, F. Matsuda, K. Saito, and M. Arita, "Consolidating metabolite identifiers to enable contextual and multi-platform metabolomics data analysis," *BMC bioinformatics*, vol. 11, no. 1, p. 214, 2010.
- [53] T. Suvitaival, S. Rogers, and S. Kaski, "Stronger findings from mass spectral data through multi-peak modeling," *BMC bioinformatics*, vol. 15, no. 1, p. 208, 2014.
- [54] R. Spicer, R. M. Salek, P. Moreno, D. Cañuelo, and C. Steinbeck, "Navigating freely-available software tools for metabolomics analysis," *Metabolomics*, vol. 13, no. 9, p. 106, 2017.
- [55] W. M. Edmands, J. Hayes, and S. M. Rappaport, "Simextargid: a comprehensive package for real-time lc-ms data acquisition and analysis," *Bioinformatics*, vol. 34, no. 20, pp. 3589–3590, 2018.
- [56] J. Chong, O. Soufan, C. Li, I. Caraus, S. Li, G. Bourque, D. S. Wishart, and J. Xia, "Metaboanalyst 4.0: towards more transparent and integrative metabolomics analysis," *Nucleic acids research*, vol. 46, no. W1, pp. W486–W494, 2018.
- [57] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000.

- [58] H. Ji, F. Zeng, Y. Xu, H. Lu, and Z. Zhang, "Kpic2: an effective framework for mass spectrometry-based metabolomics using pure ion chromatograms," *Analytical chemistry*, vol. 89, no. 14, pp. 7631–7640, 2017.
- [59] A. Rodriguez-Martinez, R. Ayala, J. M. Posma, N. Harvey, B. Jiménez, K. Sonomura, T.-A. Sato, F. Matsuda, P. Zalloua, D. Gauguier *et al.*, "pjres binning algorithm (jba): a new method to facilitate the recovery of metabolic information from pjres 1h nmr spectra," *Bioinformatics*, 2018.
- [60] Y. Wen, P. Li, J. Hao, C. Duan, J. Han, A. He, Y. Du, L. Liu, X. Liang, F. Zhang *et al.*, "Integrating genome-wide dna methylation and mrna expression profiles identified different molecular features between kashin-beck disease and primary osteoarthritis," *Arthritis research & therapy*, vol. 20, no. 1, p. 41, 2018.
- [61] M. Kutmon, M. P. van Iersel, A. Bohler, T. Kelder, N. Nunes, A. R. Pico, and C. T. Evelo, "Pathvisio 3: an extendable pathway analysis toolbox," *PLoS computational biology*, vol. 11, no. 2, p. e1004085, 2015.
- [62] L. B. Jorde, "Linkage disequilibrium as a gene-mapping tool." *American journal of human genetics*, vol. 56, no. 1, p. 11, 1995.
- [63] I. Laponogov, N. Sadawi, D. Galea, R. Mirnezami, and K. A. Veselkov, "Chemdistiller: an engine for metabolite annotation in mass spectrometry," *Bioinformatics*, vol. 34, no. 12, pp. 2096–2102, 2018.
- [64] C. D. Christensen, J.-H. S. Hofmeyr, and J. M. Rohwer, "Pyscestoolbox: a collection of metabolic pathway analysis tools," *Bioinformatics*, vol. 34, no. 1, pp. 124–125, 2017.
- [65] C. Jia, W. Guan, A. Yang, R. Xiao, W. W. Tang, C. S. Moravec, K. B. Margulies, T. P. Cappola, C. Li, and M. Li, "Metadiff: differential isoform expression analysis using random-effects meta-regression," *BMC bioinformatics*, vol. 16, no. 1, p. 208, 2015.
- [66] G. Yu and Q.-Y. He, "Reactomepa: an r/bioconductor package for reactome pathway analysis and visualization," *Molecular BioSystems*, vol. 12, no. 2, pp. 477–479, 2016.
- [67] V. De Anda, I. Zapata-Peñasco, A. C. Poot-Hernandez, L. E. Eguiarte, B. Contreras-Moreira, and V. Souza, "Mebs, a software platform to evaluate large (meta) genomic collections according to their metabolic machinery: unraveling the sulfur cycle," *GigaScience*, 2017.
- [68] V. Giri, T. V. Sivakumar, K. M. Cho, T. Y. Kim, and A. Bhaduri, "Rxnsim: a tool to compare biochemical reactions," *Bioinformatics*, vol. 31, no. 22, pp. 3712–3714, 2015.
- [69] K. Choi, L. P. Smith, J. K. Medley, and H. M. Sauro, "phrased-ml: A paraphrased, human-readable adaptation of sed-ml," *Journal of bioinformatics and computational biology*, vol. 14, no. 06, p. 1650035, 2016.
- [70] M. Poolman, "Scrumpy: metabolic modelling with python," *IEE Proceedings-Systems Biology*, vol. 153, no. 5, pp. 375–378, 2006.

-
- [71] L. Feng, Y. Xu, Y. Zhang, Z. Sun, J. Han, C. Zhang, H. Yang, D. Shang, F. Su, X. Shi *et al.*, "Subpathway-gmir: identifying mirna-mediated metabolic subpathways by integrating condition-specific genes, micrornas, and pathway topologies," *Oncotarget*, vol. 6, no. 36, p. 39151, 2015.
 - [72] O. K. Kamneva, "Genome composition and phylogeny of microbes predict their co-occurrence in the environment," *PLoS computational biology*, vol. 13, no. 2, p. e1005366, 2017.
 - [73] A. Rodriguez-Martinez, R. Ayala, J. M. Posma, A. L. Neves, D. Gauguier, J. K. Nicholson, and M.-E. Dumas, "Metabosignal: a network-based approach for topological analysis of metabotype regulation via metabolic and signaling pathways," *Bioinformatics*, vol. 33, no. 5, pp. 773–775, 2017.
 - [74] T. V. Sivakumar, V. Giri, J. H. Park, T. Y. Kim, and A. Bhaduri, "Reactpred: a tool to predict and analyze biochemical reactions," *Bioinformatics*, vol. 32, no. 22, pp. 3522–3524, 2016.
 - [75] H. Mendes-Soares, M. Mundy, L. M. Soares, and N. Chia, "Mminte: an application for predicting metabolic interactions among the microbial species in a community," *BMC bioinformatics*, vol. 17, no. 1, p. 343, 2016.
 - [76] L. M. Mulligan, J. B. Kwok, C. S. Healey, M. J. Elsdon, C. Eng, E. Gardner, D. R. Love, S. E. Mole, J. K. Moore, L. Papi *et al.*, "Germ-line mutations of the ret proto-oncogene in multiple endocrine neoplasia type 2a," *Nature*, vol. 363, no. 6428, p. 458, 1993.
 - [77] D. Jiao, Y. Ye, and H. Tang, "Probabilistic inference of biochemical reactions in microbial communities from metagenomic sequences," *PLoS computational biology*, vol. 9, no. 3, p. e1002981, 2013.
 - [78] L. Angeles-Martinez and C. Theodoropoulos, "Estimation of flux distribution in metabolic networks accounting for thermodynamic constraints: The effect of equilibrium vs. blocked reactions," *Biochemical engineering journal*, vol. 105, pp. 347–357, 2016.
 - [79] R. R. Adams, "Sed-ed, a workflow editor for computational biology experiments written in sed-ml," *Bioinformatics*, vol. 28, no. 8, pp. 1180–1181, 2012.
 - [80] C. P. Fisher, N. J. Plant, J. B. Moore, and A. M. Kierzek, "Qsspn: dynamic simulation of molecular interaction networks describing gene regulation, signalling and whole-cell metabolism in human cells," *Bioinformatics*, vol. 29, no. 24, pp. 3181–3190, 2013.
 - [81] G. Basler and Z. Nikoloski, "Jmassbalance: Mass-balanced randomization and analysis of metabolic networks," *Bioinformatics*, vol. 27, no. 19, pp. 2761–2762, 2011.
 - [82] N. Fyson, M. K. Kim, D. Lun, and C. Colijn, "Gene-centric constraint of metabolic models," *bioRxiv*, p. 116558, 2017.
-

- [83] S. Waldherr, "A guideline to model reduction by stoichiometric decomposition for biochemical network analysis," in *Proc. of the 21st International Symposium on Mathematical Theory of Networks and Systems pages*, 2014, pp. 490–495.
- [84] J. Einloft, J. Ackermann, J. Nöthen, and I. Koch, "Monalisa—visualization and analysis of functional modules in biochemical networks," *Bioinformatics*, vol. 29, no. 11, pp. 1469–1470, 2013.
- [85] S. J. Sheridan and H. L. Springfield, "Age appropriate games in the teaching of leisure skills to persons with mental retardation." 1987.
- [86] S. Basu, W. Duren, C. R. Evans, C. F. Burant, G. Michailidis, and A. Karnovsky, "Sparse network modeling and metscape-based visualization methods for the analysis of large-scale metabolomics data," *Bioinformatics*, vol. 33, no. 10, pp. 1545–1553, 2017.
- [87] L. Corrêa, R. Alves, F. Goés, C. Chaparro, and L. Thom, "Funn-mg: A metagenomic systems biology computational framework," in *Brazilian Symposium on Bioinformatics*. Springer, 2014, pp. 25–32.
- [88] K.-A. Lê Cao, I. González, and S. Déjean, "integromics: an r package to unravel relationships between two omics datasets," *Bioinformatics*, vol. 25, no. 21, pp. 2855–2856, 2009.
- [89] A. Karnovsky, T. Weymouth, T. Hull, V. G. Tarcea, G. Scardoni, C. Laudanna, M. A. Sartor, K. A. Stringer, H. Jagadish, C. Burant *et al.*, "Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data," *Bioinformatics*, vol. 28, no. 3, pp. 373–380, 2011.
- [90] E. G. Cerami, G. D. Bader, B. E. Gross, and C. Sander, "cpath: open source software for collecting, storing, and querying biological pathways," *BMC bioinformatics*, vol. 7, no. 1, p. 497, 2006.
- [91] W. Duren, T. Weymouth, T. Hull, G. S. Omenn, B. Athey, C. Burant, and A. Karnovsky, "Metdisease—connecting metabolites to diseases via literature," *Bioinformatics*, vol. 30, no. 15, pp. 2239–2241, 2014.
- [92] H. U. Zacharias, T. Rehberg, S. Mehrl, D. Richtmann, T. Wettig, P. J. Oefner, R. Spang, W. Gronwald, and M. Altenbuchinger, "Scale-invariant biomarker discovery in urine and plasma metabolite fingerprints," *Journal of proteome research*, vol. 16, no. 10, pp. 3596–3605, 2017.
- [93] S. Picart-Armada, F. Fernández-Albert, M. Vinaixa, O. Yanes, and A. Perera-Lluna, "Fella: an r package to enrich metabolomics data," *BMC bioinformatics*, vol. 19, no. 1, p. 538, 2018.
- [94] D. Verdegem, D. Lambrechts, P. Carmeliet, and B. Ghesquière, "Improved metabolite identification with midas and magma through ms/ms spectral dataset-driven parameter optimization," *Metabolomics*, vol. 12, no. 6, p. 98, 2016.

- [95] M. Katajamaa, J. Miettinen, and M. Orešič, "Mzmine: toolbox for processing and visualization of mass spectrometry based molecular profile data," *Bioinformatics*, vol. 22, no. 5, pp. 634–636, 2006.
- [96] C. A. Smith, E. J. Want, G. O'Maille, R. Abagyan, and G. Siuzdak, "Xcms: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification," *Analytical chemistry*, vol. 78, no. 3, pp. 779–787, 2006.
- [97] U. Kolitsch, H. Seifert, T. Ludwig, and F. Aldinger, "Phase equilibria and crystal chemistry in the $\text{Y}_2\text{O}_3\text{-Al}_2\text{O}_3\text{-SiO}_2$ system," *Journal of materials research*, vol. 14, no. 2, pp. 447–455, 1999.
- [98] R. C. Ramaker, E. R. Gordon, and S. J. Cooper, "R2dgc: threshold-free peak alignment and identification for 2d gas chromatography-mass spectrometry in r," *Bioinformatics*, vol. 34, no. 10, pp. 1789–1791, 2017.
- [99] D. M. Avtonomov, A. Raskind, and A. I. Nesvizhskii, "Batmass: a java software platform for lc-ms data visualization in proteomics and metabolomics," *Journal of proteome research*, vol. 15, no. 8, pp. 2500–2509, 2016.
- [100] M. Rasmussen, Y. Li, S. Lindgreen, J. S. Pedersen, A. Albrechtsen, I. Moltke, M. Metspalu, E. Metspalu, T. Kivisild, R. Gupta *et al.*, "Ancient human genome sequence of an extinct palaeo-eskimo," *Nature*, vol. 463, no. 7282, p. 757, 2010.
- [101] R. R. Silva, F. Jourdan, D. M. Salvanha, F. Letisse, E. L. Jamin, S. Guidetti-Gonzalez, C. A. Labate, and R. Z. Vêncio, "Probmetab: an r package for bayesian probabilistic annotation of lc-ms-based metabolomics," *Bioinformatics*, vol. 30, no. 9, pp. 1336–1337, 2014.
- [102] T. Naake and E. Gaquerel, "Metcirc: navigating mass spectral similarity in high-resolution ms/ms metabolomics data," *Bioinformatics*, vol. 33, no. 15, pp. 2419–2420, 2017.
- [103] K. Uppal, Q. A. Soltow, F. H. Strobel, W. S. Pittard, K. M. Gernert, T. Yu, and D. P. Jones, "xmsanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data," *BMC bioinformatics*, vol. 14, no. 1, p. 15, 2013.
- [104] W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin, "Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy," *Chemical Society Reviews*, vol. 40, no. 1, pp. 387–426, 2011.
- [105] K. Kumar, A. Espaillat, and F. Cava, "Pg-metrics: A chemometric-based approach for classifying bacterial peptidoglycan data sets and uncovering their subjacent chemical variability," *PLOS ONE*, vol. 12, p. e0186197, 10 2017.
- [106] J. A. Wushensky, T. Youngster, C. M. Mendonca, and L. Aristilde, "Flux connections between gluconate pathway, glycolysis, and pentose-phosphate pathway during carbohydrate metabolism in *bacillus megaterium qmb1551*," *Frontiers in microbiology*, vol. 9, p. 2789, 2018.