
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/>

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

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.

Trademarks:

Sabz Qalam and the "feather" logo are registered trademarks of Sabz Qalam. All other marks are property of their respective owners.

Department of Information Technology
Sabz Qalam
6 Lodge Road, Old Anarkali
Lahore, Pakistan

REF MX-211

Acknowledgment:

This work has been supported by Sabz-Qalam via grant SQ-2018-Bioinfo-1.



Sabz-Qalam is a privately held company and its Department of Information Technology currently working in IT consulting, software development, testing and support.

Department of Information Technology
Sabz Qalam
6 Lodge Road, Old Anarkali
Lahore, Pakistan

Contents

About MetumpX	1
Abstract	1
Purpose	1
Intended Audience	1
Required Background	1
How This Guide is Organized	2
Contact	2
MetumpX Package	3
Software Selection Criteria	3
Software Categorization Scheme	9
MetumpX Packaged Tools	12
Software Installation Process	17
Installing MetumpX software	17
Microsoft Windows Support	27
Installing MetumpX software on Microsoft Windows	27
Introduction to Metabolomics	37
In this Chapter	37
Introduction	37
Interactome	38
Metabolomic Study Strategies	39
Metabolomics Workflow	41
In this Chapter	41
Introduction	41
Untargeted Metabolomics Workflow	42
Sample Acquisition	43

Sample Preparation	43
Data Mining and Extraction	47
Metabolite Identification	53
Data Analysis	54
Data Visualization	62
MetumpX Software Pipeline	63

About MetumpX

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

MetumpX 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.
- MetumpX Package (Chapter 2) has tables for Software Selection Criteria, Categorization Scheme and Included Tools
- Software Installation Process (Chapter 3) describes how to install MetumpX.
- Microsoft Windows Support (Chapter 4) describes a step by step procedure to configure a virtual OS for MetumpX on Windows.
- Introduction to Metabolomics (Chapter 5) contain definitions and explanations on metabolomics, targeted and untargeted metabolomics, interactome, and study strategies.
- Metabolomics Workflow (Chapter 6) contain brief definitions and explanations on the entire software pipeline including necessary diagrams to facilitate learning of the field.

Contact

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

MetumpX Package

Software Selection Criteria

A specific criteria is used for enclosure of tools in MetumpX. Table shows this selection criteria. Softwares which are included are due to these specific reasons:

- Publication Date (later then 2010)
- Freeware License
- Linux based
- Offline

Sr. no	Software Tool Name	Pub. Date (2010-)	Offline	Linux based	Free	Installed
1	ProbMetab[1]	✓	✓	✓	✓	✓
2	intCor[2]	✓	✓	✓	✓	✓
3	CorrectOverloadedPeaks[3]	✓	✓	✓	✓	✓
4	iMet-Q[4]	✓	✓		✓	
5	AnalyzerPro[5]	✓	✓	✓		
6	ADAP-GC[6]	✓	✓	✓	✓	✓
7	ChromGenius[7]	✓	✓	✓		
8	ChromA[8]	✓		✓	✓	
9	XI3CMS[9]	✓	✓	✓	✓	✓
10	MET-COFEA[10]	✓	✓		✓	
11	MET-XAlign[11]	✓	✓		✓	
12	batchCorr[12]	✓	✓	✓	✓	✓
13	MZMine[13]	✓	✓	✓	✓	✓
14	MET-IDEA[14]	✓	✓		✓	
15	PyMS[15]	✓	✓	✓	✓	✓
16	MassCascade[16]		✓	✓	✓	
17	Mnova-MS[17]	✓	✓	✓		
18	XCMS[18]	✓	✓	✓	✓	✓
19	flagme[19]	✓	✓	✓	✓	✓
20	Elm. Metabolomics[20]	✓	✓		✓	
21	MetCirc[21]	✓	✓	✓	✓	✓
22	mzAccess[22]	✓	✓	✓		
23	SpeckTackle[23]		✓	✓	✓	
24	COMSPARI[24]		✓	✓	✓	
25	DAVE[25]	✓		✓	✓	
26	TargetSearch[26]	✓	✓	✓	✓	✓
27	HCor[27]	✓	✓	✓	✓	✓
28	MS-DIAL[28]	✓	✓		✓	
29	MaxEnt[29]	✓	✓	✓	✓	✓
30	RANSY/RAMSY[30]	✓		✓	✓	

MetumpX v2.1

31	UniDec[31]	✓	✓		✓	
32	Metab[32]	✓	✓	✓	✓	✓
33	decoMS2[33]	✓	✓	✓	✓	✓
34	GAGdecon[34]	✓	✓	✓	✓	✓
35	TMTc+[35]	✓	✓	✓		
36	BernetAI2018[36]		✓	✓	✓	
37	BUPID-Top-Down[37]	✓		✓	✓	
38	PicaudEtAl2018[38]	✓		✓	✓	
39	msXpertSuite[39]	✓	✓	✓	✓	✓
40	proFIA[40]	✓	✓	✓	✓	✓
41	apLCMS[41]	✓	✓	✓	✓	✓
42	yamss[42]	✓	✓	✓	✓	✓
43	cosmiq[43]	✓	✓	✓	✓	✓
44	AMDIS[44]	✓	✓		✓	
45	mzMatch-ISO[45]	✓	✓	✓	✓	✓
46	Elgen-MS[46]		✓	✓	✓	
47	MetaQuant[47]		✓	✓	✓	
48	CAMERA[48]	✓	✓	✓	✓	✓
49	MS-FLO[49]	✓		✓	✓	
50	JumPm[50]	✓	✓	✓		
51	FastChrom[51]	✓	✓		✓	
52	GridMass[52]		✓	✓	✓	
53	KMMDA[53]	✓	✓	✓	✓	✓
54	HayStack[54]	✓		✓	✓	
55	msPeak[55]	✓	✓	✓	✓	✓
56	GC-Analyizer[56]	✓	✓	✓		
57	MsXelerator[57]	✓	✓	✓		
58	MarkerLynx[58]	✓	✓	✓		
59	GCxCAnalyzer[59]	✓	✓	✓		
60	MetNorm[60]	✓	✓	✓	✓	✓
61	MetTailor[61]	✓	✓	✓	✓	✓
62	MetaPre[62]	✓		✓	✓	
63	NOREVA[63]	✓		✓	✓	
64	crmn[64]	✓	✓	✓	✓	✓
65	Normalizer[65]	✓		✓	✓	
66	metaX[66]	✓		✓	✓	
67	MetabR[67]		✓	✓	✓	
68	LowessNormalization[68]	✓	✓		✓	
69	MSPrep[69]	✓	✓	✓	✓	✓
70	Aloutput [70]	✓	✓		✓	
71	Ionwinze[71]	✓	✓		✓	
72	MPP [72]	✓	✓	✓		
73	R2DGC[73]	✓	✓	✓	✓	✓
74	mSPA[74]	✓	✓	✓	✓	✓
75	Maui-VIA[75]	✓	✓		✓	
76	SECIIMTools[76]	✓	✓	✓	✓	✓
77	BatMass[77]	✓	✓	✓	✓	✓
78	MetaboQC[78]	✓	✓	✓		
79	QCScreen[79]	✓	✓	✓	✓	✓
80	QC-RFSC[80]		✓	✓	✓	
81	mscompare[81]		✓	✓	✓	
82	PYQUAN[82]	✓	✓		✓	
83	MINMA[83]	✓	✓	✓	✓	✓
84	MetaboloDerivatizer[84]	✓	✓		✓	
85	SIRIUS[85]	✓	✓	✓	✓	✓
86	HAMMER[86]	✓	✓		✓	
87	ISDB-MN[87]		✓	✓	✓	

88	SweetSubstitute [88]	✓	✓		✓	
89	MetFrag[89]	✓		✓	✓	
90	MetFusion[90]	✓		✓	✓	
91	iontree[91]	✓	✓	✓	✓	✓
92	ACD/MS Fragmenter[92]		✓	✓	✓	
93	MassFrontier[93]	✓	✓	✓		
94	MWASTools[94]	✓	✓	✓	✓	✓
95	RegScan[95]		✓	✓	✓	
96	InCroMAP[96]	✓	✓	✓	✓	✓
97	PathVisio[97]	✓	✓	✓	✓	✓
98	CHem-SMP[98]	✓	✓	✓		
99	cPath[99]	✓	✓	✓	✓	✓
100	MetaMapp[100]		✓	✓	✓	
101	Mapping Tool[101]	✓	✓	✓	✓	✓
102	BLASTX [102]	✓		✓	✓	
103	PSSAlib [103]	✓	✓	✓		
104	iPath [104]	✓		✓	✓	
105	MetExplore [105]	✓		✓	✓	
106	CATABOL[106]		✓	✓	✓	
107	Paintomics[107]	✓		✓	✓	
108	ProteinLounge[108]	✓		✓	✓	
109	CellMLTools[109]		✓	✓	✓	
110	FCF[110]		✓	✓	✓	
111	PathPred[111]	✓		✓	✓	
112	SABIO-RK[112]	✓		✓	✓	
113	OptCom[113]		✓	✓	✓	
114	Subpathway-GM[114]	✓		✓	✓	
115	IPAVS[115]	✓		✓	✓	
116	GAM[116]	✓		✓	✓	
117	GLAMM[117]	✓		✓	✓	
118	PASMet[118]	✓		✓	✓	
119	NICElips[119]		✓	✓	✓	
120	MetaMapR[120]	✓		✓	✓	
121	PAPi[121]	✓	✓	✓	✓	✓
122	ReactPRED[122]	✓	✓	✓	✓	✓
123	JigCell[123]		✓	✓	✓	
124	Cell++[124]		✓	✓	✓	
125	NetNetteR[125]		✓	✓	✓	
126	MEMOSys[126]	✓		✓	✓	
127	MonaLisa[127]	✓	✓	✓	✓	✓
128	QSSPN[128]	✓	✓	✓	✓	✓
129	NetSeed[129]	✓		✓	✓	
130	tEFMA[130]	✓	✓	✓		
131	NetCmpt[131]	✓		✓	✓	
132	SED-ED [132]	✓	✓	✓	✓	✓
133	MetNetMaker [133]	✓	✓		✓	
134	RxnSim [134]	✓	✓	✓	✓	✓
135	MetExploreViz[135]	✓		✓	✓	
136	C.Calculator[136]		✓	✓	✓	
137	WebMetabase [137]	✓		✓	✓	
138	MoDify[138]	✓	✓	✓	✓	✓
139	MetaboSignal[139]	✓	✓	✓	✓	✓
140	JMassBalance[140]	✓	✓	✓	✓	✓
141	MetaDiff[141]	✓	✓	✓	✓	✓
142	FTA [142]		✓	✓	✓	
143	PySCeSToolbox[143]	✓	✓	✓	✓	✓
144	MEBS[144]	✓	✓	✓	✓	✓

MetumpX v2.1

145	Subpathway-GMIR[145]	✓	✓	✓	✓	✓
146	Prol[146]	✓		✓	✓	
147	MetaNetSam[147]	✓	✓	✓	✓	✓
148	IumpGEM [148]		✓	✓	✓	
149	redGEM[149]		✓	✓	✓	
150	CCC[150]	✓	✓	✓		
151	HuEtAl2018[151]	✓	✓	✓		
152	Kamneva[152]	✓	✓	✓	✓	✓
153	XeDetect[153]		✓	✓	✓	
154	ReactomePA[154]	✓	✓	✓	✓	✓
155	IPPAD[155]	✓		✓	✓	
156	MMinte[156]	✓	✓	✓	✓	✓
157	SED-ML[157]		✓	✓	✓	
158	NetCooperate[158]	✓		✓	✓	
159	PathRings[159]	✓		✓	✓	
160	IntPath [160]	✓		✓	✓	
161	EvoMS[161]	✓		✓	✓	
162	phraSED-ML[162]	✓	✓	✓	✓	✓
163	CARMEN[163]	✓		✓	✓	
164	ScrumPy[164]	✓	✓	✓	✓	✓
165	Pybrn[165]	✓	✓	✓	✓	✓
166	Pyabolism[166]	✓	✓	✓	✓	✓
167	KEGGREST[167]	✓	✓	✓	✓	✓
168	VIENNA-RNL[168]		✓	✓	✓	
169	Fbar[169]	✓	✓	✓	✓	✓
170	MetaCore[170]	✓		✓	✓	
171	PathwayLab[171]	✓	✓	✓		
172	MetScape[172]	✓	✓	✓	✓	✓
173	MPEA[173]	✓		✓	✓	
174	IMPaLA[174]	✓		✓	✓	
175	MBRole[175]	✓		✓	✓	
176	zeroSum[176]	✓	✓	✓	✓	✓
177	ChemRICH[177]	✓		✓	✓	
178	FELLA[178]	✓	✓	✓	✓	✓
179	BinChE[179]	✓		✓	✓	
180	MetaboliteDConv.[180]	✓	✓	✓	✓	✓
181	MetaboAnalyst[181]	✓	✓	✓	✓	✓
182	MapMan[182]		✓	✓	✓	
183	3Omics[183]	✓		✓	✓	
184	integrOmics[184]	✓	✓	✓	✓	✓
185	MetDisease[185]	✓	✓	✓	✓	✓
186	MetaBridge[186]	✓		✓	✓	
187	MetMask[187]	✓	✓	✓	✓	✓
188	ProMeTra [188]		✓	✓	✓	
189	MAGI[189]	✓		✓	✓	
190	KPIC2[190]	✓	✓	✓	✓	✓
191	MarVis-Suite[191]	✓	✓		✓	
192	MSClust[192]		✓	✓	✓	
193	MetMSLine[193]	✓	✓	✓	✓	✓
194	SimExTargid [194]	✓	✓	✓	✓	✓
195	MetaboliteDetector [195]		✓	✓	✓	
196	specmine[196]	✓	✓	✓	✓	✓
197	W4M [197]	✓		✓	✓	
198	MeltDB [198]	✓	✓		✓	
199	xMSAnalyzer [199]	✓	✓	✓	✓	✓
200	ChromaTOF[200]	✓	✓	✓		
201	MetabolomeExpress[201]	✓		✓	✓	

202	Metabox [202]	✓	✓	✓	✓	✓
203	PiMP[203]	✓		✓	✓	
204	MET-COFEI[204]	✓	✓		✓	
205	MAIT [205]	✓	✓	✓	✓	✓
206	BinVestigate[206]	✓		✓	✓	
207	CEU Mass Mediator [207]	✓		✓	✓	
208	MAGMa[208]	✓		✓	✓	
209	CSI:FingerID [209]	✓		✓	✓	
210	MS2LDA SUPPORT [210]	✓		✓	✓	
211	MetExtract [211]	✓	✓		✓	
212	T-Biolinfo [212]	✓		✓	✓	
213	MetAlign[213]	✓	✓		✓	
214	CFM-ID [214]	✓		✓	✓	
215	Ideom [215]		✓	✓	✓	
216	AStream [216]		✓	✓	✓	
217	PUTMEDID-LCMS[217]	✓	✓	✓	✓	✓
218	DECOMP [218]		✓	✓	✓	
219	MetiTTree[219]	✓		✓	✓	
220	MIA [220]	✓	✓	✓	✓	✓
221	MFSearcher[221]	✓	✓		✓	
222	ChemDistiller[222]		✓	✓	✓	
223	MSeasy[223]	✓	✓	✓	✓	✓
224	SIMPLE[224]		✓	✓	✓	
225	MAVEN[225]		✓	✓	✓	
226	SpectConnect [226]	✓		✓	✓	
227	RAMClustR[227]	✓	✓	✓	✓	✓
228	Molfind[228]	✓	✓	✓	✓	✓
229	MS2Analyzer [229]	✓	✓	✓	✓	✓
230	MS-FINDER [230]	✓	✓		✓	
231	geoRge[231]	✓	✓	✓	✓	✓
232	MetFamily[232]	✓		✓	✓	
233	eRAH[233]	✓	✓	✓	✓	✓
234	IsoMS [234]	✓	✓		✓	
235	MetDIA [235]		✓	✓	✓	
236	iMET[236]	✓		✓	✓	
237	MIDAS[237]		✓	✓	✓	
238	InterpretMSSpectrum [238]	✓	✓	✓	✓	✓
239	AssayR[239]	✓	✓	✓	✓	✓
240	MCID[240]	✓		✓	✓	
241	compMS2Miner[241]	✓		✓	✓	
242	MetShot [242]	✓	✓	✓	✓	✓
243	MINE[243]	✓		✓	✓	
244	NP-StructurePred. [244]	✓	✓	✓		
245	MetaboSearch[245]	✓	✓	✓	✓	✓
246	ALLOCator[246]	✓		✓	✓	
247	PROFANCY[247]	✓	✓		✓	
248	SpiderMass[248]	✓	✓		✓	
249	MZedDB[249]	✓		✓	✓	
250	BinBase [250]		✓	✓	✓	
251	PowerGet[251]	✓	✓		✓	
252	AMDORAP [252]	✓	✓	✓	✓	✓
253	MBIdent [253]	✓	✓	✓	✓	✓
254	peakANOVA[254]	✓	✓	✓	✓	✓
255	MI-Pack[255]	✓	✓	✓	✓	✓
256	SetupX [256]		✓	✓	✓	
257	FeatureFinderMetabo [257]		✓	✓	✓	
258	MassMetaSite[258]		✓	✓	✓	

259	MetDNA[259]	✓		✓	✓	
260	DASI[260]	✓	✓	✓		
261	MetaboList[261]	✓	✓	✓	✓	✓
262	MetaMS[262]	✓	✓	✓	✓	✓
263	SIEVE[263]	✓	✓	✓		
264	SimMet[264]	✓		✓	✓	
265	Apex[265]	✓	✓	✓		
266	Nontarget[266]	✓	✓	✓	✓	✓
267	NIST MS Search[267]	✓	✓		✓	
268	EI-Maven [268]	✓	✓	✓	✓	✓

NOTE The Authors have relaxed the selection criteria and have included 103 tools out of 268 in MetumpX. The remaining tools either does not lie on the basic selection criteria or have following notable issues:

- Software is free but its dependencies are paid
- Obsolete or not available
- Credentials or Subscription needed to avail the software

Software Categorization Scheme

Metabolomics Softwares are mainly workflows and one software can lie in more then one metabolomic pipeline category. Table shows a software categorization scheme. The software is placed in the lowest category it is present.

Sr. No.	Software	Noise Filtering	Chromatogram Alignment	Peak Alignment	Retention Time Correction	Spectral Deconvolution	Peak Detection	Data Normalization	Statistical Analysis	Quality Control	Metabolite Quantification	Data Imputation	In-silico Fragmentation	Metabolite Identification	Spectral Visualization	Clustering Analysis	mGWAS	Mapping	Network Analysis	Enrichment Analysis	Integrative Analysis
1	C.O.Peaks[3]	✓																			
2	specmine[196]	✓						✓	✓	✓		✓		✓							
3	intCor[2]	✓	✓	✓																	
4	batchCorr[12]		✓																		
5	mSPA[74]			✓																	
6	AMDORAP[252]	✓	✓	✓			✓														
7	MI-Pack[255]			✓																	
8	Metab[32]						✓														
9	decoMS2[28]						✓									✓					
10	GAGdecon[34]						✓														
11	msXpertSuite[39]						✓														
12	ADAP-GC[6]						✓	✓													
13	MaxEnt[29]						✓														
14	HCor[27]						✓														
15	MetMSLine[193]	✓	✓				✓									✓					
16	X13CMS[9]		✓				✓														
17	proFIA[40]	✓					✓									✓					
18	cosmiq[43]		✓				✓								✓						
19	mzMatch-ISO[45]						✓										✓				
20	PyMS[15]						✓										✓				
21	TargetSearch[26]						✓										✓				
22	msPeak[55]						✓										✓				
23	Metabox[202]						✓	✓											✓	✓	✓
24	MetNorm[60]	✓					✓														
25	crmn[64]	✓					✓														
26	KMMDA[53]						✓	✓													
27	MSPrep[69]							✓	✓							✓					
28	flagme[9]	✓					✓	✓	✓								✓				
29	SECIMTools[76]							✓	✓	✓											
30	QCScreen[79]									✓											
31	MetTailor[61]		✓												✓						
32	MetaQuant[47]							✓	✓						✓			✓			
33	apLCMS[41]		✓					✓							✓			✓			
34	MINMA[83]														✓						
35	SIRIUS[85]															✓					
36	iontree[91]															✓					
37	MetShot[242]																✓				
38	Molfind[228]																✓				
39	MIA[220]																✓				
40	MetaMS[262]																✓				
41	MSeasy[223]																✓				

MetumpX v2.1

42	RAMClustR[227]									✓									
43	MetaboSearch[245]									✓									
44	EI-Maven[268]									✓									
45	geoRge[231]									✓									
46	eRAH[233]									✓									
47	MetaboList[261]t									✓									
48	IMSSpectrum[238]									✓									
49	AssayR[239]									✓									
50	MS2Analyzer[229]									✓									
51	Nontarget[266]									✓									
52	MetMask[187]									✓									
53	peakANOVA[254]									✓									
54	PUTMEDID-LCMS[217]									✓									
55	SimExTargid[94]	✓	✓	✓		✓	✓		✓		✓	✓	✓						
56	MetaboAnalyst[181]					✓	✓				✓	✓	✓			✓	✓		
57	CAMERA[48]					✓					✓	✓	✓						
58	KPIC2[190]		✓		✓			✓				✓							
59	MWASTools[94]															✓			
60	InCroMAP[96]																✓		
61	PathVisio[97]																✓		
62	Mapping Tool[101]																✓		
63	ChemDistiller[222]																✓		
64	PycesToolbox[143]																✓		
65	MetaDiff[141]																✓		
66	ReactomePA[154]																✓		
67	MEBS[144]																✓		
68	RxnSim[134]																✓		
69	phraSED-ML[162]																✓		
70	ScrumPy[164]																✓		
71	Subpathway-GMir[145]																✓		
72	Kamneva[152]																✓		
73	MetaboSignal[139]																✓		
74	ReactPRED[122]																✓		
75	Mmine[156]																✓		
76	PAPI[121]																✓		
77	MetaNetSam[147]																✓		
78	Fbar[169]																✓		
79	SED-ED[132]																✓		
80	QSSPN[128]																✓		
81	JMassBalance[140]																✓		
82	Pyabolism[166]																✓		
83	Pybrn[165]																✓		
84	MonaLisa[127]																✓		
85	MoDitify[138]																✓		
86	C.Calculator[136]																✓		
87	KEGGREST[167]																✓		
88	integrOmics[184]																	✓	
89	MetScape[172]																✓	✓	✓
90	cPath[99]																✓	✓	✓
91	MetDisease[185]																✓	✓	✓
92	zeroSum[176]																✓		
93	FELLA[178]																✓		
94	MetaboliteIDConv.[180]																✓		
95	MZMine[3]	✓	✓			✓	✓	✓		✓						✓	✓	✓	
96	XCMS[18]		✓	✓			✓	✓		✓		✓			✓	✓	✓		
97	yamss[42]		✓					✓											
98	R2DGC[73]		✓	✓			✓		✓						✓	✓			

99	BatMass[7]								✓				✓						
100	MAIT[205]		✓				✓		✓			✓	✓	✓	✓				✓
101	ProbMetab[1]	✓	✓				✓		✓			✓	✓	✓	✓				
102	MetCirc[21]			✓								✓	✓	✓	✓				
103	xMSAnalyzer[199]		✓			✓	✓			✓			✓	✓	✓				

MetumpX Packaged Tools

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
Data Pre-processing				
Noise Filtering				
1	C.O.Peaks[3]	4.8	1.2.17	2019
2	specmine[196]	15.2	2.0.3	2018
3	intCor[2]	3.0	1.03.0	2014
Chromatogram Alignment				
4	batchCorr[12]	17.2	0.2.1	2018
Peak Alignment				
5	mSPA[74]	0.1	1.0.0	2011
6	AMDORAP[252]	40.9	1.0.6	2012
7	MI-Pack[255]	28.9	1.0.0	2015
Spectral Deconvolution				
8	Metab[32]	3.2	1.18.0	2019
9	decoMS2[28]	5.2	0.1.0	2013
10	GAGdecon[34]	0.1	1.0.0	2018
11	msXpertSuite[39]	0.7	4.1.0	2019
12	ADAP-GC[6]	3.2	3.0.0	2017
13	MaxEnt[29]	12.4	3.4.1	2017
Retention Time Correction				
14	HCor[27]	0.1.0	1.01.0	2014
Data Processing				
Peak Detection				
15	MetMSLine[193]	1.6	1.2.1	2017
16	X13CMS[9]	0.1	1.4.0	2014

17	proFIA[40]	2.0	1.10.0	2019
18	cosmiq[43]	17.5	1.18.0	2019
19	mzMatch-ISO[45]	0.1	1.0.0	2019
20	PyMS[15]	0.45	1.0.0	2012
21	TargetSearch[26]	0.69	1.40.3	2019
22	msPeak[55]	32.3	1.0.0	2013
Data Normalization				
23	Metabox[202]	53.0	1.2.0	2016
24	MetNorm[60]	0.5	0.1.0	2015
25	crmn[64]	2.4	0.0.20	2014
Statistical Analysis				
26	KMMDA[53]	0.5	1.0.0	2018
27	MSPrep[69]	1.0	0.0.2	2018
28	flagme[19]	22.0	1.40.0	2019
Quality Control				
29	SECIMTools[76]	0.6	1.0.0	2018
30	QCScreen[79]	12.1	1.0.0	2018
Metabolite Quantification				
31	MetTailor[61]	0.9	2.0.0	2015
32	MetaQuant[47]	12.0	1.0.0	2010
33	apLCMS[41]	14.1	6.6.3	2019
Data Imputation				
34	MINMA[83]	2.6	0.1.0	2017
In-silico Fragmentation				
35	SIRIUS[85]	37.0	4.0.1	2019
36	iontree[91]	0.9	1.23.1	2018
Metabolite Identification				
37	MetShot[242]	0.9	0.3.2	2018
38	Molfind[228]	36.2	1.9.0	2013
39	MIA[220]	2.7	1.0.0	2017

40	MetaMS[262]	3.7	1.20.0	2019
41	MSeasy[223]	5.8	5.3.3	2013
42	RAMClustR[227]	51.6	0.4.1	2019
43	MetaboSearch[245]	45.7	1.0.0	2012
44	EI-Maven[268]	91.4	9.0.0	2019
45	geoRge[231]	13.7	1.0.0	2017
46	eRAH[233]	3.7	1.1.0	2018
47	MetaboList[261]	0.3	1.4.0	2019
48	I.MSSpect.[238]	0.2	1.2.0	2018
49	AssayR[239]	71.6	0.0.9	2017
50	MS2Analyzer[229]	3.0	2.1.0	2016
51	Nontarget[266]	3.5	1.9.0	2019
52	MetMask[187]	4.43	0.5.3	2017
53	peakANOVA[254]	0.5	1.0.0	2015
54	PUTMEDID[217]	191.2	1.0.0	2011
Data Clustering Analysis				
55	SimExTargid[194]	49.8	0.2.1	2017
56	MetaboAnalyst[181]	49.3	4.0.0	2019
57	CAMERA[48]	2.1	1.40.0	2019
58	KPIC2[190]	15.7	2.4.0	2019
Data Network Analysis				
mGWAS				
59	MWASTools[94]	56.3	1.8.0	2019
Metabolite Mapping				
60	InCroMAP[96]	54.6	1.5.0	2012
61	PathVisio[97]	17.6	3.3.0	2019
62	Mapping Tool[101]	15.0	1.3.0	2013
63	ChemDistiller[222]	133.2	0.1.0	2018
Metabolic Network Analysis				
64	PySCeSToolbox [143]	5.1	0.9.6	2018

65	MetaDiff[141]	6.0	1.0.0	2016
66	ReactomePA[154]	11.6	1.28.0	2019
67	MEBS[144]	73.2	1.0.0	2017
68	RxnSim[134]	21.6	1.0.3	2013
69	phraSED-ML[162]	1.8	1.0.3	2018
70	ScrumPy[164]	1093.9	1.0.0	2018
71	Subpathway[145]	1.7	3.0.0	2013
72	Kamneva 2016[152]	101.8	1.0.0	2016
73	MetaboSignal[139]	169.1	1.14.0	2019
74	ReactPRED[122]	52.4	1.0.0	2016
75	Mminte[156]	52.2	1.0.0	2017
76	PAPi[121]	0.6	1.24.0	2019
77	MetaNetSam[147]	3.8	1.1.0	2015
78	Fbar[169]	2.1	0.5.2	2018
79	SED-ED[132]	6.4	2.2.3	2016
80	QSSPN[128]	2.1	1.0.0	2015
81	JMassBalance[140]	3.7	1.0.0	2013
82	Pyabolism[166]	0.4	1.0.0	2017
83	Pybrn[165]	0.6	0.4.3	2016
84	MonaLisa[127]	17.8	5.1.0	2016
85	MoDentify[138]	0.6	0.99.0	2019
86	C.Calculator[136]	16.4	1.0.0	2010
87	KEGGREST[167]	13.67	1.24.0	2019
Data Integration Analysis				
88	integrOmics[184]	0.1	2.55.0	2012
89	MetScape[172]	17.2	3.1.3	2017
90	cPath[99]	1.32	2.0.0	2019
91	MetDisease[185]	15.2	1.1.0	2014
Data Enrichment Analysis				
92	zeroSum[176]	1.0	2.0.0	2019

93	FELLA[178]	3.1	1.4.1	2019
94	M.IDConvertor[180]	0.2	1.0.0	2010
Data Visualization				
95	MZMine[13]	148.7	2.0.0	2019
96	XCMS[18]	3.5	3.7.1	2018
97	yamss[42]	15.5	1.9.1	2018
98	R2DGC[73]	0.6	1.0.3	2017
99	BatMass[77]	0.1	0.3.0	2018
100	MAIT[205]	36.2	1.18.0	2019
101	ProbMetab[1]	0.2	1.0.0	2013
102	MetCirc[21]	5.0	1.14.0	2017
103	xMSAnalyzer[199]	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 MetumpX require a proper internet connection to proceed, otherwise the installation terminates.

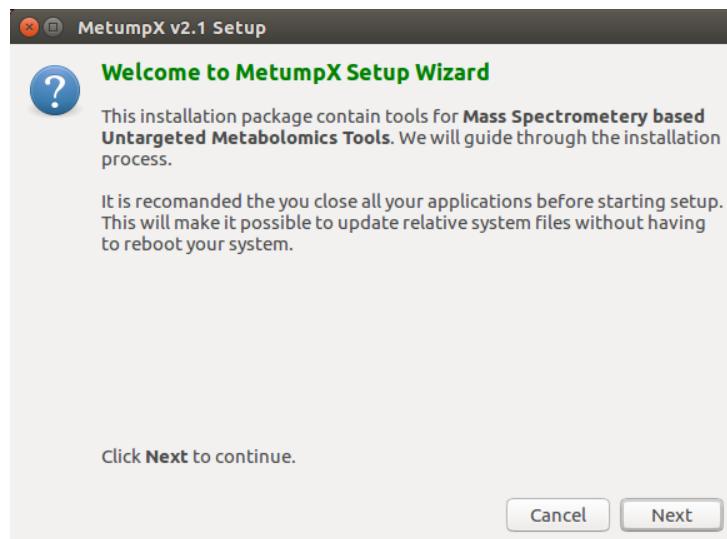


Figure 1: Installation Welcome Screen

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

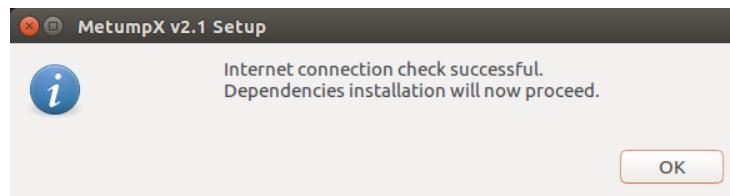


Figure 2: Internet Check Successful Screen

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

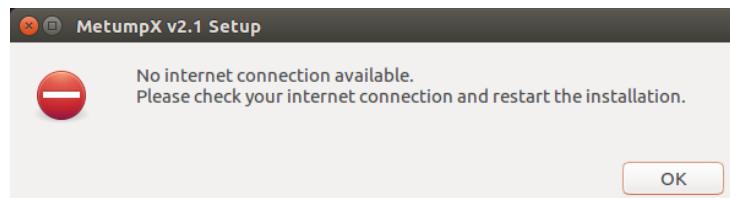


Figure 3: Internet Check failed screen

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

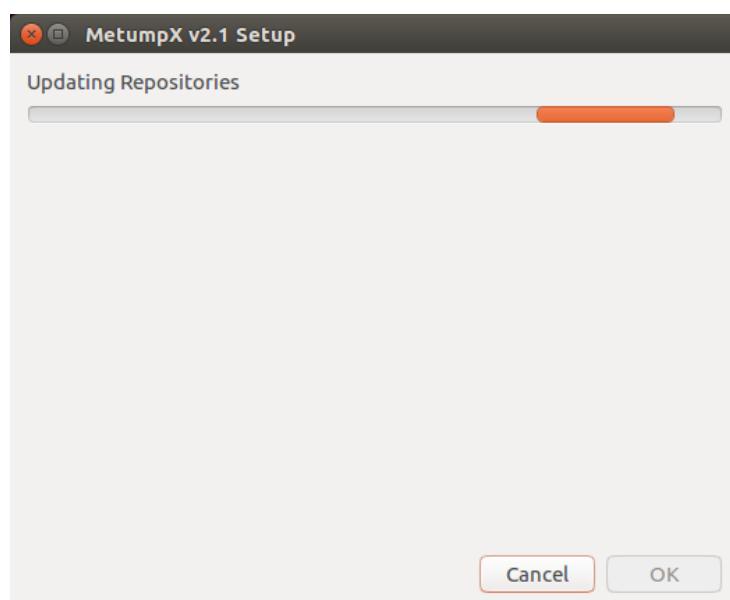


Figure 4: Installation Screen

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

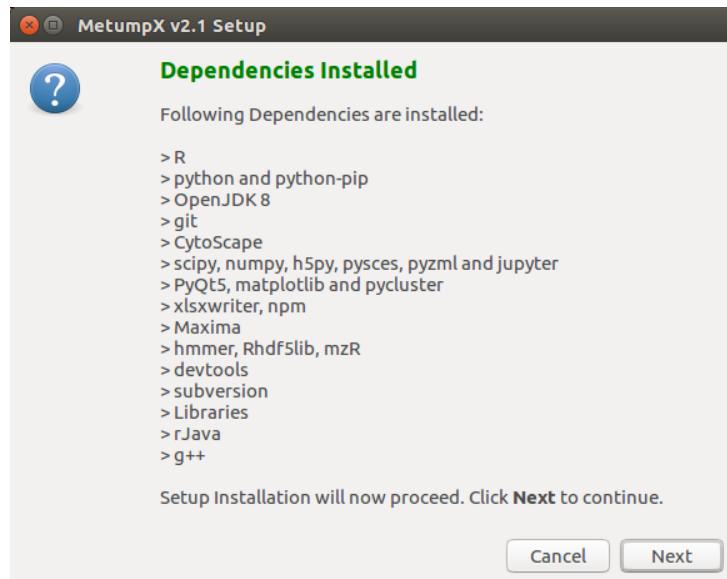


Figure 5: Dependencies installation finished screen

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

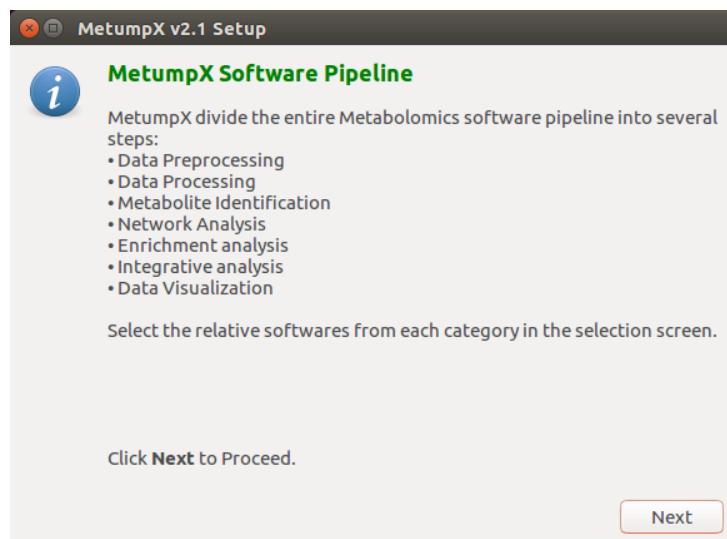


Figure 6: Pipeline information screen

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

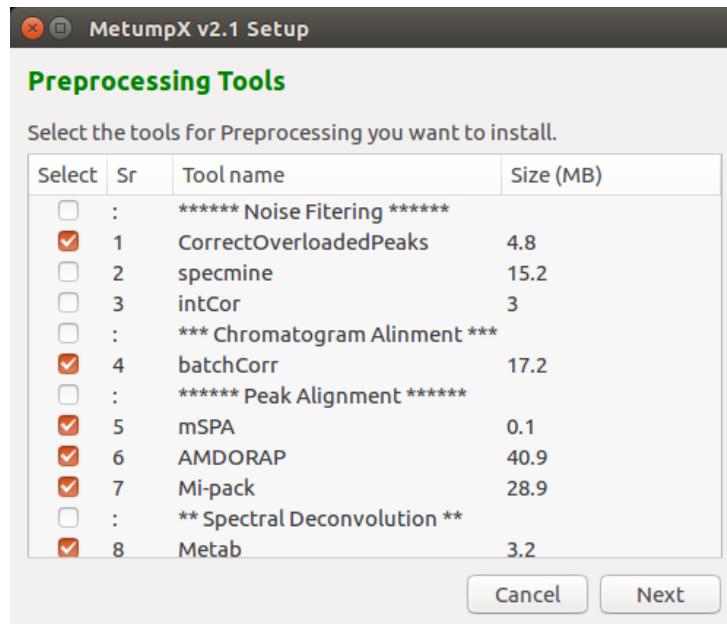


Figure 7: Preprocessing tools selection screen

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

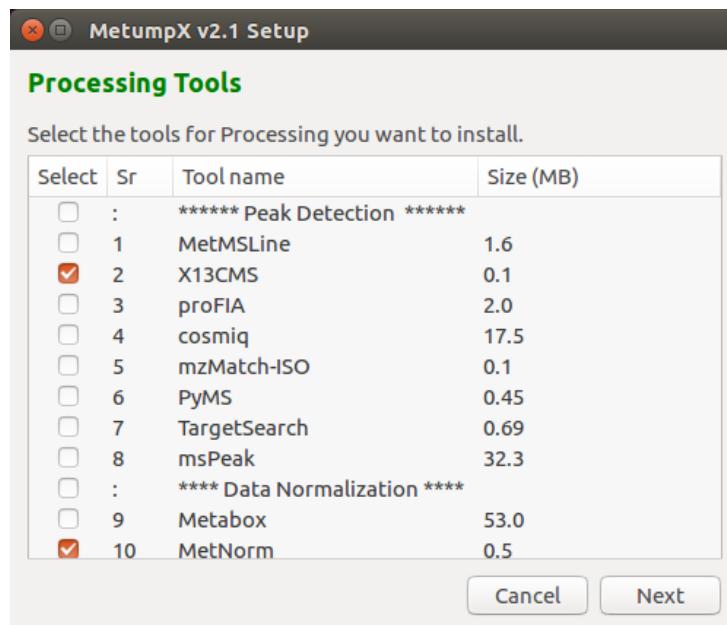


Figure 8: Processing tools selection screen

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

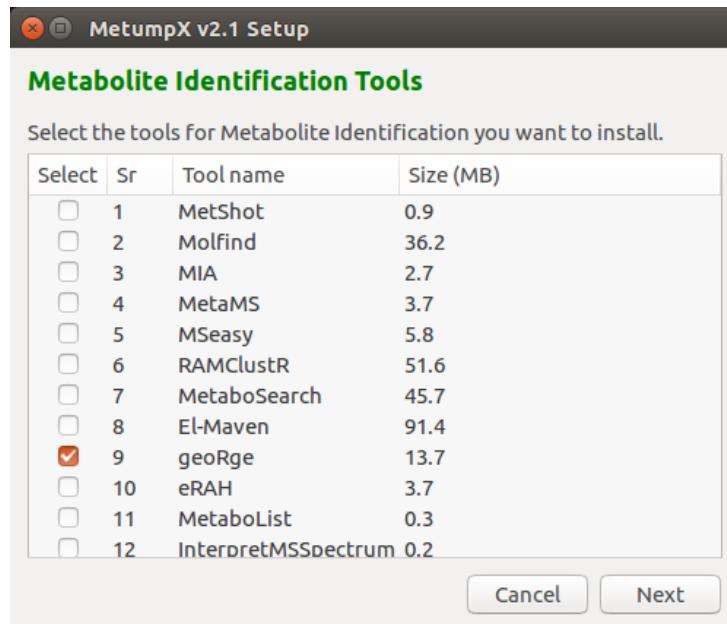


Figure 9: Metabolite identification tools selection screen

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

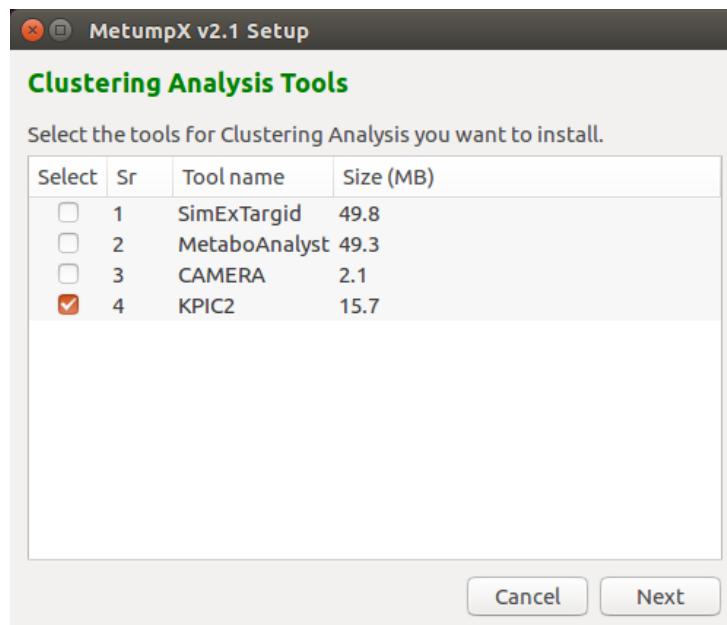


Figure 10: Clustering analysis tools selection screen

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

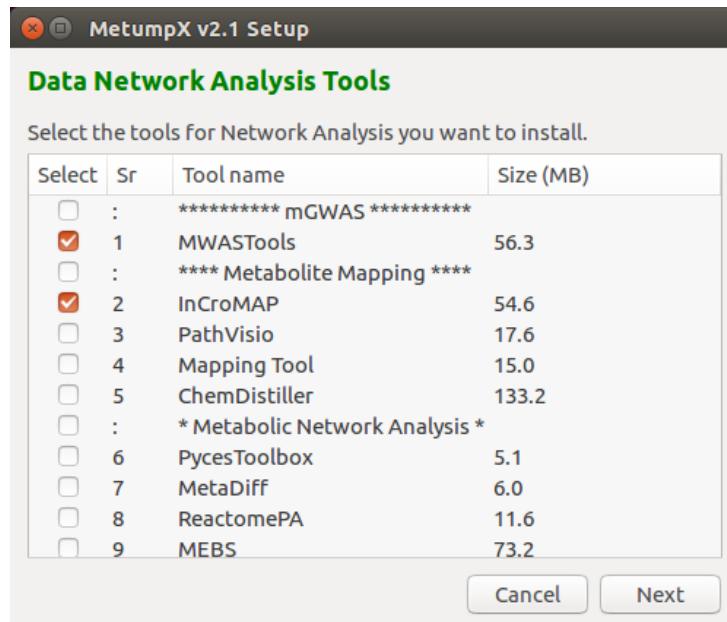


Figure 11: Data network analysis tools selection screen

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

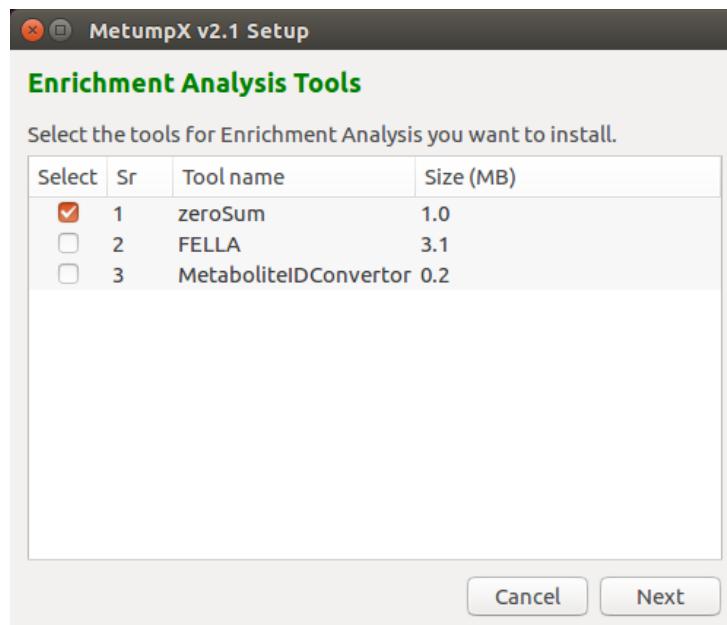


Figure 12: Enrichment analysis tools selection screen

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

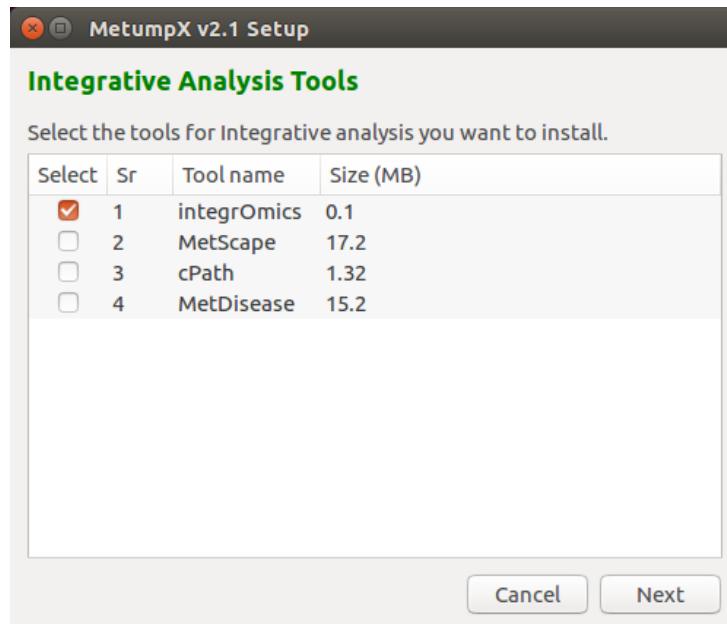


Figure 13: Integrative analysis tools selection screen

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

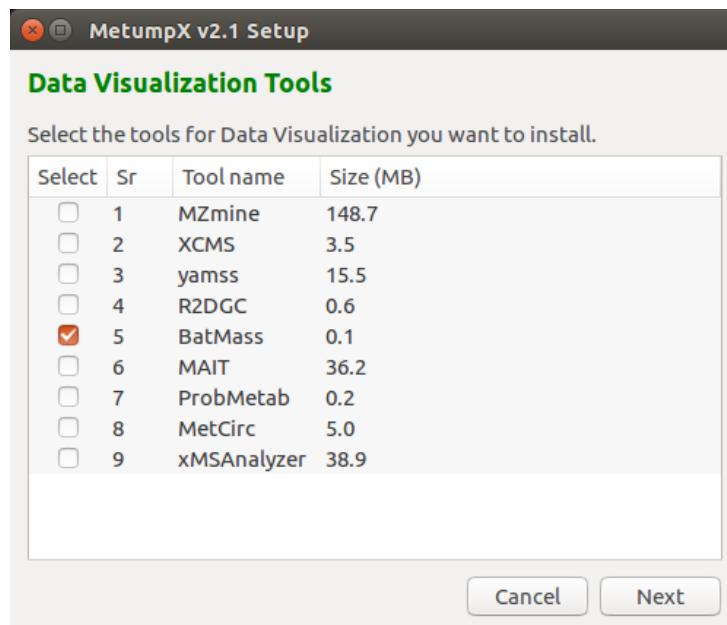


Figure 14: Data visualization tools selection screen

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

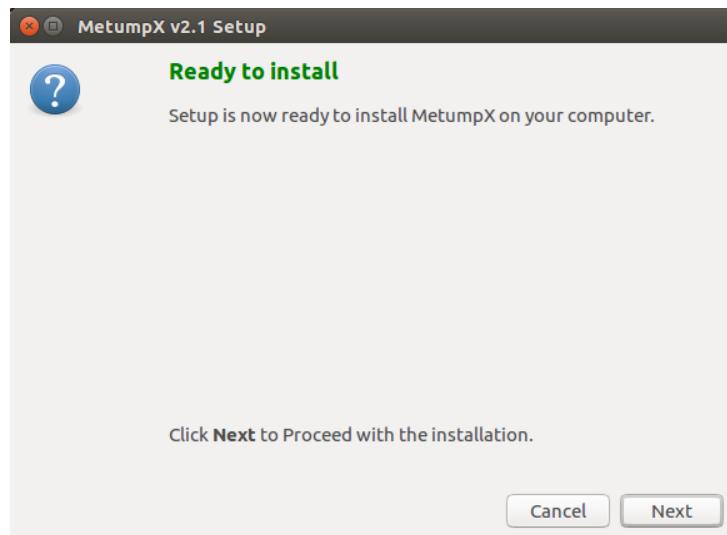


Figure 15: Ready to Install Screen

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

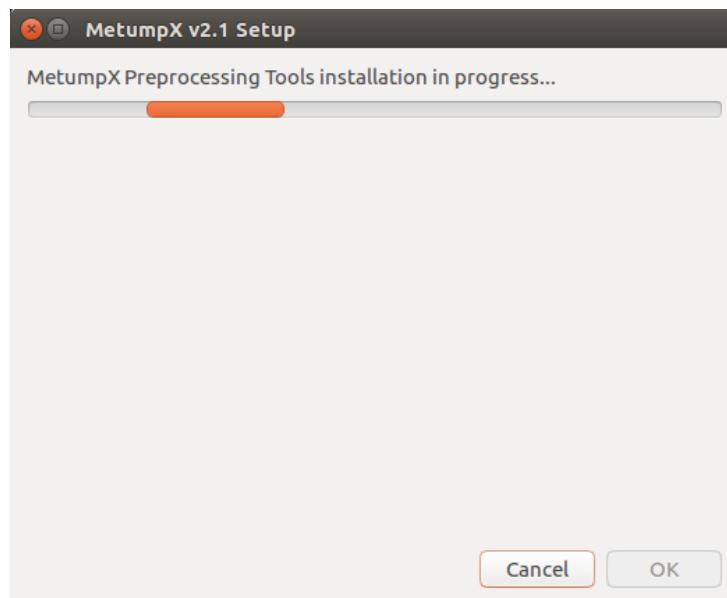


Figure 16: Tools installation progress screen

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

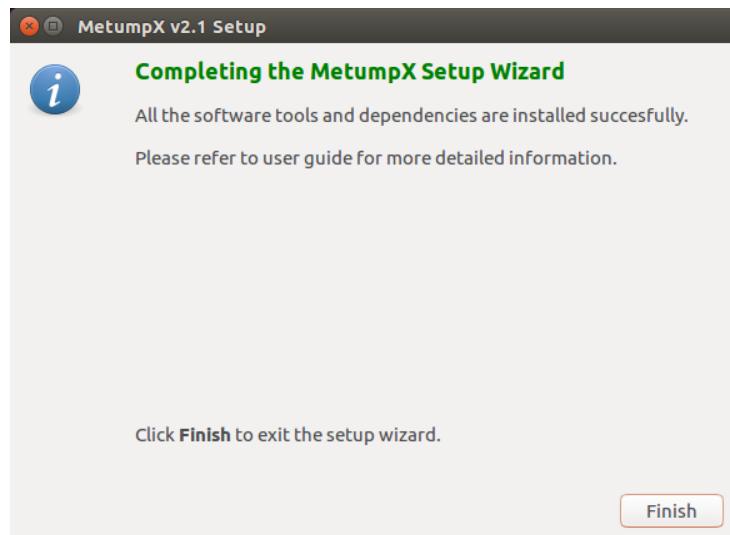


Figure 17: Finishing Installation Screen

Microsoft Windows Support

Installing MetumpX software on Microsoft Windows

MetumpX is software package for Linux, specifically Ubuntu users, but Windows users can use MetumpX 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://releases.ubuntu.com/18.04.3/ubuntu-18.04.3-desktop-amd64.iso>

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

<http://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.

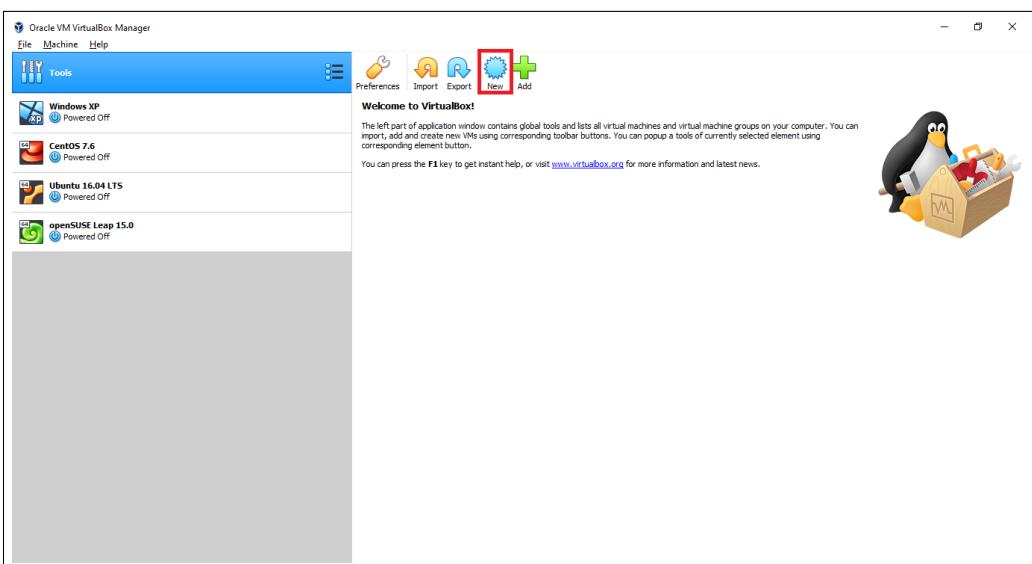


Figure 1: New Virtual OS

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

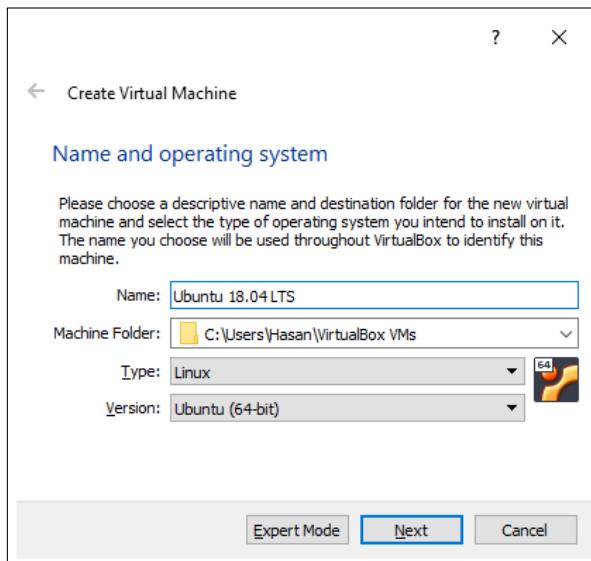


Figure 2: Assigning Name

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.

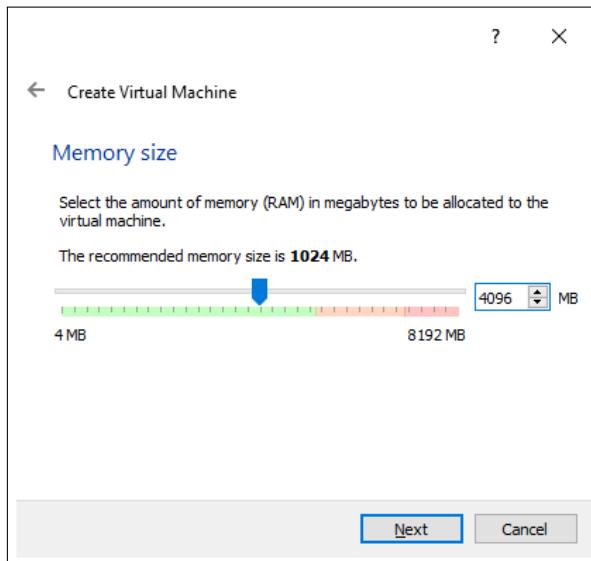


Figure 3: Memory Allocation

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**.

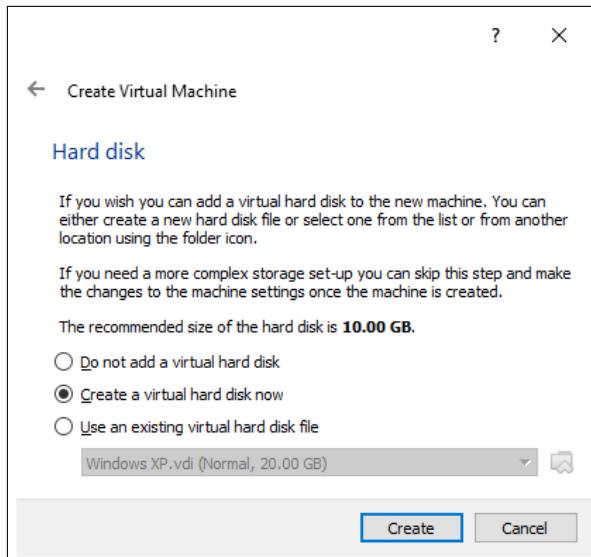


Figure 4: Creating a virtual hard disk

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

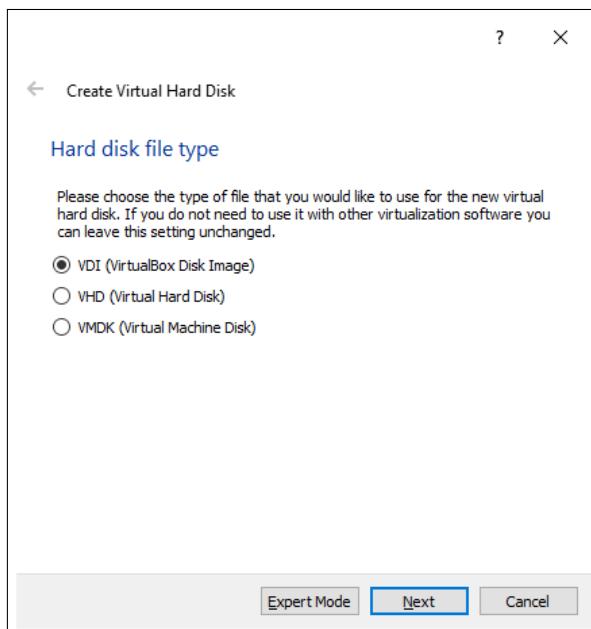


Figure 5: Hard disk file type

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.*

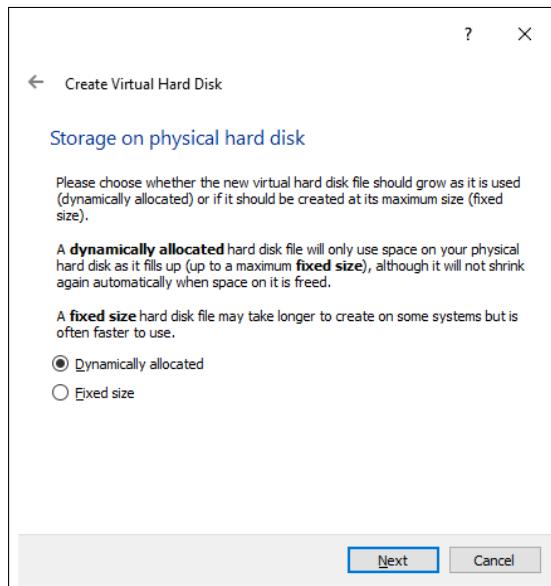


Figure 6: Storage type

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

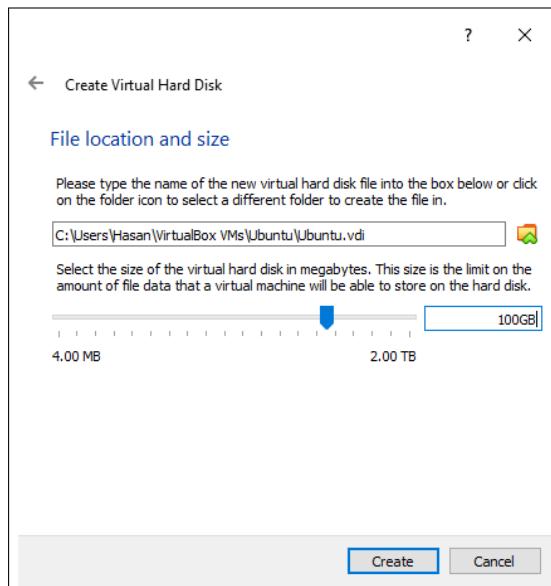


Figure 7: Hard disk size

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

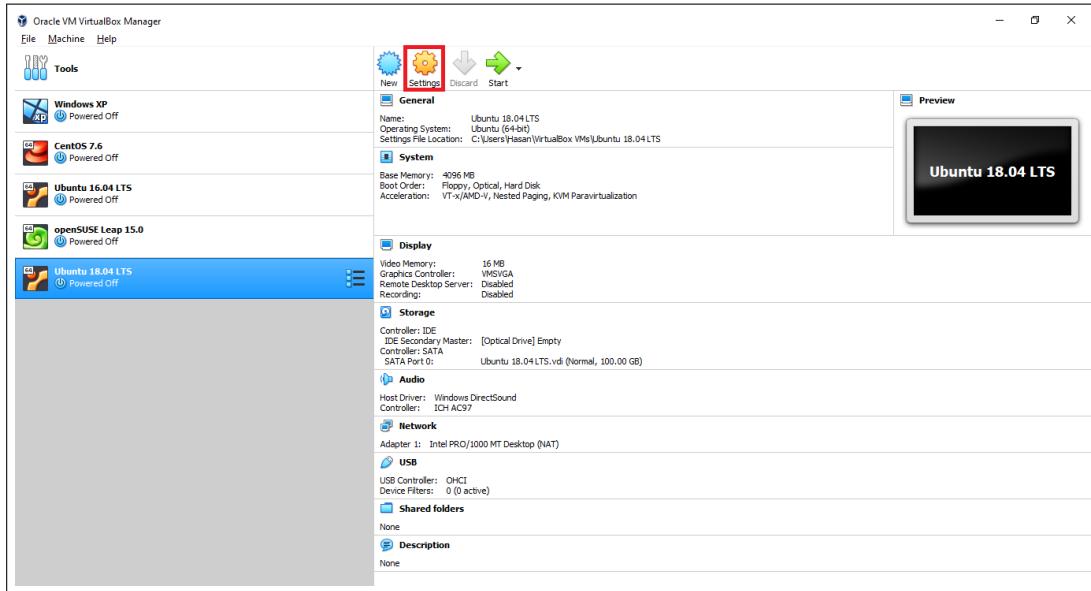


Figure 8: Select Settings

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

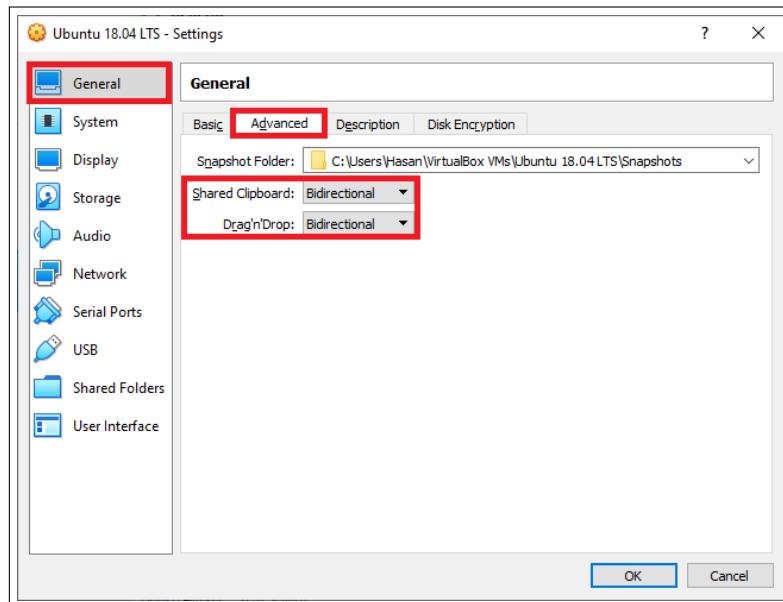


Figure 9: Advanced Settings

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**.

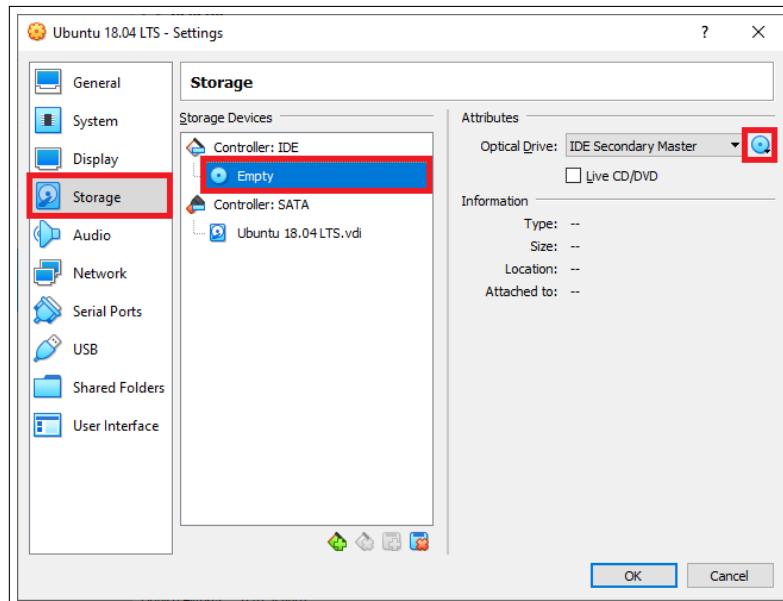


Figure 10: Providing Image file

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

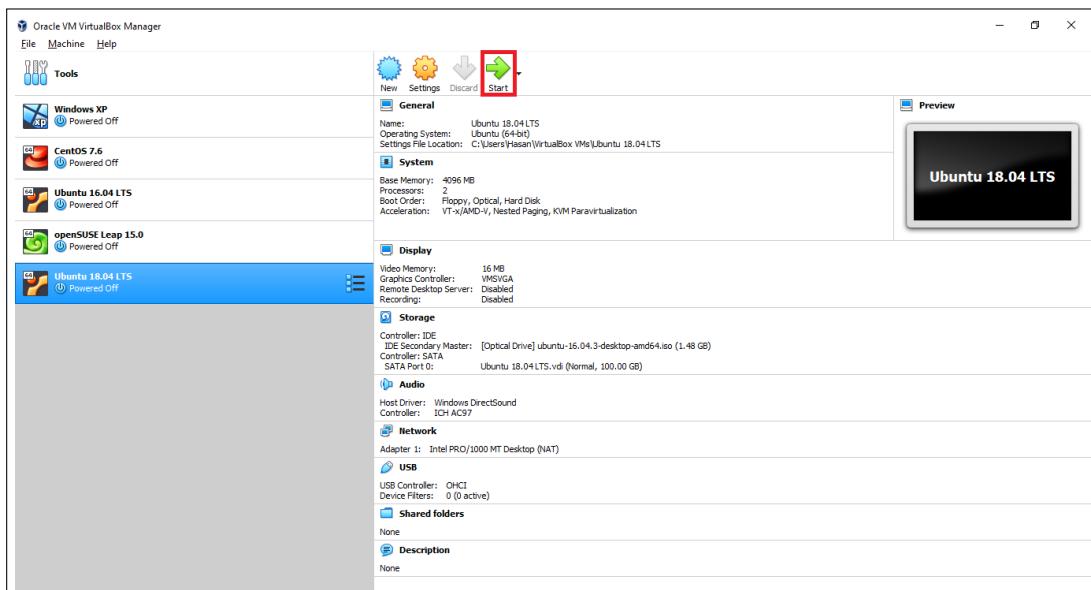


Figure 11: Starting Virtual OS

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

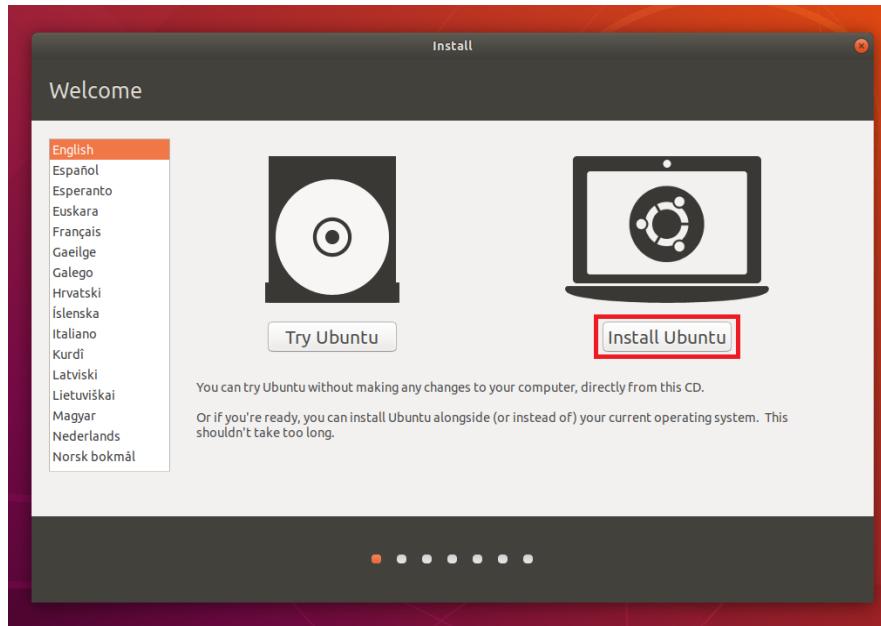


Figure 12: Installing Ubuntu

16. Continue with Normal Installation.

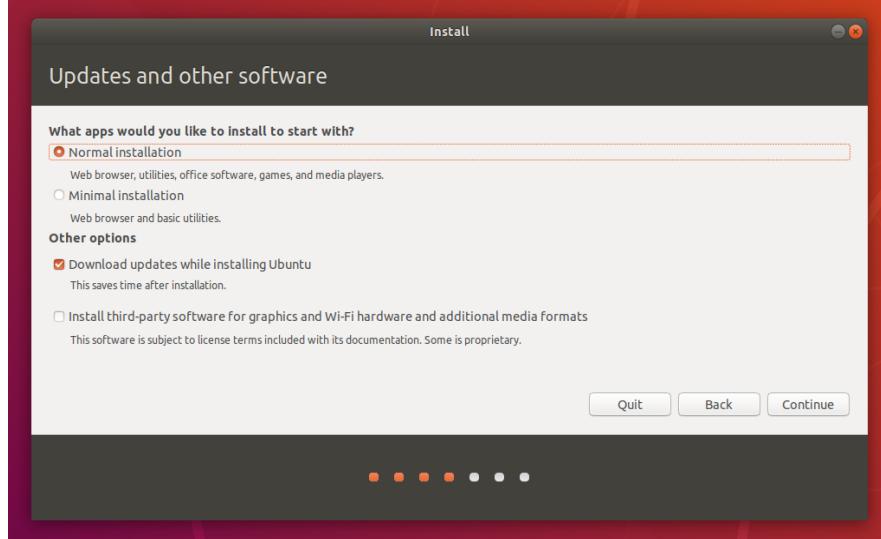


Figure 13: Update Screen

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

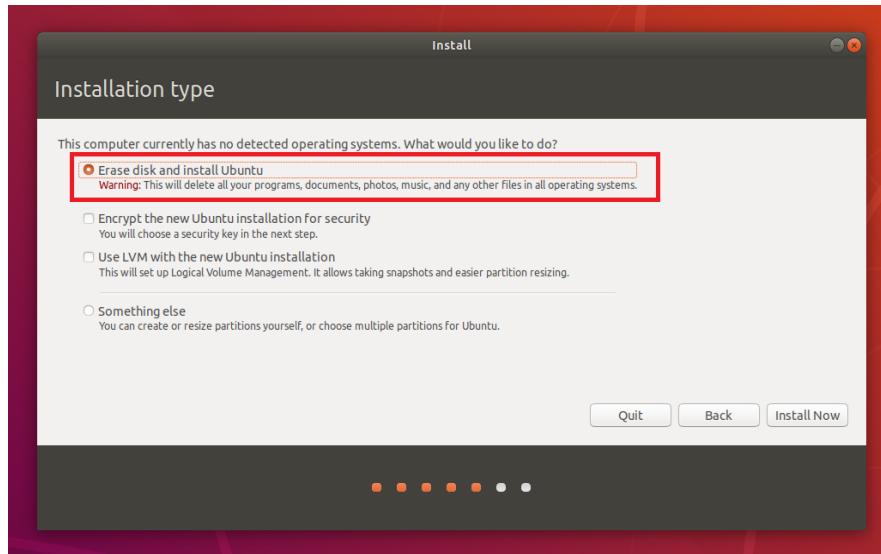


Figure 14: Installation type Screen

18. Select Continue.

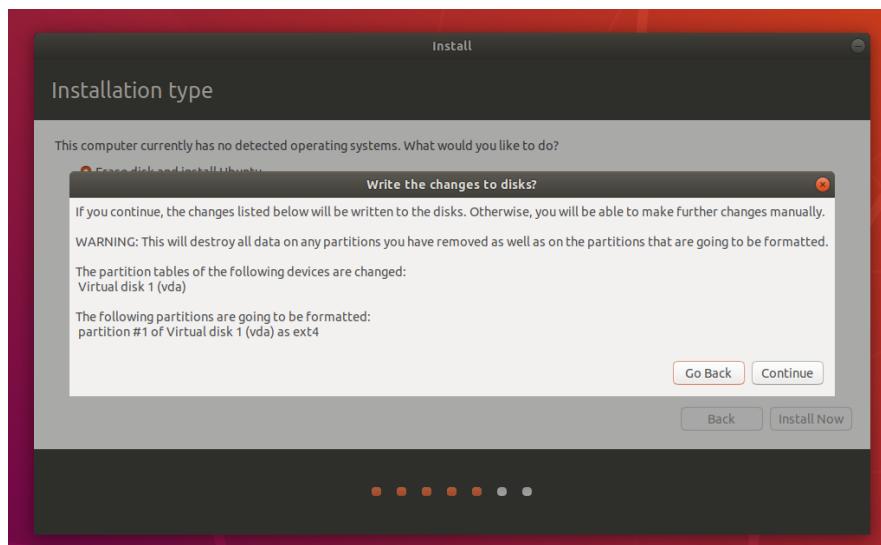


Figure 15: Confirmation Screen

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

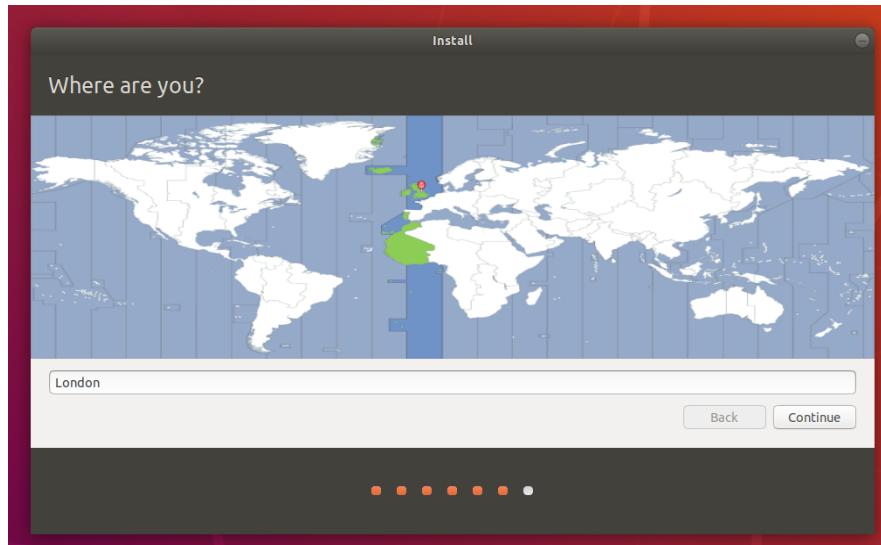


Figure 16: Location selection screen

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

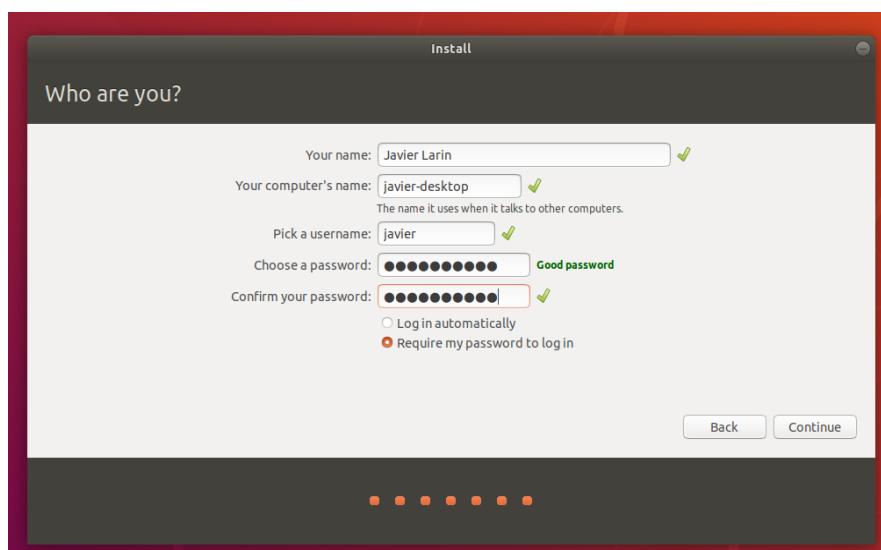


Figure 17: Intro Screen

21. Installation will Continue.

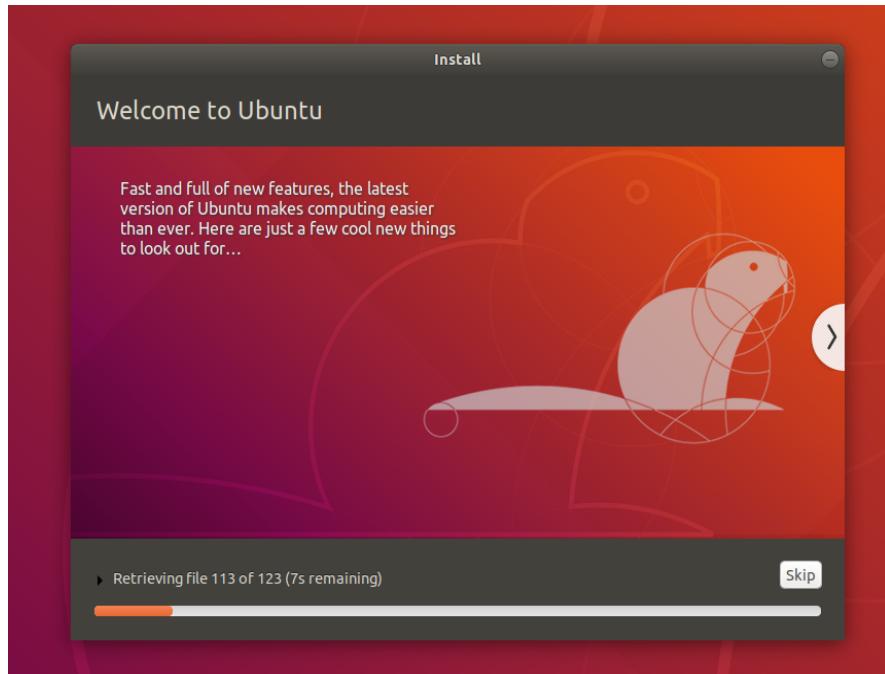


Figure 18: Installation Screen

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

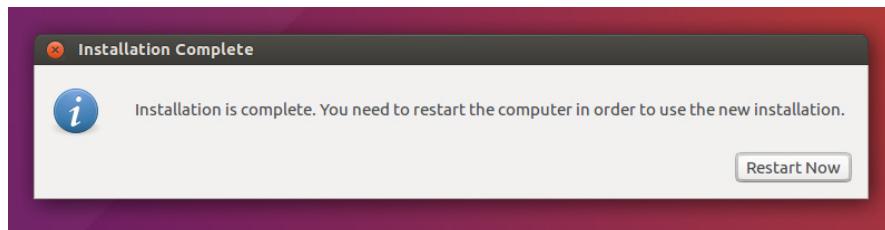


Figure 19: Complete Installation Screen

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.[269]

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**.

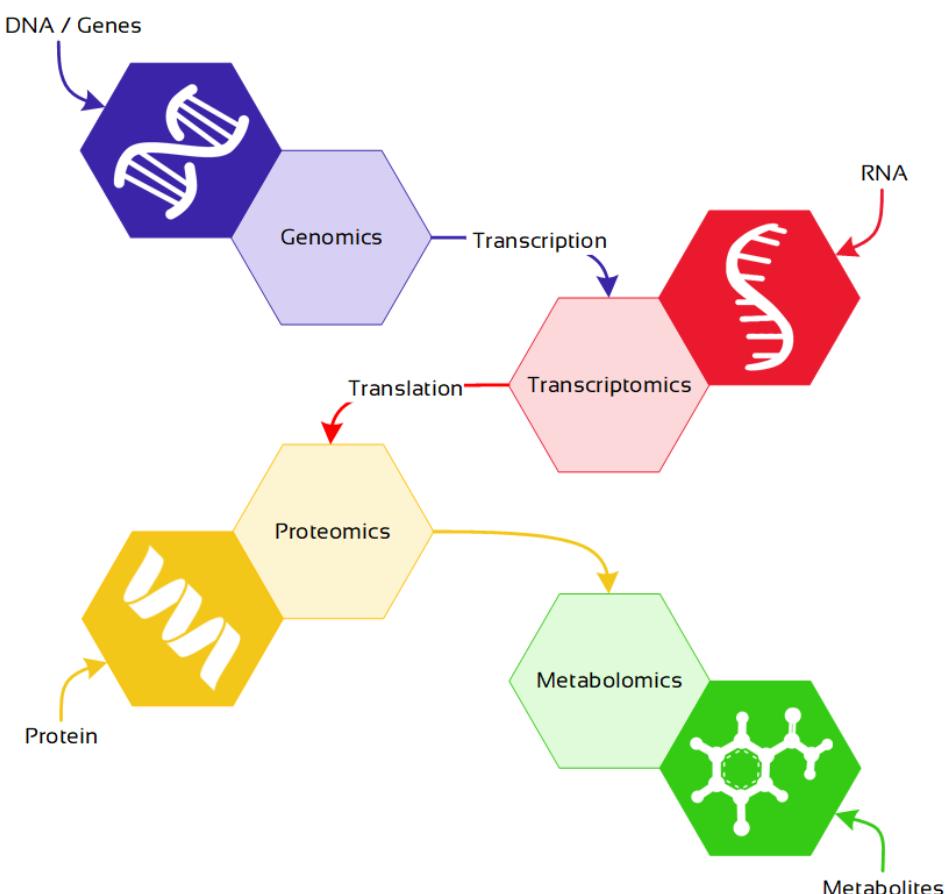


Figure 1: An overview of the four major “omics” fields, from genomics to metabolomics.

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:

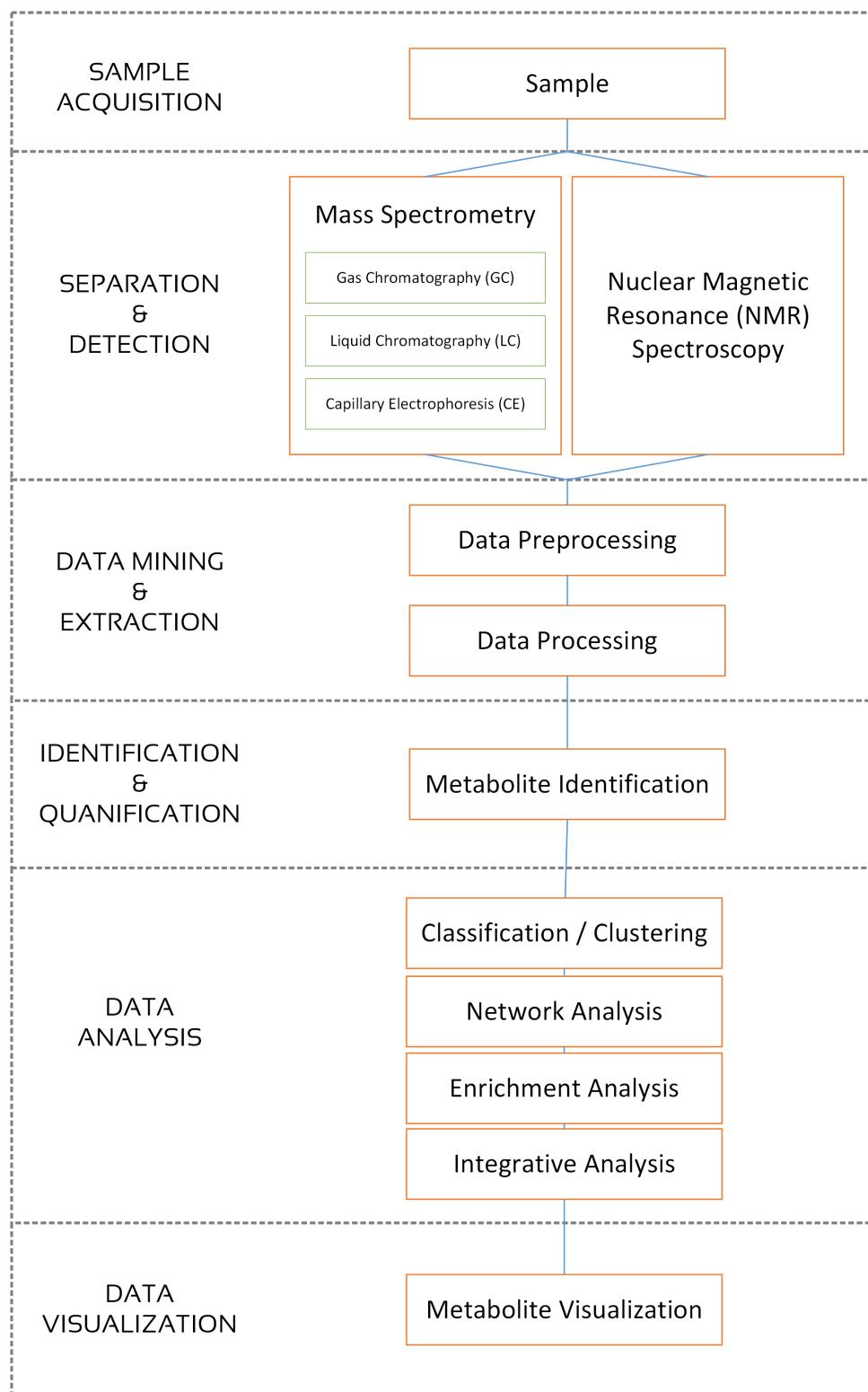


Figure 1: Metabolomics Analytical Workflow

Sample Acquisition

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

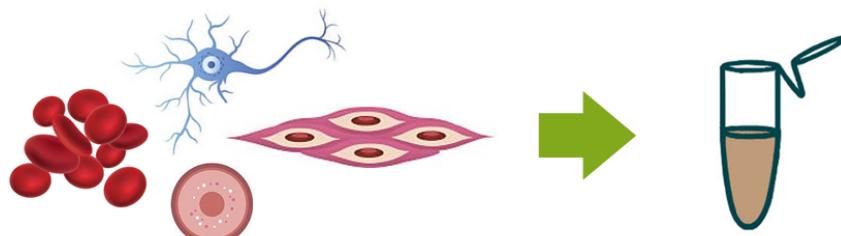


Figure 2: Sample Acquisition from Cells, Tissues and Fluids

Sample Preparation

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

- storage
- extraction
- separation
- detection

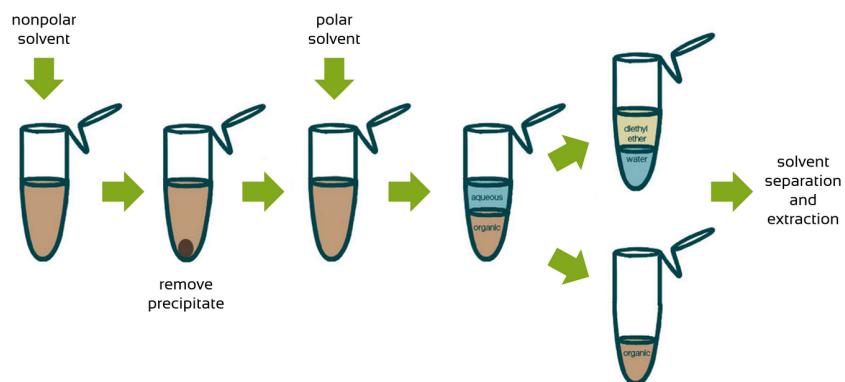


Figure 3: Main Steps involved in sample preparation

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.

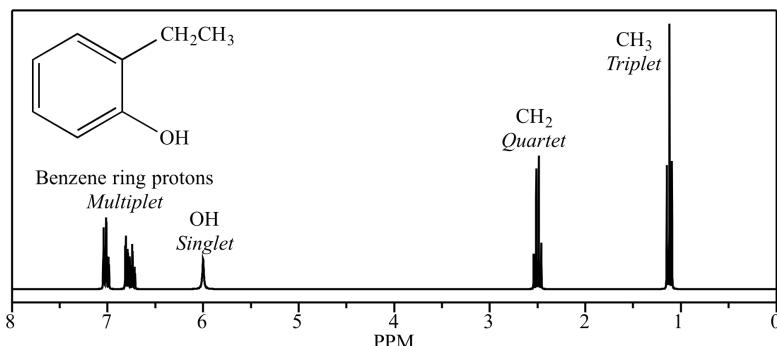


Figure 4: NMR spectrum for 2-ethyl-phenol

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](#).

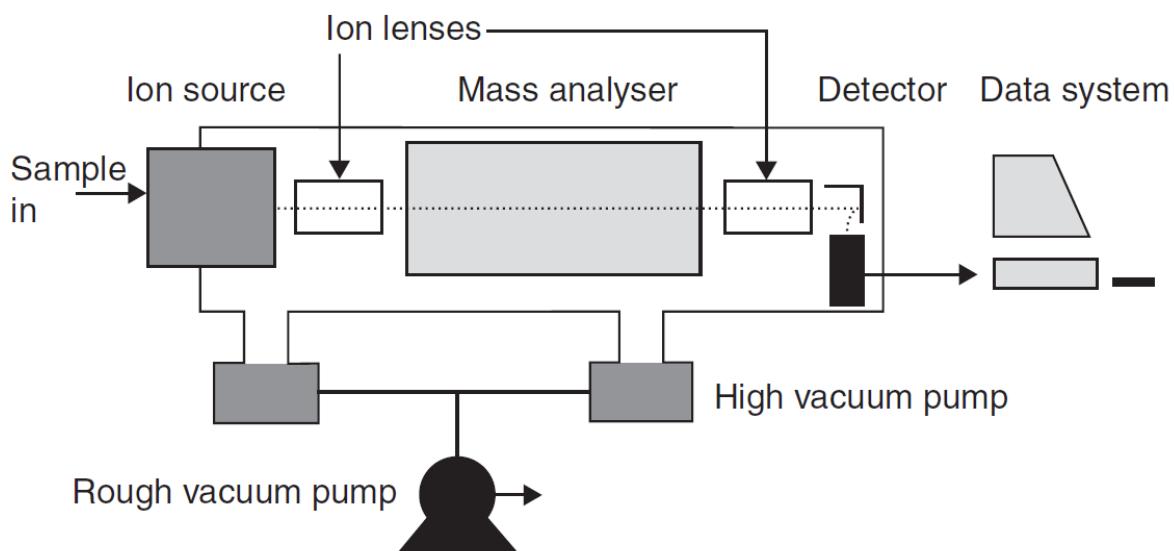


Figure 5: Standard Mass Spectrometer

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](#).

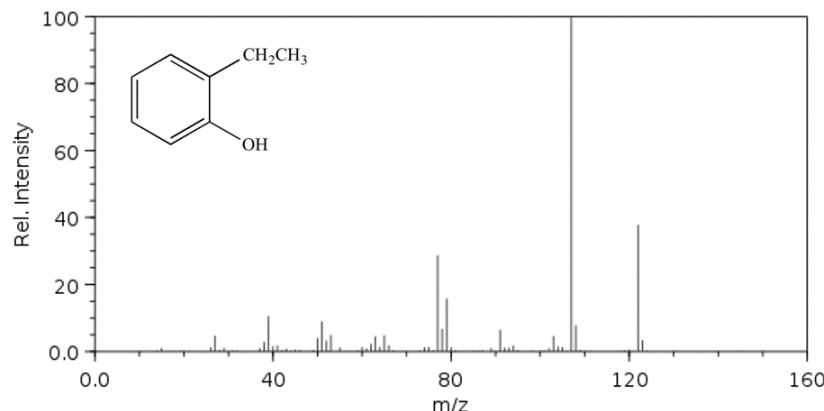


Figure 6: MS spectrum for 2-ethyl-phenol

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 are 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 a 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](#).

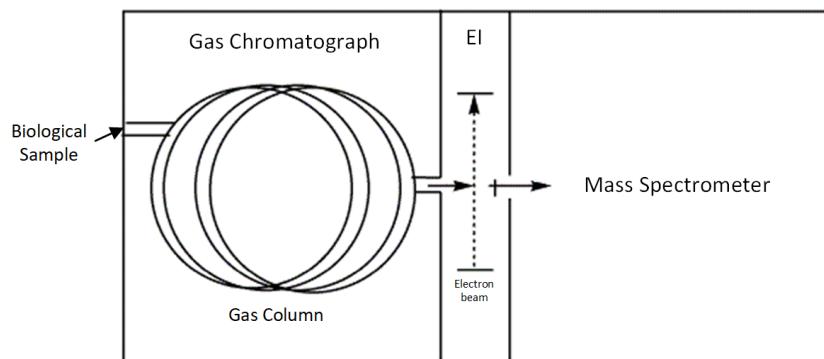


Figure 7: Gas Chromatograph Mass Spectrometer

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.

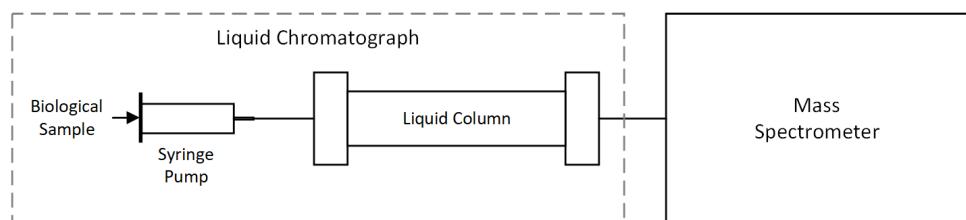


Figure 8: Liquid Chromatograph Mass Spectrometer

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.

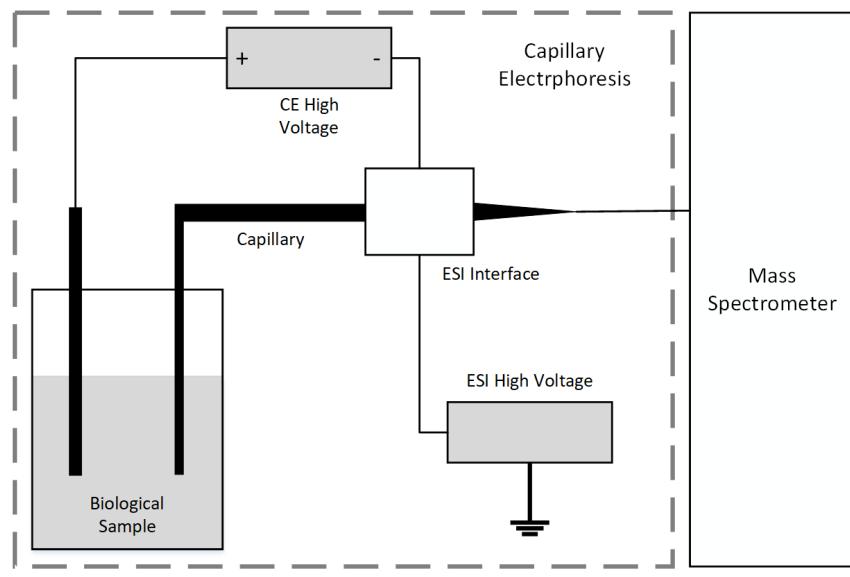


Figure 9: Capillary Electrophoresis Mass Spectrometer

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

Variables (m/z)										
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	0.7485717	7.138574	1.4357149	5.425450082	12.737057	0.95482230	0.4449665	
X20100920_64_NOR_.49	18.477925	0.80084012	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	0.3835749	8.819085130	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	0.3934801	
X20100920_12_RIN_.30	15.079683	0.66735232	0.7217023	7.878826	1.2114839	5.099457599	14.034617	1.18819566	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:

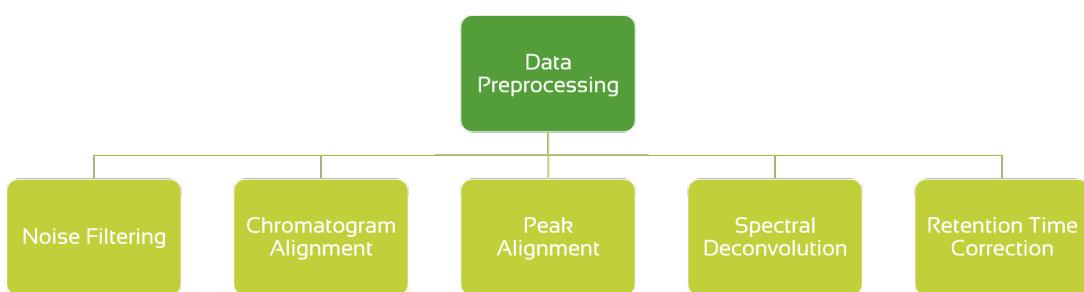


Figure 11: Data Pre-processing

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.

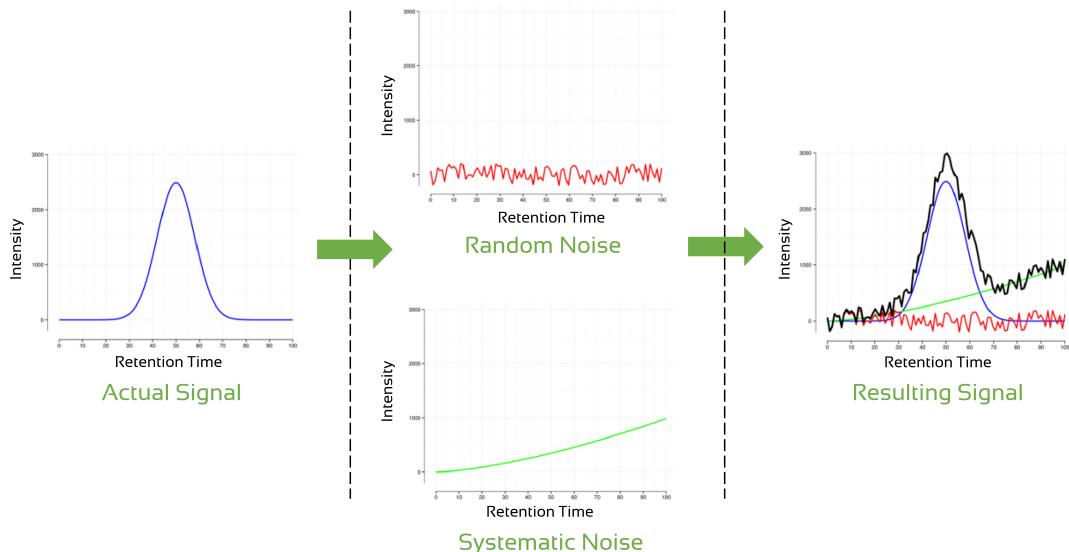


Figure 12: A summary of noise components contributing to signal distortions

- **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 $f(x')$ [270].

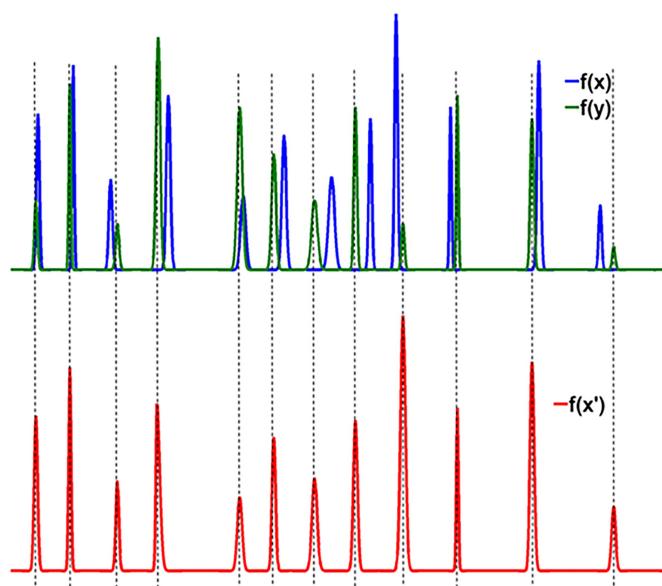


Figure 13: Chromatogram Alignment

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.

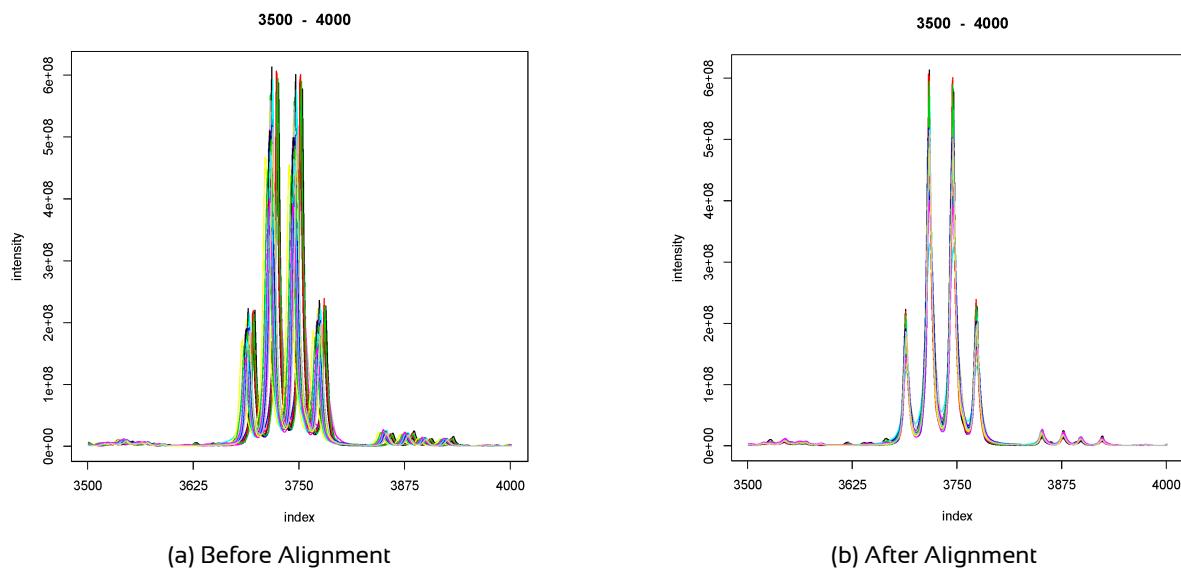


Figure 14: Peak Alignment

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.

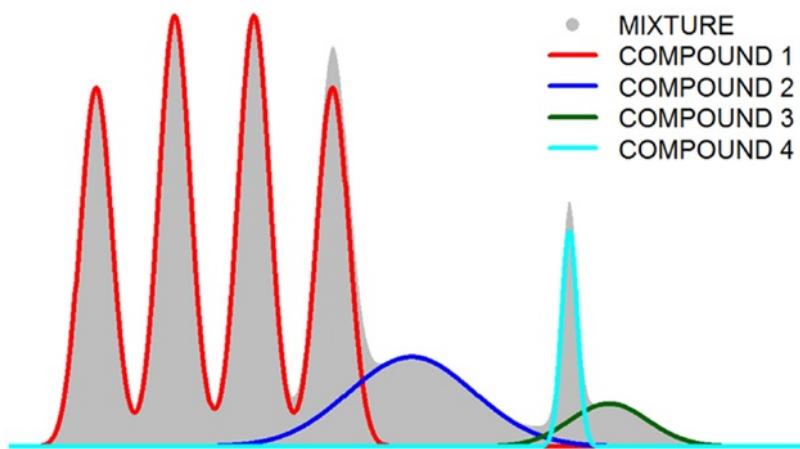


Figure 15: Spectra (gray shaded area) with decomposed (i.e., deconvoluted) multiple components corresponding to different metabolite compounds.

STEP 5 - Retention Time Correction

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

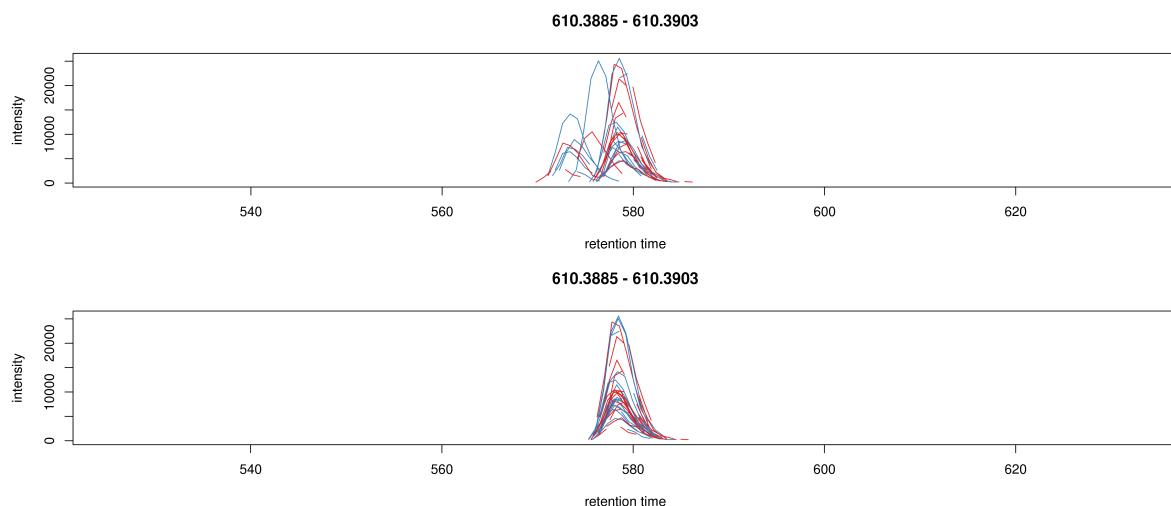


Figure 16: Retention Time Correction

Data Pre-processing

Data Processing is divided into several substeps as follows:

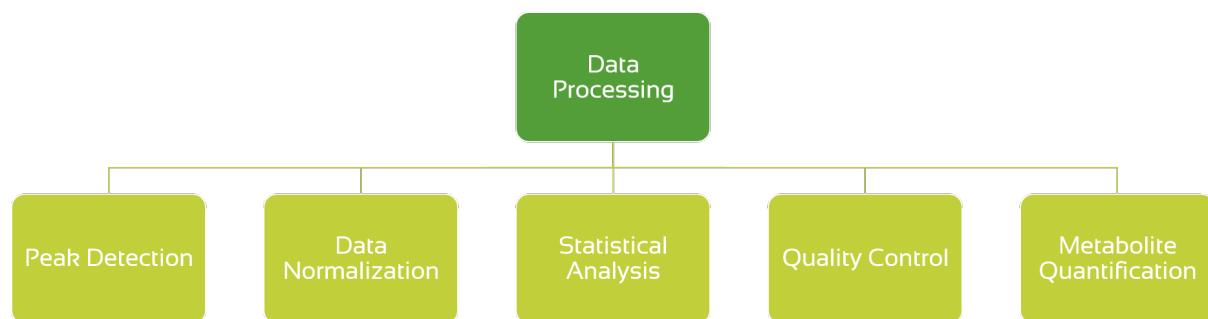


Figure 17: Data Pre-processing

STEP 1 - Peak Detection

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

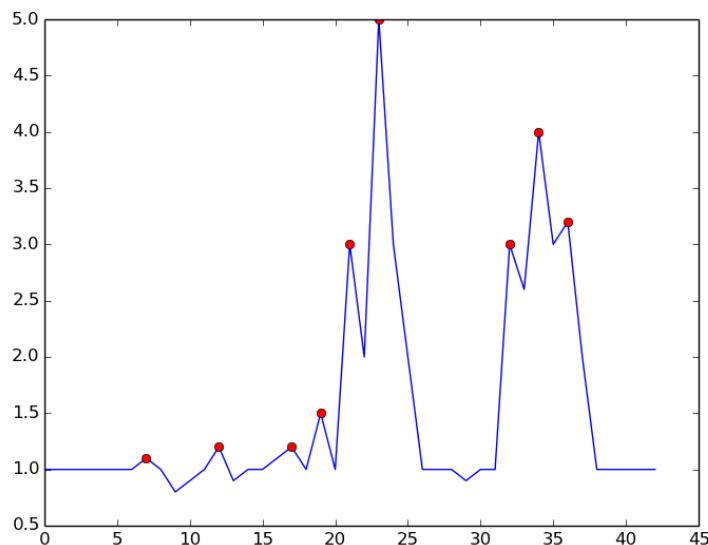


Figure 18: Peak Detection

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.

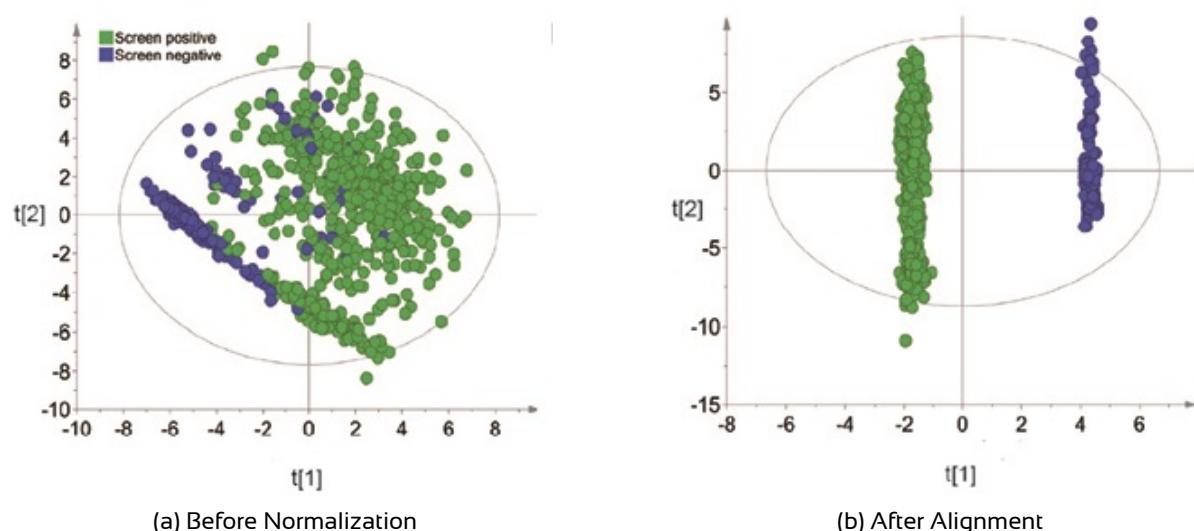


Figure 19: Data Normalization

STEP 3 - 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.

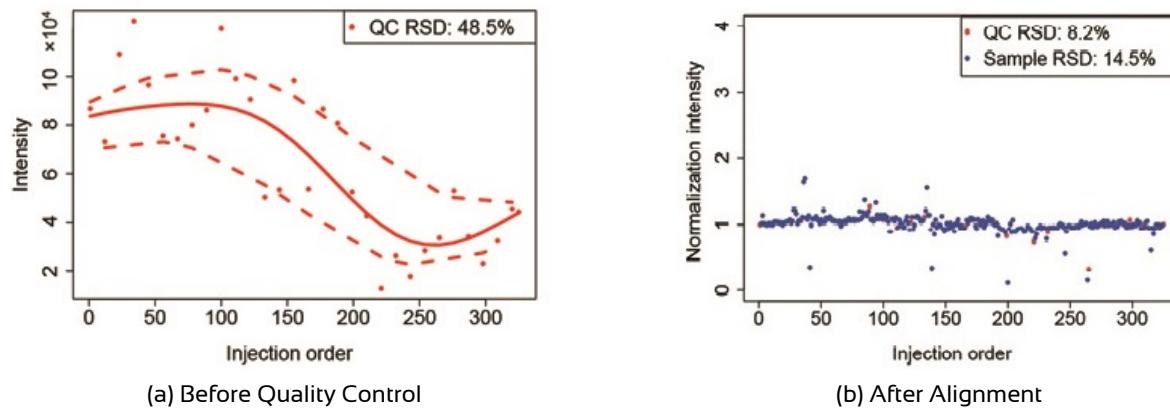


Figure 20: Data Normalization

To establish 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 4 - 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.

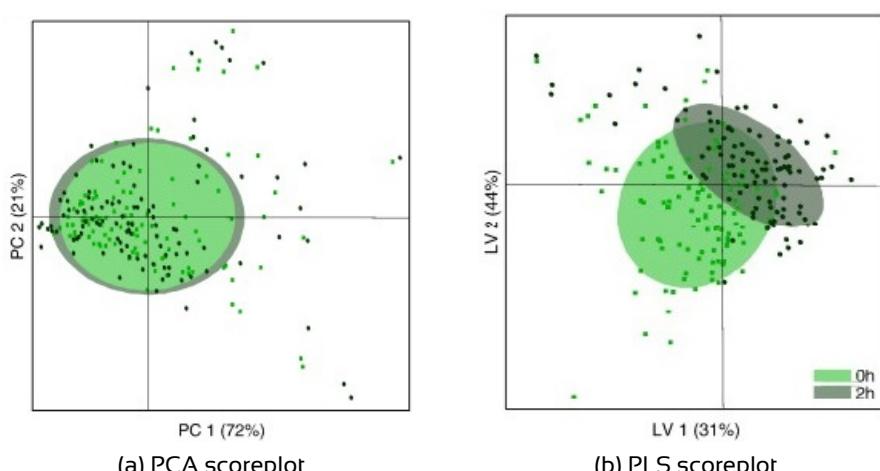


Figure 21: Statistical scoreplots

- **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 5 - Metabolite Quantification

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

STEP 6 - Data Imputation

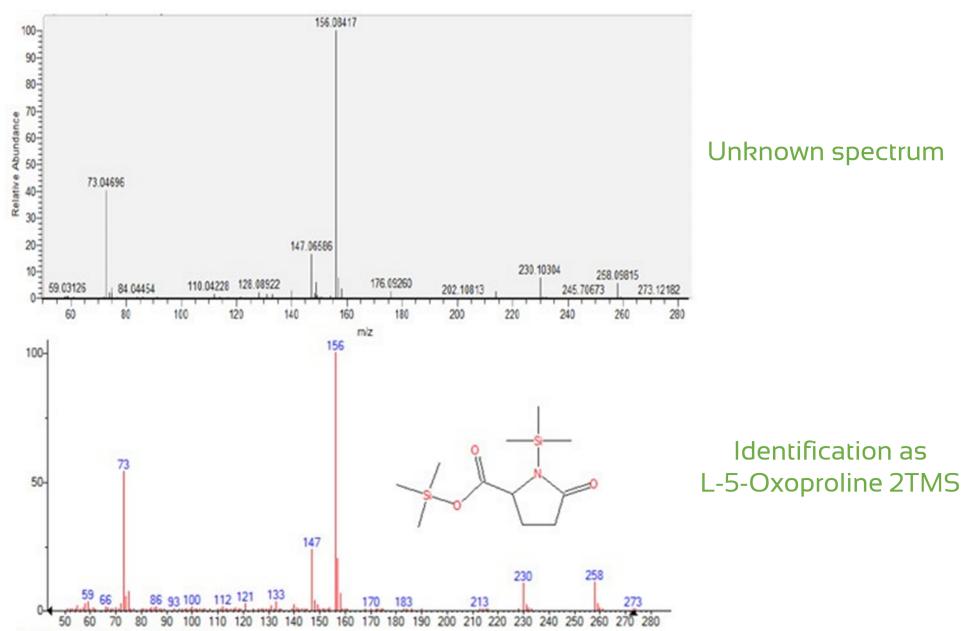
Data Imputation is the process of handling the missing values in mass spectrometry. Typically, there are three types of missing values, missing not at random (MNAR), missing at random (MAR), and missing completely at random (MCAR).

STEP 7 - 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.



Data Analysis

Data Processing and Metabolite Identification can help researchers achieve a systems level understanding of metabolism but these studies produce vast swaths of data which are often only lightly interpreted.

Recently, A large number of computational tools have been developed which enable much deeper analysis of metabolomics data. This analysis in which a finalized dataset is subject to higher level analysis using information obtained from biochemical databases is called **Secondary Analysis** [271].

Several types of Secondary Data Analysis is done as follows:

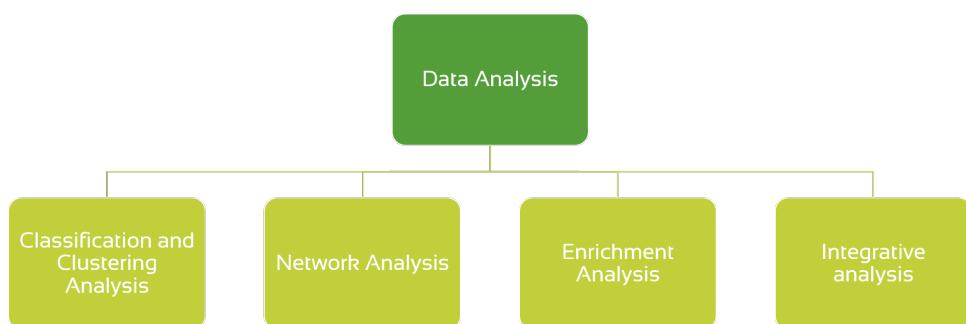


Figure 23: Data Analysis

Classification and Clustering Analysis

Clustering of intensity profiles (multivariate intensity patterns) from mass spectrometry measurements is an unsupervised approach to analyze metabolic data. There are two main approaches to clustering:

- **Metabolite-based Clustering** - all intensities of a metabolite under certain experimental conditions provide an intensity vector representation for multivariate analysis [272].
- **Sample-based Clustering** - only few intensity vectors according to the number of conditions and repetitions are considered for multivariate analysis [272].

Thus, the two clustering approaches correspond to different views on a given matrix of intensity measurements as shown in [Figure 24](#). A row represents a sample for sample-based clustering, while a column corresponds to a (putative) metabolite for metabolite-based clustering. Colors represent different intensity values [272].

	MC 1	MC 2	MC 3	MC 4	MC 5	...
Condition 1	Sample 1					
	Sample 2					
	Sample 3					
Condition 2						
	Sample 1					
	Sample 2					
	Sample 3					
Condition 3						
	Sample 1					
	Sample 2					
	Sample 3					
...

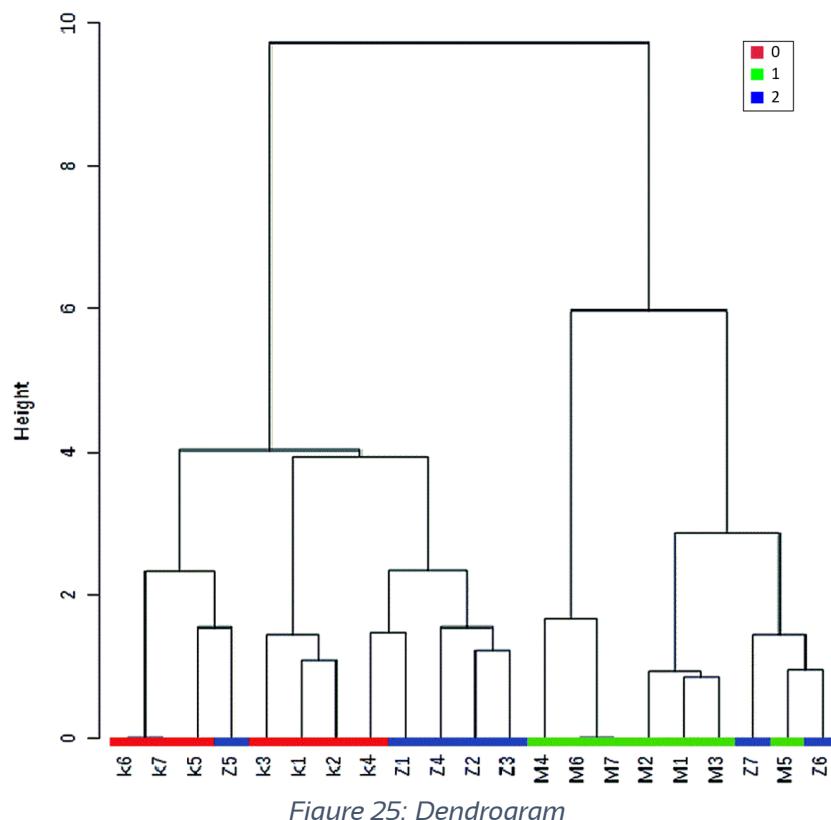
Figure 24: Illustration of differences between sample-based clustering and metabolite-based clustering

There are several clustering algorithms. Important ones are described below:

Hierarchical clustering (HCL)

This method clusters the data forming a tree diagram, or dendrogram as shown in [Figure 25](#), which shows the relationships between samples. It use the following procedures:

1. Calculate the similarity of the two samples using a specific metric, such as Pearson correlation, Euclidean, mutual information and covariance values.
2. Align the most similar samples as neighbors or pair them as a single cluster;
3. Reiterate step 1 and 2 until all samples are aligned.

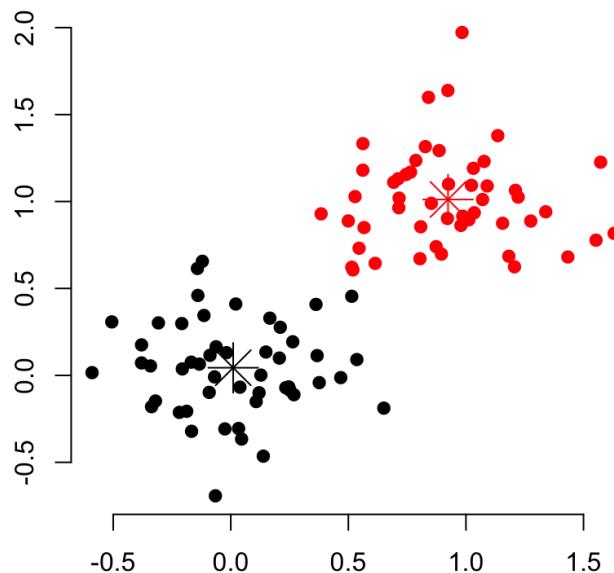


K-means Clustering

The k-means algorithm has become one of the most widely used clustering approaches finding many applications in post-genomics, especially in the analysis of transcriptomic data. It use the following procedures:

1. K initial “means” are randomly generated within the data domain.
2. K clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.
3. The centroid of each of the k clusters becomes the new mean.
4. Repeat steps 2 and 3 are until convergence has been reached.

Following plot shows k-means clustering with k equals to 2.

Figure 26: K-means Clustering with $k=2$

Self-organising maps (SOMs)

The SOM maps a high n-dimensional distribution of data to a low dimensional (usually two-dimensional) array of nodes which enables not only visualisation but also unsupervised classification.

The weight matrix is determined according to the clustering approach (Metabolite-based or Sample-based) as shown in Figure 24. The SOM concept is described in Figure 27.

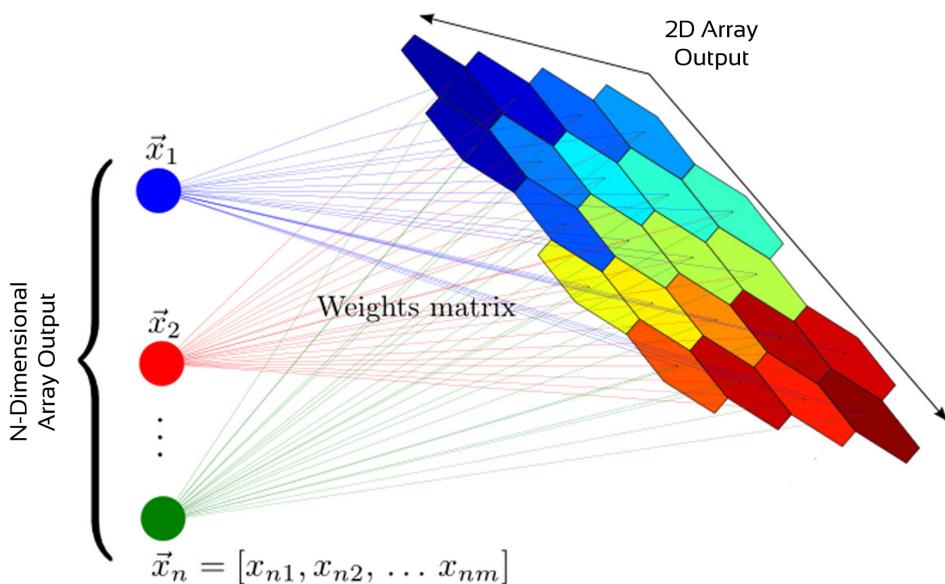


Figure 27: Self-organizing maps concept

1D-SOM matrix after metabolite-based clustering with 33 prototypes is shown in Figure 28. The horizontal and vertical dimensions correspond to prototypes and experimental conditions, respectively. The color of matrix elements represent (average) intensity values according to the color map on the right hand side [272].

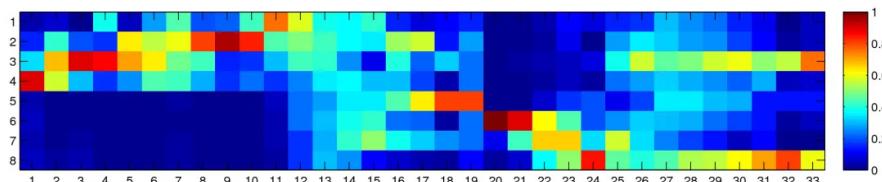


Figure 28: 1D-SOM matrix

Network Modeling and Analysis

A **metabolic pathway** is a linked series of chemical reactions occurring within a cell. The metabolites (reactants, products, and intermediates of an enzymatic reaction) are modified by a sequence of chemical reactions catalyzed by enzymes.

A **metabolic network** is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. As such, these networks comprise the chemical reactions of metabolism, the metabolic pathways, as well as the regulatory interactions that guide these reactions.

Metabolic network modeling, reconstruction, simulation and analysis gives in-depth insight into the molecular mechanisms of a particular organism.

Metabolite Databases

The increase in the number and importance of metabolic networks has come with the need for carefully designed databases to store/organize metabolic networks. Some important databases are given in the following table with their respective links:

Sr. No.	Database Name	Link	Scope				
			Enzymes	Genes	Reactions	Pathways	Metabolites
1	KEGG	http://kegg.jp/	✓	✓	✓	✓	✓
2	BioCyc	http://biocyc.org/	✓	✓	✓	✓	✓
3	MetaCyc	http://metacyc.org/	✓		✓	✓	✓
4	ENZYME	http://enzyme.expasy.org/	✓		✓		✓
5	BRENDA	http://brenda-enzymes.org/	✓		✓		✓
6	BiGG	http://BiGG.ucsd.edu/		✓		✓	✓
7	Reactome	http://reactome.org/	✓	✓		✓	✓
8	KaPPA-View4	http://kpv.kazusa.or.jp/		✓		✓	✓

Figure 29 shows Glycolytic Pathway of Homo Sapiens (green) with Drug/Disease (pink/blue).

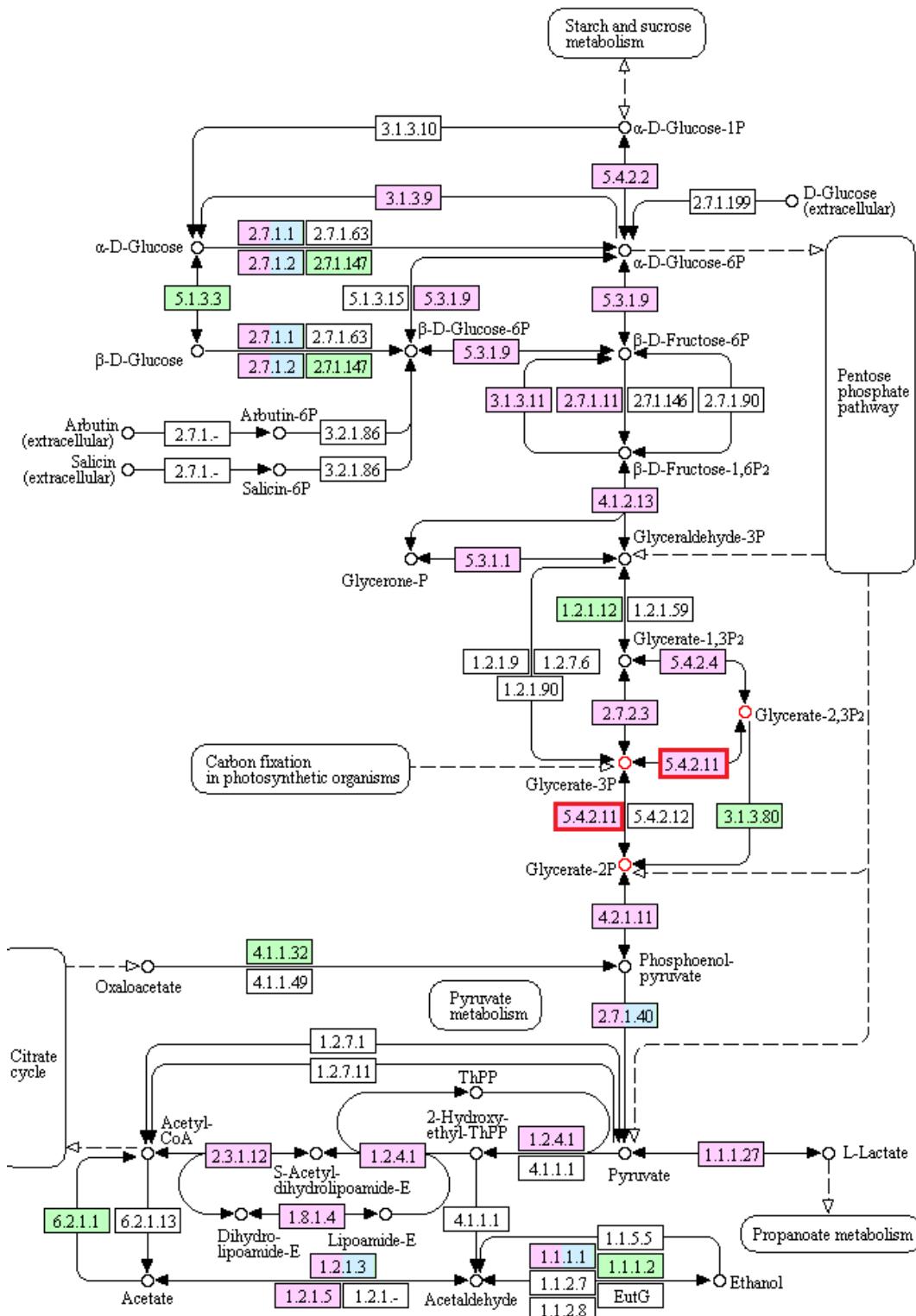


Figure 29: MAPOO10 KEGG Database - Glycolysis of Homo Sapiens

Network Analysis can be further divided into three steps:

STEP 1 - Metabolite Pathway Mapping

Metabolite mapping is a process of Integration of biochemical pathway and chemical relationships to map all detected metabolites in network graphs. The identified metabolites are compared with the existing database and mapped onto the existing network maps. Figure 30 shows mapping of C00025, C00043, C00064, C00073, C00082, C00487, C00588 and C00670 compounds.

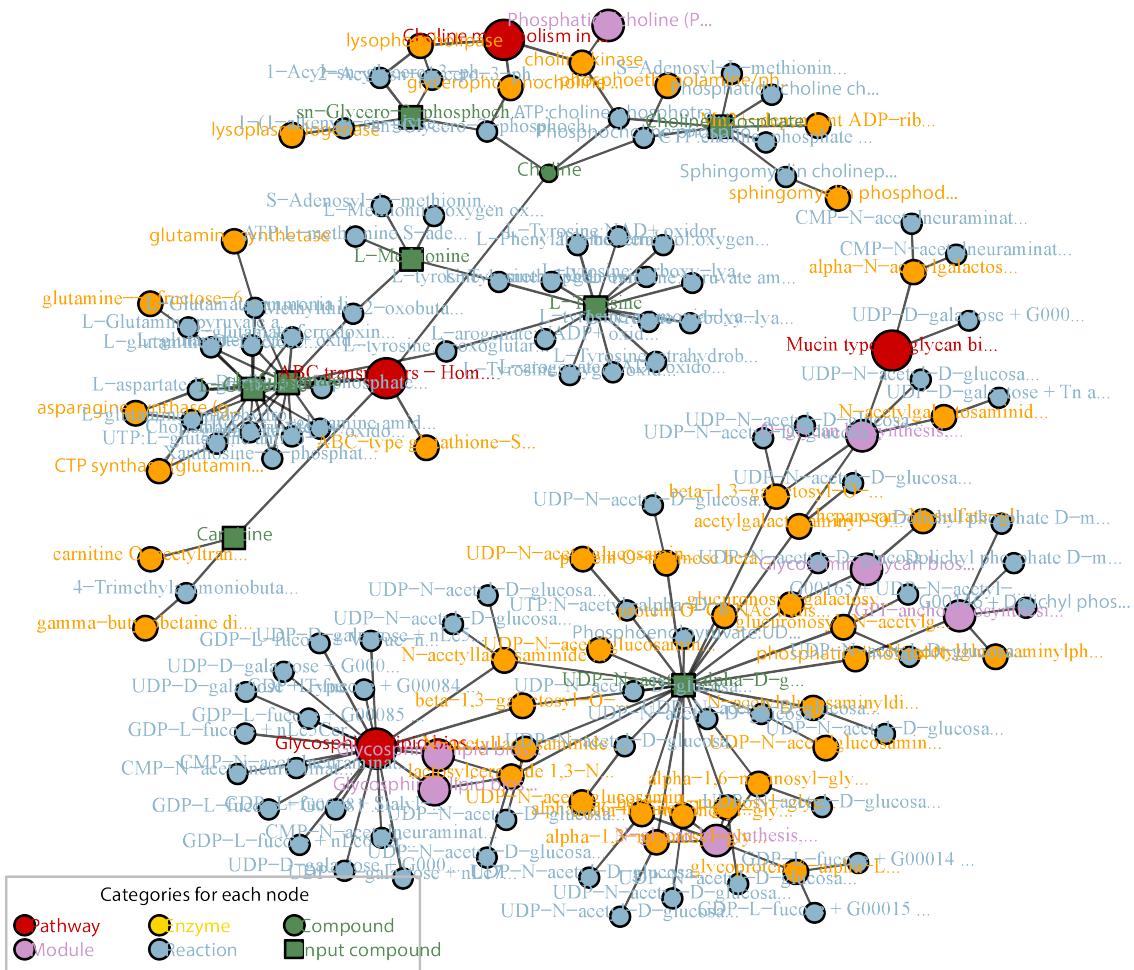


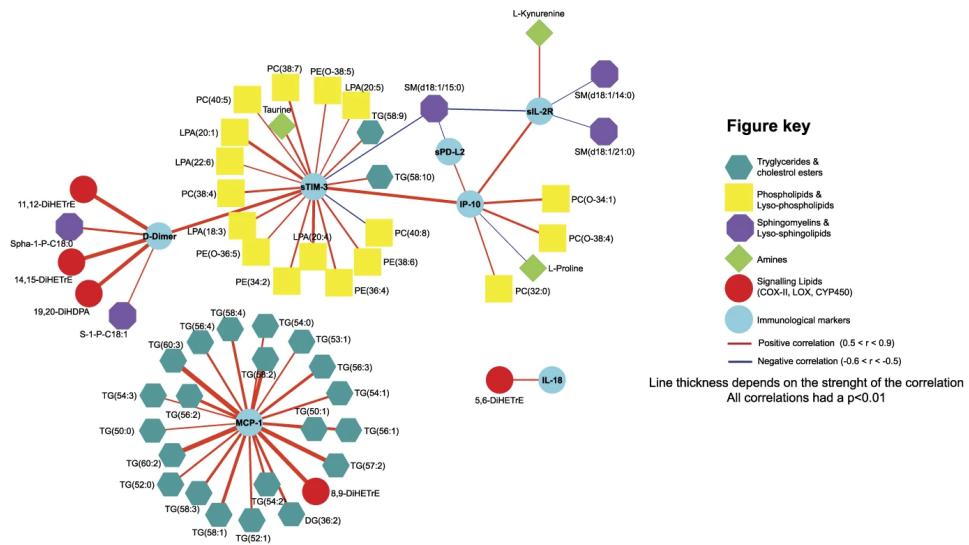
Figure 30: Metabolite Pathway Mapping

STEP 2 - Metabolic Network Analysis

Metabolic networks analysis can be used to detect correlation and comorbidity patterns in diseased patients [273, 274].

If the same metabolite is shared between two metabolic reactions, the scarcity or abundance of that metabolite may affect the fluxes of both reactions, potentially coupling their activity. We consider two metabolic reactions linked if they process a common metabolite (correlation), i.e. if they are adjacent to each other in a metabolic reaction map.

For example, in [Figure 29](#), if the phosphoglycerate mutase (5.4.2.11, Analyzed in pink colour) is not active, the production (or consumption) of glycerate-2P, and in turn of phosphoenolpyruvate, is expected to also be altered. For the HIV Infection, correlation of metabolites is shown in [Figure 31](#).

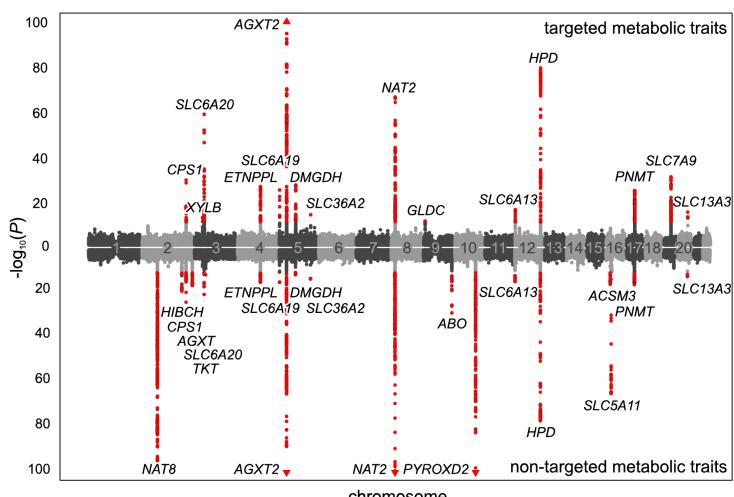


[Figure 31: Correlation of HIV Infection](#)

STEP 3 - metabolite Genome-wide Association Study (mGWAS)

Genome wide association studies (GWAS) are hypothesis-free methods for identifying associations between genetic regions (loci) and traits (including diseases), whereas **metabolomics GWAS (mGWAS)** investigate how genetic variation effect metabolite levels (metabolism and complex diseases). It maps the loci, position of a genes on a chromosome, responsible for natural variations in a target phenotype.

A **Manhattan plot**, as shown in [Figure 32](#), is a type of scatter plot, usually used to display data with a large number of data-points, many of non-zero amplitude, and with a distribution of higher-magnitude values. The plot is commonly used in GWAS studies.



[Figure 32: Manhattan plot of genetic associations to targeted and non-targeted traits](#)

Enrichment analysis

Metabolite Set Enrichment Analysis (MSEA) use a collection of predefined metabolite pathways, network information and disease states obtained from the databases which help metabolomics researchers identify and interpret patterns of metabolite concentration changes in a biologically meaningful way [275].

According to the user data input there are three main types of enrichment analysis:

- **Over Representation Analysis (ORA)** - Accepts list of metabolite names only.
- **Single Sample Profiling (SSP)** - Accepts a two column list of metabolite names and concentration data from a single sample
- **Quantitative Enrichment Analysis (QEA)** - Accepts a table with concentration data from multiple samples.

Data generated is the tabular form and can be visualized as shown in Figure 33

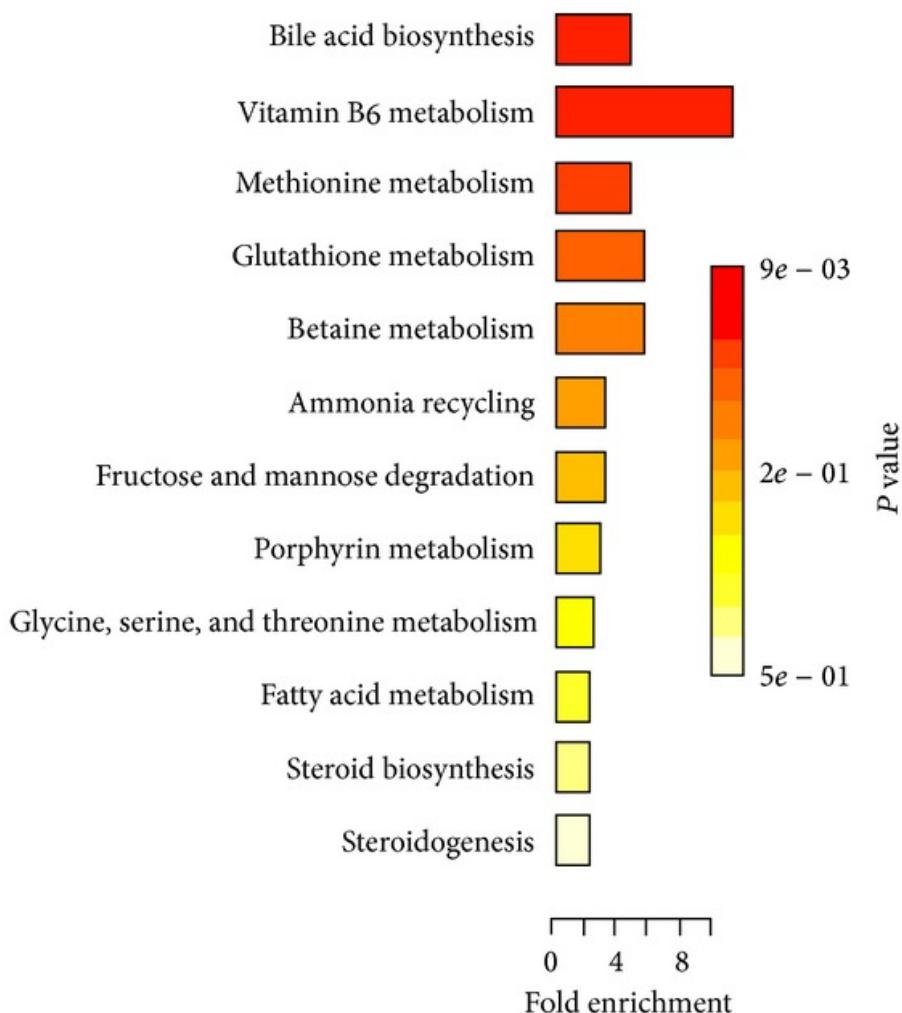


Figure 33: MSEA of Colorectal carcinoma (CRC)

Integrative analysis

The use of a **multi-omics**, systems biology approach to study cellular metabolism is key to understanding disease etiology and outcome. Integration of omics data types can help identify biological pathways affected and provides a more complete picture disease biology and prognosis, and can even provide new drug targets.

Integrative Analysis enables the integration of transcriptomic and proteomic data with genome-scale metabolic network models to predict enzymes' metabolic flux. The prediction of metabolic fluxes based on high-throughput molecular data sources could help to advance our understanding of cellular metabolism.

Integration of metabolomics and transcriptomics revealed altered lipid metabolism pathways in PDAC. A weighted network analysis showed that the turquoise-colored module was highly conserved in both test and validation cohort as shown in [Figure 34](#). 8 fatty acids (indicated by stars) with high connectivity in both cohorts represent the main hubs in the metabolite network that were chosen for integration analysis.

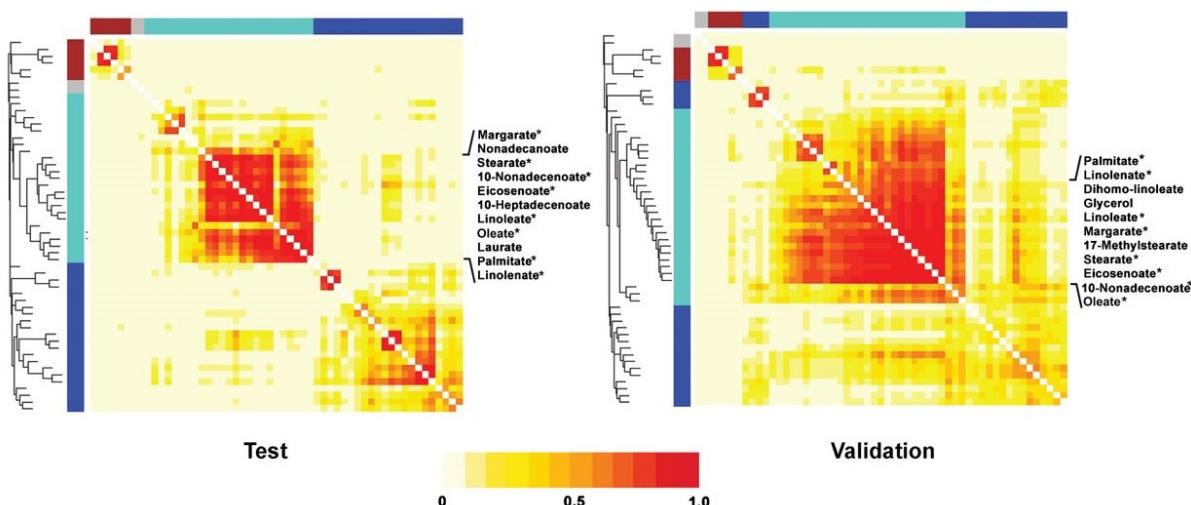


Figure 34: Integration Analysis of lipid metabolism

Data Visualization

Visualization of the metabolic spectrum, chromatograms, metabolic pathways and networks, dendrograms, heatmaps, SOMs and clusters. All the visualization of these things is already shown in each explanation.

MetumpX Software Pipeline

As discussed, Untargeted MS-based Metabolomics software pipeline is divided into several steps. These steps are visualized using a pipeline as shown in Figure 35 so that the tools can be accessed according to user's needs. MetumpX contain sufficient tools for each category. Download size of each software is mentioned for user's ease.

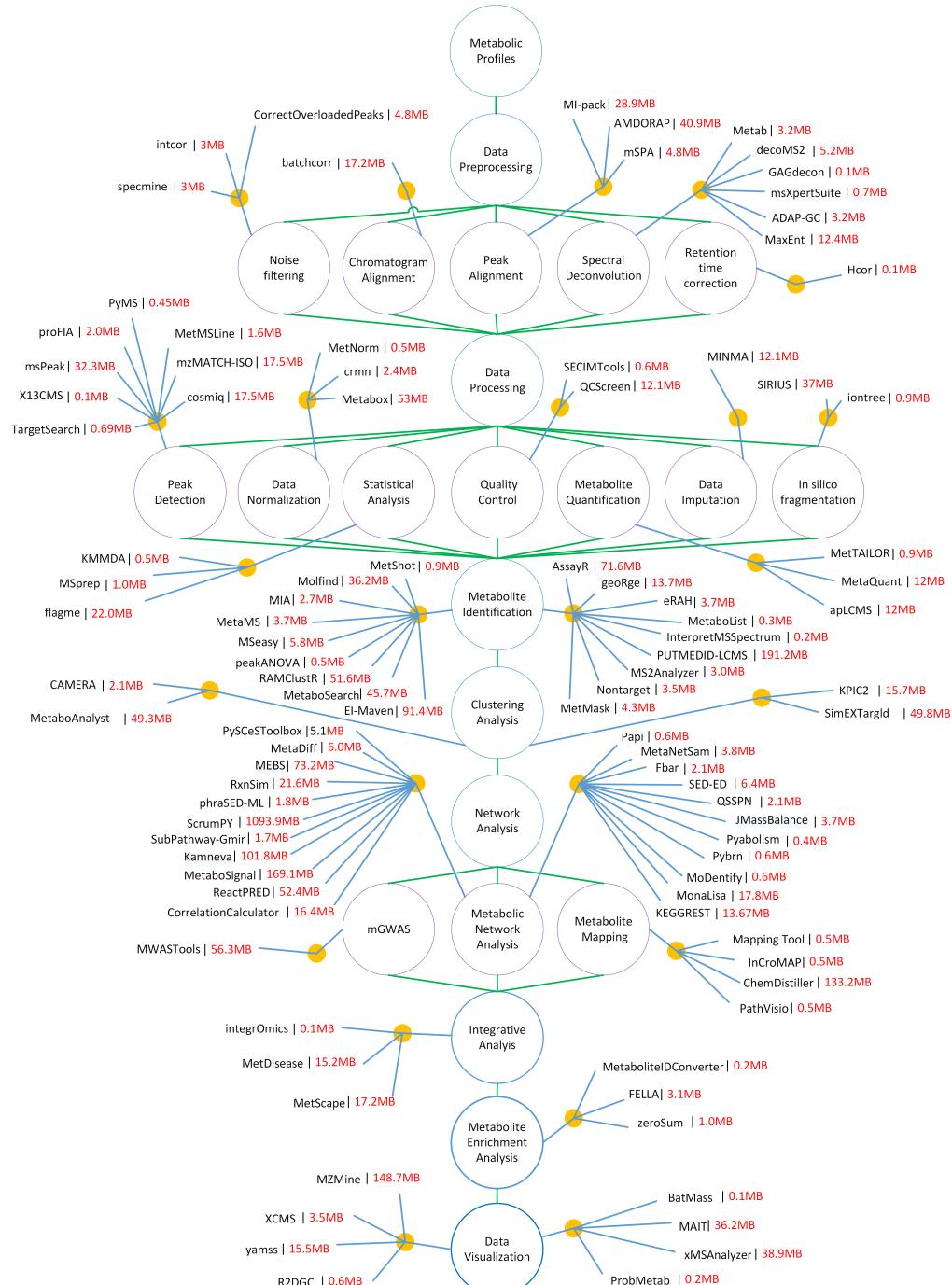


Figure 35: Software Pipeline representing individual steps along with software tools packaged within MetumpX. for details on installation.

Bibliography

- [1] 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.
- [2] D. Ghosh and H. Choi, "metaarray package for meta-analysis of microarray data," *R package*. <http://www.r-project.org>, 2006.
- [3] H. Poppe, "Overloading and interaction phenomena in electrophoretic separations," *Analytical Chemistry*, vol. 64, no. 17, pp. 1908–1919, 1992.
- [4] H.-Y. Chang, C.-T. Chen, T. M. Lih, K.-S. Lynn, C.-G. Juo, W.-L. Hsu, and T.-Y. Sung, "imet-q: a user-friendly tool for label-free metabolomics quantitation using dynamic peak-width determination," *PloS one*, vol. 11, no. 1, p. e0146112, 2016.
- [5] C. Baldari, V. Bonavolontà, G. P. Emerenziani, M. C. Gallotta, A. J. Silva, and L. Guidetti, "Accuracy, reliability, linearity of accutrend and lactate pro versus ebio plus analyzer," *European journal of applied physiology*, vol. 107, no. 1, pp. 105–111, 2009.
- [6] 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.
- [7] E. Tyteca, M. Talebi, R. Amos, S. H. Park, M. Taraji, Y. Wen, R. Szucs, C. A. Pohl, J. W. Dolan, and P. R. Haddad, "Towards a chromatographic similarity index to establish localized quantitative structure-retention models for retention prediction: use of retention factor ratio," *Journal of Chromatography A*, vol. 1486, pp. 50–58, 2017.
- [8] R. G. Edwards and B. Joo, "The chroma software system for lattice qcd," *arXiv preprint hep-lat/0409003*, 2004.
- [9] 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.
- [10] W. Zhang, J. Chang, Z. Lei, D. Huhman, L. W. Sumner, and P. X. Zhao, "Met-cofea: a liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation," *Analytical chemistry*, vol. 86, no. 13, pp. 6245–6253, 2014.

- [11] W. Zhang, Z. Lei, D. Huhman, L. W. Sumner, and P. X. Zhao, "Met-xalign: a metabolite cross-alignment tool for lc/ms-based comparative metabolomics," *Analytical chemistry*, vol. 87, no. 18, pp. 9114–9119, 2015.
 - [12] 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.
 - [13] 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.
 - [14] C. D. Broeckling, I. R. Reddy, A. L. Duran, X. Zhao, and L. W. Sumner, "Met-idea: data extraction tool for mass spectrometry-based metabolomics," *Analytical chemistry*, vol. 78, no. 13, pp. 4334–4341, 2006.
 - [15] 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.
 - [16] C.-A. Darveau, R. K. Suarez, R. D. Andrews, and P. W. Hochachka, "Allometric cascade as a unifying principle of body mass effects on metabolism," *Nature*, vol. 417, no. 6885, p. 166, 2002.
 - [17] D. Reddy, C. Sekar, C. Gaur, D. B. J. Causon, and M. Pillai, "Identification & confirmation of structurally related degradation products of simvastatin."
 - [18] 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.
 - [19] 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 oncprotein," *Molecular and cellular biology*, vol. 24, no. 23, pp. 10 470–10 478, 2004.
 - [20] K. Jobu, C. Sun, S. Yoshioka, J. Yokota, M. Onogawa, C. Kawada, K. Inoue, T. Shuin, T. Sendo, and M. Miyamura, "Metabolomics study on the biochemical profiles of odor elements in urine of human with bladder cancer," *Biological and Pharmaceutical Bulletin*, vol. 35, no. 4, pp. 639–642, 2012.
 - [21] 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.
 - [22] D. N. Cole, "Controlling the spread of campsites at popular wilderness destinations," *Journal of Soil and Water Conservation*, vol. 37, no. 5, pp. 291–295, 1982.
 - [23] S. Beisken, P. Conesa, K. Haug, R. M. Salek, and C. Steinbeck, "Specktackle: Javascript charts for spectroscopy," *Journal of cheminformatics*, vol. 7, no. 1, p. 17, 2015.
 - [24] E. Want and P. Masson, "Processing and analysis of gc/lc-ms-based metabolomics data," in *Metabolic profiling*. Springer, 2011, pp. 277–298.
-

-
- [25] D. A. Shirley, "High-resolution x-ray photoemission spectrum of the valence bands of gold," *Physical Review B*, vol. 5, no. 12, p. 4709, 1972.
 - [26] 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.
 - [27] D. M. Small, "The effects of glyceride structure on absorption and metabolism," *Annual review of nutrition*, vol. 11, no. 1, pp. 413–434, 1991.
 - [28] 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.
 - [29] 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.
 - [30] C. Szabo, C. Ransy, K. Módis, M. Andriamihaja, B. Murghes, C. Coletta, G. Olah, K. Yanagi, and F. Bouillaud, "Regulation of mitochondrial bioenergetic function by hydrogen sulfide. part i. biochemical and physiological mechanisms," *British journal of pharmacology*, vol. 171, no. 8, pp. 2099–2122, 2014.
 - [31] Y. Inoue, T. Kubokawa, and K. Sato, "Photocatalytic activity of alkali-metal titanates combined with ruthenium in the decomposition of water," *The Journal of Physical Chemistry*, vol. 95, no. 10, pp. 4059–4063, 1991.
 - [32] 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.
 - [33] I. Nikolskiy, N. G. Mahieu, Y.-J. Chen, R. Tautenhahn, and G. J. Patti, "An untargeted metabolomic workflow to improve structural characterization of metabolites," *Anal. Chem*, vol. 85, no. 16, pp. 7713–7719, 2013.
 - [34] 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.
 - [35] M. Sonnett, E. Yeung, and M. Wuhr, "Accurate, sensitive, and precise multiplexed proteomics using the complement reporter ion cluster," *Analytical chemistry*, vol. 90, no. 8, pp. 5032–5039, 2018.
 - [36] M.-F. Bernet, D. Brassart, J.-R. Neeser, and A. Servin, "Lactobacillus acidophilus la 1 binds to cultured human intestinal cell lines and inhibits cell attachment and cell invasion by enterovirulent bacteria." *Gut*, vol. 35, no. 4, pp. 483–489, 1994.
 - [37] W. Cui, H. W. Rohrs, and M. L. Gross, "Top-down mass spectrometry: recent developments, applications and perspectives," *Analyst*, vol. 136, no. 19, pp. 3854–3864, 2011.
-

- [38] V. Picaud, J.-F. Giovannelli, C. Truntzer, J.-P. Charrier, A. Giremus, P. Grangeat, and C. Mercier, "Linear maldi-tof simultaneous spectrum deconvolution and baseline removal," *BMC bioinformatics*, vol. 19, no. 1, p. 123, 2018.
- [39] 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.
- [40] 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.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] S. Vey and A. Voigt, "Amdis: adaptive multidimensional simulations," *Computing and Visualization in Science*, vol. 10, no. 1, pp. 57–67, 2007.
- [45] 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.
- [46] P. C. das Neves, P. H. N. Rangel, and V. d. A. Cutrim, "Selección recurrente para la producción de arroz híbrido." Centro Internacional de Agricultura Tropical (CIAT), 1997.
- [47] 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.
- [48] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000.
- [49] B. C. DeFelice, S. S. Mehta, S. Samra, T. Cajka, B. Wancewicz, J. F. Fahrmann, and O. Fiehn, "Mass spectral feature list optimizer (ms-flo): A tool to minimize false positive peak reports in untargeted liquid chromatography–mass spectroscopy (lc-ms) data processing," *Analytical chemistry*, vol. 89, no. 6, pp. 3250–3255, 2017.
- [50] T. Jansen and I. Wegener, "On the analysis of evolutionary algorithms—a proof that crossover really can help," in *European Symposium on Algorithms*. Springer, 1999, pp. 184–193.

-
- [51] L. G. Johnsen, T. Skov, U. Houlberg, and R. Bro, "An automated method for baseline correction, peak finding and peak grouping in chromatographic data," *Analyst*, vol. 138, no. 12, pp. 3502–3511, 2013.
 - [52] V. Treviño, I.-L. Yañez-Garza, C. E. Rodriguez-López, R. Urrea-López, M.-L. Garza-Rodriguez, H.-A. Barrera-Saldaña, J. G. Tamez-Peña, R. Winkler, and R.-I. Díaz de-la Garza, "Gridmass: a fast two-dimensional feature detection method for lc/ms," *Journal of Mass Spectrometry*, vol. 50, no. 1, pp. 165–174, 2015.
 - [53] 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. VI-120.
 - [54] L. K. Grover, "Quantum mechanics helps in searching for a needle in a haystack," *Physical review letters*, vol. 79, no. 2, p. 325, 1997.
 - [55] 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.
 - [56] N. H. Beltrán, M. A. Duarte-Mermoud, V. A. S. Vicencio, S. A. Salah, and M. A. Bustos, "Chilean wine classification using volatile organic compounds data obtained with a fast gc analyzer," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 11, pp. 2421–2436, 2008.
 - [57] M. G. Kok, M. M. Ruijken, J. R. Swann, I. D. Wilson, G. W. Somsen, and G. J. de Jong, "Anionic metabolic profiling of urine from antibiotic-treated rats by capillary electrophoresis–mass spectrometry," *Analytical and bioanalytical chemistry*, vol. 405, no. 8, pp. 2585–2594, 2013.
 - [58] W. Zhou, S.-L. Su, J.-A. Duan, J.-M. Guo, D.-W. Qian, E.-X. Shang, and J. Zhang, "Characterization of the active constituents in shixiao san using bioactivity evaluation followed by uplc-qtof and markerlynx analysis," *Molecules*, vol. 15, no. 9, pp. 6217–6230, 2010.
 - [59] D. Mao, R. Lookman, H. Van De Weghe, R. Weltens, G. Vanermen, N. De Brucker, and L. Diels, "Estimation of ecotoxicity of petroleum hydrocarbon mixtures in soil based on hplc–gcxgc analysis," *Chemosphere*, vol. 77, no. 11, pp. 1508–1513, 2009.
 - [60] 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.
 - [61] 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.
 - [62] N. Zhong, J.-Z. Dong, Y. Yao, and S. Ohsuga, "Gastric cancer data mining with ordered information," in *International Conference on Rough Sets and Current Trends in Computing*. Springer, 2002, pp. 467–478.
-

- [63] B. Li, J. Tang, Q. Yang, S. Li, X. Cui, Y. Li, Y. Chen, W. Xue, X. Li, and F. Zhu, "Noreva: normalization and evaluation of ms-based metabolomics data," *Nucleic acids research*, vol. 45, no. W1, pp. W162–W170, 2017.
- [64] 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.
- [65] B. M. Sullivan and S.-P. Hsu, "Biometric score normalizer," Jul. 8 2003, uS Patent 6,591,224.
- [66] B. Wen, Z. Mei, C. Zeng, and S. Liu, "metax: a flexible and comprehensive software for processing metabolomics data," *BMC bioinformatics*, vol. 18, no. 1, p. 183, 2017.
- [67] B. Ernest, J. R. Gooding, S. R. Campagna, A. M. Saxton, and B. H. Voy, "Metabr: an r script for linear model analysis of quantitative metabolomic data," *BMC research notes*, vol. 5, no. 1, p. 596, 2012.
- [68] J. A. Berger, S. Hautaniemi, A.-K. Järvinen, H. Edgren, S. K. Mitra, and J. Astola, "Optimized lowess normalization parameter selection for dna microarray data," *BMC bioinformatics*, vol. 5, no. 1, p. 194, 2004.
- [69] G. Hughes, C. Cruickshank-Quinn, R. Reisdorph, S. Lutz, I. Petracche, 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.
- [70] H. Tsugawa, T. Bamba, M. Shinohara, S. Nishiumi, M. Yoshida, and E. Fukusaki, "Practical non-targeted gas chromatography/mass spectrometry-based metabolomics platform for metabolic phenotype analysis," *Journal of bioscience and bioengineering*, vol. 112, no. 3, pp. 292–298, 2011.
- [71] T. Kokubun and L. D'Costa, "Direct and unbiased information recovery from liquid chromatography-mass spectrometry raw data for phenotype-differentiating metabolites based on screening window coefficient of ion currents," *Analytical chemistry*, vol. 85, no. 18, pp. 8684–8691, 2013.
- [72] A. Geremia, C. V. Arancibia-Cárcamo, M. P. Fleming, N. Rust, B. Singh, N. J. Mortensen, S. P. Travis, and F. Powrie, "Il-23-responsive innate lymphoid cells are increased in inflammatory bowel disease," *Journal of Experimental Medicine*, vol. 208, no. 6, pp. 1127–1133, 2011.
- [73] 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.
- [74] 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.
- [75] P. Kuich, N. Hoffmann, and S. Kempa, "Maui-via: a user-friendly software for visual identification, alignment, correction, and quantification of gas chromatography-mass spectrometry data," *Frontiers in bioengineering and biotechnology*, vol. 2, p. 84, 2015.

- [76] 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.
- [77] 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.
- [78] M. Calderón-Santiago, M. A. López-Bascón, Á. Peralbo-Molina, and F. Priego-Capote, "Metaboqc: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls," *Talanta*, vol. 174, pp. 29–37, 2017.
- [79] A. M. Simader, B. Kluger, N. K. N. Neumann, C. Bueschl, M. Lemmens, G. Lirk, R. Krška, 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.
- [80] H. Luan, F. Ji, Y. Chen, and Z. Cai, "stattarget: A streamlined tool for signal drift correction and interpretations of quantitative mass spectrometry-based omics data," *Analytica chimica acta*, vol. 1036, pp. 66–72, 2018.
- [81] B. Hoekman, R. Breitling, F. Suits, R. Bischoff, and P. Horvatovich, "mscompare: a framework for quantitative analysis of label-free lc-ms data for comparative candidate biomarker studies," *Molecular & Cellular Proteomics*, vol. 11, no. 6, pp. M111–O15 974, 2012.
- [82] M. M. Smits, R. Carleer, and J. V. Colpaert, "Pyquan: A rapid workflow around the amdis deconvolution software for high throughput analysis of pyrolysis gc/ms data," *Journal of analytical and applied pyrolysis*, vol. 118, pp. 335–342, 2016.
- [83] 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.
- [84] T. Matsuo, H. Tsugawa, H. Miyagawa, and E. Fukusaki, "Integrated strategy for unknown ei-ms identification using quality control calibration curve, multivariate analysis, ei-ms spectral database, and retention index prediction," *Analytical chemistry*, vol. 89, no. 12, pp. 6766–6773, 2017.
- [85] 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.
- [86] D. Shen and C. Davatzikos, "Hammer: hierarchical attribute matching mechanism for elastic registration," in *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*. IEEE, 2001, pp. 29–36.
- [87] P.-M. Allard, G. Genta-Jouve, and J.-L. Wolfender, "Deep metabolome annotation in natural products research: towards a virtuous cycle in metabolite identification," *Current opinion in chemical biology*, vol. 36, pp. 40–49, 2017.
- [88] S. Clerens, W. Van den Ende, P. Verhaert, L. Geenen, and L. Arckens, "Sweet substitute: A software tool for in silico fragmentation of peptide-linked n-glycans," *Proteomics*, vol. 4, no. 3, pp. 629–632, 2004.

- [89] S. Wolf, S. Schmidt, M. Müller-Hannemann, and S. Neumann, "In silico fragmentation for computer assisted identification of metabolite mass spectra," *BMC bioinformatics*, vol. 11, no. 1, p. 148, 2010.
- [90] M. Gerlich and S. Neumann, "Metfusion: integration of compound identification strategies," *Journal of Mass Spectrometry*, vol. 48, no. 3, pp. 291–298, 2013.
- [91] 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.
- [92] A. Pelander, E. Tyrkkö, and I. Ojanperä, "In silico methods for predicting metabolism and mass fragmentation applied to quetiapine in liquid chromatography/time-of-flight mass spectrometry urine drug screening," *Rapid Communications in Mass Spectrometry: An International Journal Devoted to the Rapid Dissemination of Up-to-the-Minute Research in Mass Spectrometry*, vol. 23, no. 4, pp. 506–514, 2009.
- [93] G. G. Raffelt, J. Redondo, and N. V. Maira, "The mev mass frontier of axion physics," *Physical Review D*, vol. 84, no. 10, p. 103008, 2011.
- [94] 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.
- [95] T. Haller, M. Kals, T. Esko, R. Mägi, and K. Fischer, "Regscan: a gwas tool for quick estimation of allele effects on continuous traits and their combinations," *Briefings in bioinformatics*, vol. 16, no. 1, pp. 39–44, 2013.
- [96] 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.
- [97] 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.
- [98] R. M. Schueler, "Self-deploying trusses containing shape-memory polymers," 2008.
- [99] 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.
- [100] D. K. Barupal, P. K. Haldiya, G. Wohlgemuth, T. Kind, S. L. Kothari, K. E. Pinkerton, and O. Fiehn, "Metamapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity," *BMC bioinformatics*, vol. 13, no. 1, p. 99, 2012.
- [101] L. B. Jorde, "Linkage disequilibrium as a gene-mapping tool." *American journal of human genetics*, vol. 56, no. 1, p. 11, 1995.
- [102] D. H. Huson and C. Xie, "A poor man's blastx—high-throughput metagenomic protein database search using pauda," *Bioinformatics*, vol. 30, no. 1, pp. 38–39, 2013.

-
- [I03] O. Ostrenko, P. Incardona, R. Ramaswamy, L. Brusch, and I. F. Sbalzarini, "pssalib: The partial-propensity stochastic chemical network simulator," *PLoS computational biology*, vol. 13, no. 12, p. e1005865, 2017.
 - [I04] A. Joyce, U. Z. Ijaz, C. Nzeteu, A. Vaughan, S. L. Shirran, C. H. Botting, C. Quince, V. O'Flaherty, and F. Abram, "Linking microbial community structure and function during the acidified anaerobic digestion of grass," *Frontiers in microbiology*, vol. 9, p. 540, 2018.
 - [I05] A. Marco-Ramell, M. Palau-Rodriguez, A. Alay, S. Tulipani, M. Urpi-Sarda, A. Sanchez-Pla, and C. Andres-Lacueva, "Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data," *BMC bioinformatics*, vol. 19, no. 1, p. 1, 2018.
 - [I06] S. Dimitrov, V. Kamenska, J. Walker, W. Windle, R. Purdy, M. Lewis, and O. Mekenyan, "Predicting the biodegradation products of perfluorinated chemicals using catabol," *SAR and QSAR in Environmental Research*, vol. 15, no. 1, pp. 69–82, 2004.
 - [I07] R. Hernández-de Diego, S. Tarazona, C. Martínez-Mira, L. Balzano-Nogueira, P. Furio-Tari, G. J. Pappas Jr, and A. Conesa, "Paintomics 3: a web resource for the pathway analysis and visualization of multi-omics data," *Nucleic acids research*, vol. 46, no. W1, pp. W503–W509, 2018.
 - [I08] N. Lavender, D. W. Hein, G. Brock, and L. C. R. Kidd, "Evaluation of oxidative stress response related genetic variants, pro-oxidants, antioxidants and prostate cancer," *AIMS medical science*, vol. 2, no. 4, p. 271, 2015.
 - [I09] A. A. Cuellar, C. M. Lloyd, P. F. Nielsen, D. P. Bullivant, D. P. Nickerson, and P. J. Hunter, "An overview of cellml 1.1, a biological model description language," *Simulation*, vol. 79, no. 12, pp. 740–747, 2003.
 - [I10] S. Wigby, L. K. Sirot, J. R. Linklater, N. Buehner, F. C. Calboli, A. Bretman, M. F. Wolfner, and T. Chapman, "Seminal fluid protein allocation and male reproductive success," *Current Biology*, vol. 19, no. 9, pp. 751–757, 2009.
 - [I11] Y. Moriya, D. Shigemizu, M. Hattori, T. Tokimatsu, M. Kotera, S. Goto, and M. Kanehisa, "Pathpred: an enzyme-catalyzed metabolic pathway prediction server," *Nucleic acids research*, vol. 38, no. suppl_2, pp. W138–W143, 2010.
 - [I12] U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, L. Jong, E. Algaa, A. Weidemann, H. Sauer-Danzwith, S. Mir *et al.*, "Sabio-rk—database for biochemical reaction kinetics," *Nucleic acids research*, vol. 40, no. D1, pp. D790–D796, 2011.
 - [I13] A. R. Zomorrodi, M. M. Islam, and C. D. Maranas, "d-optcom: dynamic multi-level and multi-objective metabolic modeling of microbial communities," *ACS synthetic biology*, vol. 3, no. 4, pp. 247–257, 2014.
 - [I14] C. Li, J. Han, Q. Yao, C. Zou, Y. Xu, C. Zhang, D. Shang, L. Zhou, C. Zou, Z. Sun *et al.*, "Subpathway-gm: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways," *Nucleic acids research*, vol. 41, no. 9, pp. e101–e101, 2013.

- [115] P. K. Sreenivasaiah, S. Rani, J. Cayetano, N. Arul, and D. H. Kim, "Ipavs: Integrated pathway resources, analysis and visualization system," *Nucleic acids research*, vol. 40, no. D1, pp. D803–D808, 2011.
 - [116] D. B. Warheit, B. R. Laurence, K. L. Reed, D. H. Roach, G. A. Reynolds, and T. R. Webb, "Comparative pulmonary toxicity assessment of single-wall carbon nanotubes in rats," *Toxicological sciences*, vol. 77, no. 1, pp. 117–125, 2004.
 - [117] J. T. Bates, D. Chivian, and A. P. Arkin, "Glamm: genome-linked application for metabolic maps," *Nucleic acids research*, vol. 39, no. suppl_2, pp. W400–W405, 2011.
 - [118] K. Sriyudthsak, R. F. Mejia, M. Arita, and M. Y. Hirai, "Pasmet: a web-based platform for prediction, modelling and analyses of metabolic systems," *Nucleic acids research*, vol. 44, no. W1, pp. W205–W211, 2016.
 - [119] N. Hadadi, K. C. Soh, M. Seijo, A. Zisaki, X. Guan, M. R. Wenk, and V. Hatzimanikatis, "A computational framework for integration of lipidomics data into metabolic pathways," *Metabolic engineering*, vol. 23, pp. 1–8, 2014.
 - [120] D. Grapov, K. Wanichthanarak, and O. Fiehn, "Metamapr: pathway independent metabolomic network analysis incorporating unknowns," *Bioinformatics*, vol. 31, no. 16, pp. 2757–2760, 2015.
 - [121] 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.
 - [122] 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.
 - [123] M. Vass, N. Allen, C. A. Shaffer, N. Ramakrishnan, L. T. Watson, and J. J. Tyson, "The jigcell model builder and run manager," *Bioinformatics*, vol. 20, no. 18, pp. 3680–3681, 2004.
 - [124] C. Sanford, M. L. Yip, C. White, and J. Parkinson, "Cell++—simulating biochemical pathways," *Bioinformatics*, vol. 22, no. 23, pp. 2918–2925, 2006.
 - [125] F. Jourdan, R. Breitling, M. P. Barrett, and D. Gilbert, "Metanetter: inference and visualization of high-resolution metabolomic networks," *Bioinformatics*, vol. 24, no. 1, pp. 143–145, 2007.
 - [126] S. Pabinger, R. Snajder, T. Hardiman, M. Willi, A. Dander, and Z. Trajanoski, "Memosys 2.0: an update of the bioinformatics database for genome-scale models and genomic data," *Database*, vol. 2014, 2014.
 - [127] 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.
 - [128] 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.
-

-
- [129] R. Carr and E. Borenstein, "Netseed: a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment," *Bioinformatics*, vol. 28, no. 5, pp. 734–735, 2012.
 - [130] M. P. Gerstl, C. Jungreuthmayer, and J. Zanghellini, "tefma: computing thermodynamically feasible elementary flux modes in metabolic networks," *Bioinformatics*, vol. 31, no. 13, pp. 2232–2234, 2015.
 - [131] A. Kreimer, A. Doron-Faigenboim, E. Borenstein, and S. Freilich, "Netcmpt: a network-based tool for calculating the metabolic competition between bacterial species," *Bioinformatics*, vol. 28, no. 16, pp. 2195–2197, 2012.
 - [132] 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.
 - [133] T. Forth, G. A. McConkey, and D. R. Westhead, "Metnetmaker: a free and open-source tool for the creation of novel metabolic networks in sbml format," *Bioinformatics*, vol. 26, no. 18, pp. 2352–2353, 2010.
 - [134] 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.
 - [135] M. Chazalviel, C. Frainay, N. Poupin, F. Vinson, B. Merlet, Y. Gloaguen, L. Cottret, and F. Jourdan, "Metexploreviz: web component for interactive metabolic network visualization," *Bioinformatics*, vol. 34, no. 2, pp. 312–313, 2017.
 - [136] 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.
 - [137] T. Radchenko, F. Fontaine, L. Morettoni, and I. Zamora, "Webmetabase: cleavage sites analysis tool for natural and unnatural substrates from diverse data source," *Bioinformatics*, 2018.
 - [138] S. J. Sheridan and H. L. Springfield, "Age appropriate games in the teaching of leisure skills to persons with mental retardation." 1987.
 - [139] 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.
 - [140] G. Basler and Z. Nikoloski, "Jmassbalance: Mass-balanced randomization and analysis of metabolic networks," *Bioinformatics*, vol. 27, no. 19, pp. 2761–2762, 2011.
 - [141] 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.
 - [142] E. C. Samuelsson, F. A. Victor, G. Tibblin, and K. F. Svärdsudd, "Signs of genital prolapse in a swedish population of women 20 to 59 years of age and possible related factors," *American journal of obstetrics and gynecology*, vol. 180, no. 2, pp. 299–305, 1999.
 - [143] 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.
-

- [144] 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.
- [145] 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, micrnas, and pathway topologies," *Oncotarget*, vol. 6, no. 36, p. 39151, 2015.
- [146] J. Ruiz-Armenta and R. Prol-Ledesma, "Techniques for enhancing the spectral response of hydrothermal alteration minerals in thematic mapper images of central mexico," *International Journal of Remote Sensing*, vol. 19, no. 10, pp. 1981–2000, 1998.
- [147] 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.
- [148] M. Ataman and V. Hatzimanikatis, "lumpgem: Systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites," *PLoS computational biology*, vol. 13, no. 7, p. e1005513, 2017.
- [149] M. Ataman, D. F. H. Gardiol, G. Fengos, and V. Hatzimanikatis, "redgem: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models," *PLoS computational biology*, vol. 13, no. 7, p. e1005444, 2017.
- [150] T. Kieser, "Factors affecting the isolation of ccc dna from streptomyces lividans and escherichia coli," *Plasmid*, vol. 12, no. 1, pp. 19–36, 1984.
- [151] T. Hu, K. Oksanen, W. Zhang, E. Randell, A. Furey, G. Sun, and G. Zhai, "An evolutionary learning and network approach to identifying key metabolites for osteoarthritis," *PLoS computational biology*, vol. 14, no. 3, p. e1005986, 2018.
- [152] 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.
- [153] N. M. Shchegolkova, G. S. Krasnov, A. A. Belova, A. A. Dmitriev, S. L. Kharitonov, K. M. Klimina, N. V. Melnikova, and A. V. Kudryavtseva, "Microbial community structure of activated sludge in treatment plants with different wastewater compositions," *Frontiers in microbiology*, vol. 7, p. 90, 2016.
- [154] 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.
- [155] J. Fayn and P. Rubel, "Caviar: a serial ecg processing system for the comparative analysis of vcgs and their interpretation with auto-reference to the patient," *Journal of electrocardiology*, vol. 21, pp. S173–S176, 1988.
- [156] 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.

-
- [157] D. Waltemath, R. Adams, F. T. Bergmann, M. Hucka, F. Kolpakov, A. K. Miller, I. I. Moraru, D. Nickerson, S. Sahle, J. L. Snoep *et al.*, "Reproducible computational biology experiments with sed-ml—the simulation experiment description markup language," *BMC systems biology*, vol. 5, no. 1, p. 198, 2011.
 - [158] R. Levy, R. Carr, A. Kreimer, S. Freilich, and E. Borenstein, "Netcooperate: a network-based tool for inferring host-microbe and microbe-microbe cooperation," *BMC bioinformatics*, vol. 16, no. 1, p. 164, 2015.
 - [159] Y. Zhu, L. Sun, A. Garbarino, C. Schmidt, J. Fang, and J. Chen, "Pathrings: a web-based tool for exploration of ortholog and expression data in biological pathways," *BMC bioinformatics*, vol. 16, no. 1, p. 165, 2015.
 - [160] H. Zhou, J. Jin, H. Zhang, B. Yi, M. Wozniak, and L. Wong, "Intpath—an integrated pathway gene relationship database for model organisms and important pathogens," in *BMC systems biology*, vol. 6, no. 2. BioMed Central, 2012, p. S2.
 - [161] M. F. Gerard, G. Stegmayer, and D. H. Milone, "Evoms: An evolutionary tool to find de novo metabolic pathways," *Biosystems*, vol. 134, pp. 43–47, 2015.
 - [162] 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.
 - [163] J. Schneider, F.-J. Vorhölter, E. Trost, J. Blom, Y. Musa, H. Neuweiger, K. Niehaus, S. Schatschneider, A. Tauch, A. Goesmann *et al.*, "Carmen-comparative analysis and in silico reconstruction of organism-specific metabolic networks," *Genet Mol Res*, vol. 9, no. 3, pp. 1660–1672, 2010.
 - [164] M. Poolman, "Scrumpy: metabolic modelling with python," *IEE Proceedings-Systems Biology*, vol. 153, no. 5, pp. 375–378, 2006.
 - [165] 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.
 - [166] N. Fyson, M. K. Kim, D. Lun, and C. Colijn, "Gene-centric constraint of metabolic models," *bioRxiv*, p. 116558, 2017.
 - [167] 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.
 - [168] I. L. Hofacker and P. F. Stadler, "Vienna-rnl," 2005.
 - [169] 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.
 - [170] S. Ekins, A. Bugrim, L. Brovold, E. Kirillov, Y. Nikolsky, E. Rakhamatulin, S. Sorokina, A. Ryabov, T. Serebryiskaya, A. Melnikov *et al.*, "Algorithms for network analysis in systems-adme/tox using the metacore and metadrug platforms," *Xenobiotica*, vol. 36, no. 10-11, pp. 877–901, 2006.
-

- [171] Q. Zhang, S. Bhattacharya, D. E. Kline, R. B. Crawford, R. B. Conolly, R. S. Thomas, N. E. Kaminski, and M. E. Andersen, "Stochastic modeling of b lymphocyte terminal differentiation and its suppression by dioxin," *BMC systems biology*, vol. 4, no. 1, p. 40, 2010.
- [172] 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.
- [173] M. Kankainen, P. Gopalacharyulu, L. Holm, and M. Orešič, "Mpea—metabolite pathway enrichment analysis," *Bioinformatics*, vol. 27, no. 13, pp. 1878–1879, 2011.
- [174] A. Kamburov, R. Cavill, T. M. Ebbels, R. Herwig, and H. C. Keun, "Integrated pathway-level analysis of transcriptomics and metabolomics data with impala," *Bioinformatics*, vol. 27, no. 20, pp. 2917–2918, 2011.
- [175] M. Chagoyen and F. Pazos, "Mbrole: enrichment analysis of metabolomic data," *Bioinformatics*, vol. 27, no. 5, pp. 730–731, 2011.
- [176] 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.
- [177] D. K. Barupal and O. Fiehn, "Chemical similarity enrichment analysis (chemrich) as alternative to biochemical pathway mapping for metabolomic datasets," *Scientific reports*, vol. 7, no. 1, p. 14567, 2017.
- [178] 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.
- [179] M. Kaliszewski, F. Athias-Binche, and E. E. Lindquist, "Parasitism and parasitoidism in tarsonemina (acari: Heterostigmata) and evolutionary considerations," in *Advances in Parasitology*. Elsevier, 1995, vol. 35, pp. 335–367.
- [180] 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.
- [181] 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.
- [182] B. Usadel, F. Poree, A. Nagel, M. Lohse, A. CZEDIK-EYSENBERG, and M. Stitt, "A guide to using mapman to visualize and compare omics data in plants: a case study in the crop species, maize," *Plant, cell & environment*, vol. 32, no. 9, pp. 1211–1229, 2009.
- [183] T.-C. Kuo, T.-F. Tian, and Y. J. Tseng, "3omics: a web-based systems biology tool for analysis, integration and visualization of human transcriptomic, proteomic and metabolomic data," *BMC systems biology*, vol. 7, no. 1, p. 64, 2013.

- [I84] 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.
- [I85] 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.
- [I86] S. J. Hinshaw, A. HY Lee, E. E. Gill, and R. EW Hancock, "Metabridge: enabling network-based integrative analysis via direct protein interactors of metabolites," *Bioinformatics*, vol. 34, no. 18, pp. 3225–3227, 2018.
- [I87] 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.
- [I88] H. Neuweger, M. Persicke, S. P. Albaum, T. Bekel, M. Dondrup, A. T. Hüser, J. Winnebald, J. Schneider, J. Kalinowski, and A. Goesmann, "Visualizing post genomics data-sets on customized pathway maps by prometra—aeration-dependent gene expression and metabolism of corynebacterium glutamicum as an example," *BMC systems biology*, vol. 3, no. 1, p. 82, 2009.
- [I89] E. K. Speliotes, C. J. Willer, S. I. Berndt, K. L. Monda, G. Thorleifsson, A. U. Jackson, H. L. Allen, C. M. Lindgren, J. Luan, R. Mägi *et al.*, "Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index," *Nature genetics*, vol. 42, no. 11, p. 937, 2010.
- [I90] 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.
- [I91] A. Kaever, M. Landesfeind, K. Feussner, A. Mosblech, I. Heilmann, B. Morgenstern, I. Feussner, and P. Meinicke, "Marvis-pathway: integrative and exploratory pathway analysis of non-targeted metabolomics data," *Metabolomics*, vol. 11, no. 3, pp. 764–777, 2015.
- [I92] Y. Tikunov, S. Laptenok, R. Hall, A. Bovy, and R. De Vos, "Msclust: a tool for unsupervised mass spectra extraction of chromatography-mass spectrometry ion-wise aligned data," *Metabolomics*, vol. 8, no. 4, pp. 714–718, 2012.
- [I93] 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.
- [I94] 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.
- [I95] K. Hiller, J. Hangebrauk, C. Jager, J. Spura, K. Schreiber, and D. Schomburg, "Metabolitedetector: comprehensive analysis tool for targeted and nontargeted gc/ms based metabolome analysis," *Analytical chemistry*, vol. 81, no. 9, pp. 3429–3439, 2009.

- [196] 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.
- [197] F. Giacomoni, G. Le Corguillé, M. Monsoor, M. Landi, P. Pericard, M. Pétéra, C. Duperier, M. Tremblay-Franco, J.-F. Martin, D. Jacob *et al.*, "Workflow4metabolomics: a collaborative research infrastructure for computational metabolomics," *Bioinformatics*, vol. 31, no. 9, pp. 1493–1495, 2014.
- [198] N. Kessler, H. Neuweger, A. Bonte, G. Langenkämper, K. Niehaus, T. W. Nattkemper, and A. Goesmann, "Metldb 2.0—advances of the metabolomics software system," *Bioinformatics*, vol. 29, no. 19, pp. 2452–2459, 2013.
- [199] 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.
- [200] A. M. Miguez, M. P. McNerney, and M. P. Styczynski, "Metabolomics analysis of the toxic effects of the production of lycopene and its precursors," *Frontiers in microbiology*, vol. 9, p. 760, 2018.
- [201] A. J. Carroll, M. R. Badger, and A. H. Millar, "The metabolomeexpress project: enabling web-based processing, analysis and transparent dissemination of gc/ms metabolomics datasets," *BMC bioinformatics*, vol. 11, no. 1, p. 376, 2010.
- [202] 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.
- [203] C. Williamson and T. Cluse-Tolar, "Pimp-controlled prostitution: Still an integral part of street life," *Violence against women*, vol. 8, no. 9, pp. 1074–1092, 2002.
- [204] L. P. de Souza, T. Naake, T. Tohge, and A. R. Fernie, "From chromatogram to analyte to metabolite. how to pick horses for courses from the massive web-resources for mass spectral plant metabolomics," *GigaScience*, 2017.
- [205] 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.
- [206] Z. Lai, H. Tsugawa, G. Wohlgemuth, S. Mehta, M. Mueller, Y. Zheng, A. Ogiwara, J. Meissen, M. Showalter, K. Takeuchi *et al.*, "Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics," *Nature methods*, vol. 15, no. 1, p. 53, 2018.
- [207] A. G. de la Fuente, J. Godzien, M. F. López, F. J. Rupérez, C. Barbas, and A. Otero, "Knowledge-based metabolite annotation tool: Ceu mass mediator," *Journal of pharmaceutical and biomedical analysis*, vol. 154, pp. 138–149, 2018.
- [208] J. A. Winchester and P. A. Floyd, "Geochemical discrimination of different magma series and their differentiation products using immobile elements," *Chemical geology*, vol. 20, pp. 325–343, 1977.

- [209] K. Dührkop, H. Shen, M. Meusel, J. Rousu, and S. Böcker, "Searching molecular structure databases with tandem mass spectra using csi: Fingerid," *Proceedings of the National Academy of Sciences*, vol. 112, no. 41, pp. 12 580–12 585, 2015.
- [210] J. Wandy, Y. Zhu, J. J. van der Hooft, R. Daly, M. P. Barrett, and S. Rogers, "Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry," *Bioinformatics*, vol. 34, no. 2, pp. 317–318, 2017.
- [211] C. Bueschl, B. Kluger, N. K. Neumann, M. Doppler, V. Maschietto, G. G. Thallinger, J. Meng-Reiterer, R. Krska, and R. Schuhmacher, "Metextract ii: A software suite for stable isotope-assisted untargeted metabolomics," *Analytical chemistry*, vol. 89, no. 17, pp. 9518–9526, 2017.
- [212] E. Kligun, B. Ostretsov, A. Titievsky, M. Farkov, S. Alamouti, and L. Brodsky, "Adaptation of the pine fungal pathogen *grosmannia clavigera* to monoterpenes: Biochemical mechanisms revealed by rna-seq analysis," *Forest Pathology*, vol. 47, no. 6, p. e12372, 2017.
- [213] A. Lommen and H. J. Kools, "Malign 3.0: performance enhancement by efficient use of advances in computer hardware," *Metabolomics*, vol. 8, no. 4, pp. 719–726, 2012.
- [214] F. Allen, A. Pon, M. Wilson, R. Greiner, and D. Wishart, "Cfm-id: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra," *Nucleic acids research*, vol. 42, no. W1, pp. W94–W99, 2014.
- [215] D. J. Creek, A. Jankevics, K. E. Burgess, R. Breitling, and M. P. Barrett, "Ideom: an excel interface for analysis of lc-ms-based metabolomics data," *Bioinformatics*, vol. 28, no. 7, pp. 1048–1049, 2012.
- [216] A. Alonso, A. Julià, A. Beltran, M. Vinaixa, M. Diaz, L. Ibañez, X. Correig, and S. Marsal, "Astream: an r package for annotating lc/ms metabolomic data," *Bioinformatics*, vol. 27, no. 9, pp. 1339–1340, 2011.
- [217] 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.
- [218] F. Wilhelms, "Explaining the dielectric properties of firn as a density-and-conductivity mixed permittivity (decomp)," *Geophysical Research Letters*, vol. 32, no. 16, 2005.
- [219] M. Rojas-Chertó, M. van Vliet, J. E. Peironcely, R. van Doorn, M. Kooyman, T. te Beek, M. A. van Driel, T. Hankemeier, and T. Reijmers, "Metitree: a web application to organize and process high-resolution multi-stage mass spectrometry metabolomics data," *Bioinformatics*, vol. 28, no. 20, pp. 2707–2709, 2012.
- [220] 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.
- [221] N. Sakurai, T. Narise, J.-S. Sim, C.-M. Lee, C. Ikeda, N. Akimoto, and S. Kanaya, "Uc2 search: using unique connectivity of uncharged compounds for metabolite annotation by database searching in mass spectrometry-based metabolomics," *Bioinformatics*, vol. 34, no. 4, pp. 698–700, 2017.

- [222] 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.
- [223] F. Nicolè, Y. Guitton, 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.
- [224] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical review letters*, vol. 77, no. 18, p. 3865, 1996.
- [225] L. F. Feick and L. L. Price, "The market maven: A diffuser of marketplace information," *Journal of marketing*, vol. 51, no. 1, pp. 83–97, 1987.
- [226] M. P. Styczynski, J. F. Moxley, L. V. Tong, J. L. Walther, K. L. Jensen, and G. N. Stephanopoulos, "Systematic identification of conserved metabolites in gc/ms data for metabolomics and biomarker discovery," *Analytical chemistry*, vol. 79, no. 3, pp. 966–973, 2007.
- [227] 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.
- [228] 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.
- [229] 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.
- [230] H. Tsugawa, T. Kind, R. Nakabayashi, D. Yukihira, W. Tanaka, T. Cajka, K. Saito, O. Fiehn, and M. Arita, "Hydrogen rearrangement rules: computational ms/ms fragmentation and structure elucidation using ms-finder software," *Analytical chemistry*, vol. 88, no. 16, pp. 7946–7958, 2016.
- [231] 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.
- [232] H. Treutler, H. Tsugawa, A. Porzel, K. Gorzolka, A. Tissier, S. Neumann, and G. U. Balcke, "Discovering regulated metabolite families in untargeted metabolomics studies," *Analytical chemistry*, vol. 88, no. 16, pp. 8082–8090, 2016.
- [233] 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.

- [234] R. Zhou, C.-L. Tseng, T. Huan, and L. Li, "Isoms: automated processing of lc-ms data generated by a chemical isotope labeling metabolomics platform," *Analytical chemistry*, vol. 86, no. 10, pp. 4675–4679, 2014.
- [235] H. Li, Y. Cai, Y. Guo, F. Chen, and Z.-J. Zhu, "Metdia: targeted metabolite extraction of multiplexed ms/ms spectra generated by data-independent acquisition," *Analytical chemistry*, vol. 88, no. 17, pp. 8757–8764, 2016.
- [236] H. Salvesen, S. Carter, M. Mannelqvist, A. Dutt, G. Getz, I. a. Stefansson, M. Raeder, M. L. Sos, I. Engelsen, J. Trovik *et al.*, "Integrated genomic profiling of endometrial carcinoma associates aggressive tumors with indicators of pi3 kinase activation," *Proceedings of the National Academy of Sciences*, vol. 106, no. 12, pp. 4834–4839, 2009.
- [237] S. J. McIlroy, R. H. Kirkegaard, B. McIlroy, M. Nierychlo, J. M. Kristensen, S. M. Karst, M. Albertsen, and P. H. Nielsen, "Midas 2.0: an ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups," *Database*, vol. 2017, 2017.
- [238] 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.
- [239] 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.
- [240] P. W. Jones, "St. george's respiratory questionnaire: Mcid," *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 2, no. 1, pp. 75–79, 2005.
- [241] W. M. Edmands, L. Petrick, D. K. Barupal, A. Scalbert, M. J. Wilson, J. K. Wickliffe, and S. M. Rappaport, "compms2miner: An automatable metabolite identification, visualization, and data-sharing r package for high-resolution lc-ms data sets," *Analytical chemistry*, vol. 89, no. 7, pp. 3919–3928, 2017.
- [242] 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.
- [243] Y. Fukuda, T. Hayakawa, E. Ichihara, K. Inoue, K. Ishihara, H. Ishino, Y. Itow, T. Kajita, J. Kameda, S. Kasuga *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, no. 8, p. 1562, 1998.
- [244] Y.-C. Harn, B.-H. Su, Y.-L. Ku, O. A. Lin, C.-F. Chou, and Y. J. Tseng, "Np-structurepredictor: Prediction of unknown natural products in plant mixtures," *Journal of chemical information and modeling*, vol. 57, no. 12, pp. 3138–3148, 2017.
- [245] 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.
- [246] N. Kessler, F. Walter, M. Persicke, S. P. Albaum, J. Kalinowski, A. Goesmann, K. Niehaus, and T. W. Nattkemper, "Allocator: an interactive web platform for the analysis of metabolomic lc-esi-ms datasets, enabling semi-automated, user-revised compound annotation and mass isotopomer ratio analysis," *PLoS One*, vol. 9, no. 11, p. e113909, 2014.

- [247] D. Shang, C. Li, Q. Yao, H. Yang, Y. Xu, J. Han, J. Li, F. Su, Y. Zhang, C. Zhang *et al.*, "Prioritizing candidate disease metabolites based on global functional relationships between metabolites in the context of metabolic pathways," *PLoS one*, vol. 9, no. 8, p. e104934, 2014.
- [248] R. Winkler, "Spidermass: semantic database creation and tripartite metabolite identification strategy," *Journal of Mass Spectrometry*, vol. 50, no. 3, pp. 538–541, 2015.
- [249] J. Draper, D. P. Enot, D. Parker, M. Beckmann, S. Snowdon, W. Lin, and H. Zubair, "Metabolite signal identification in accurate mass metabolomics data with mzeddb, an interactive m/z annotation tool utilising predicted ionisation behaviour'rules'," *BMC bioinformatics*, vol. 10, no. 1, p. 227, 2009.
- [250] K. Skogerson, G. Wohlgemuth, D. K. Barupal, and O. Fiehn, "The volatile compound binbase mass spectral database," *BMC bioinformatics*, vol. 12, no. 1, p. 321, 2011.
- [251] N. Sakurai, T. Ara, M. Enomoto, T. Motegi, Y. Morishita, A. Kurabayashi, Y. Iijima, Y. Ogata, D. Nakajima, H. Suzuki *et al.*, "Tools and databases of the komics web portal for preprocessing, mining, and dissemination of metabolomics data," *BioMed research international*, vol. 2014, 2014.
- [252] 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.
- [253] J. Jeong, X. Shi, X. Zhang, S. Kim, and C. Shen, "An empirical bayes model using a competition score for metabolite identification in gas chromatography mass spectrometry," *BMC bioinformatics*, vol. 12, no. 1, p. 392, 2011.
- [254] T. Suvitalval, S. Rogers, and S. Kaski, "Stronger findings from mass spectral data through multi-peak modeling," *BMC bioinformatics*, vol. 15, no. 1, p. 208, 2014.
- [255] 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.
- [256] M. Scholz and O. Fiehn, "Setup x-a public study design database for metabolomic projects," in *Biocomputing 2007*. World Scientific, 2007, pp. 169–180.
- [257] E. Kenar, H. Franken, S. Forcisi, K. Wörmann, H.-U. Häring, R. Lehmann, P. Schmitt-Kopplin, A. Zell, and O. Kohlbacher, "Automated label-free quantification of metabolites from liquid chromatography–mass spectrometry data," *Molecular & cellular proteomics*, vol. 13, no. 1, pp. 348–359, 2014.
- [258] B. Bonn, C. Leandersson, F. Fontaine, and I. Zamora, "Enhanced metabolite identification with mse and a semi-automated software for structural elucidation," *Rapid Communications in Mass Spectrometry*, vol. 24, no. 21, pp. 3127–3138, 2010.
- [259] E. Compe and J.-M. Egly, "Tfih: when transcription met dna repair," *Nature reviews Molecular cell biology*, vol. 13, no. 6, p. 343, 2012.

- [260] R. Bannuru, N. Natov, U. Dasi, C. Schmid, and T. McAlindon, "Therapeutic trajectory following intra-articular hyaluronic acid injection in knee osteoarthritis—meta-analysis," *Osteoarthritis and cartilage*, vol. 19, no. 6, pp. 611–619, 2011.
- [261] 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.
- [262] 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.
- [263] A. Corma, "From microporous to mesoporous molecular sieve materials and their use in catalysis," *Chemical reviews*, vol. 97, no. 6, pp. 2373–2420, 1997.
- [264] S. Claus, I. Puscalau-Girtu, P. Walther, T. Syrovets, T. Simmet, C. Haupt, and M. Fändrich, "Cell-to-cell transfer of saal protein in a cell culture model of systemic aa amyloidosis," *Scientific reports*, vol. 7, p. 45683, 2017.
- [265] J. W. Schopf, "Microfossils of the early archean apex chert: new evidence of the antiquity of life," *Science*, vol. 260, no. 5108, pp. 640–646, 1993.
- [266] 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.
- [267] Y. Simon-Manso, M. S. Lowenthal, L. E. Kilpatrick, M. L. Sampson, K. H. Telu, P. A. Rudnick, W. G. Mallard, D. W. Bearden, T. B. Schock, D. V. Tchekhovskoi *et al.*, "Metabolite profiling of a nist standard reference material for human plasma (srm 1950): Gc-ms, lc-ms, nmr, and clinical laboratory analyses, libraries, and web-based resources," *Analytical chemistry*, vol. 85, no. 24, pp. 11725–11731, 2013.
- [268] 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.
- [269] 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.
- [270] 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.
- [271] S. C. Booth, A. M. Weljie, and R. J. Turner, "Computational tools for the secondary analysis of metabolomics experiments," *Computational and structural biotechnology journal*, vol. 4, no. 5, p. e201301003, 2013.
- [272] P. Meinicke, T. Lingner, A. Kaever, K. Feussner, C. Göbel, I. Feussner, P. Karlovsky, and B. Morgenstern, "Metabolite-based clustering and visualization of mass spectrometry data using one-dimensional self-organizing maps," *Algorithms for Molecular Biology*, vol. 3, no. 1, p. 9, 2008.

- [273] R. Steuer, J. Kurths, O. Fiehn, and W. Weckwerth, "Observing and interpreting correlations in metabolomic networks," *Bioinformatics*, vol. 19, no. 8, pp. 1019–1026, 2003.
- [274] D.-S. Lee, J. Park, K. Kay, N. A. Christakis, Z. Oltvai, and A.-L. Barabási, "The implications of human metabolic network topology for disease comorbidity," *Proceedings of the National Academy of Sciences*, vol. 105, no. 29, pp. 9880–9885, 2008.
- [275] J. Xia and D. S. Wishart, "Msea: a web-based tool to identify biologically meaningful patterns in quantitative metabolomic data," *Nucleic acids research*, vol. 38, no. suppl_2, pp. W71–W77, 2010.