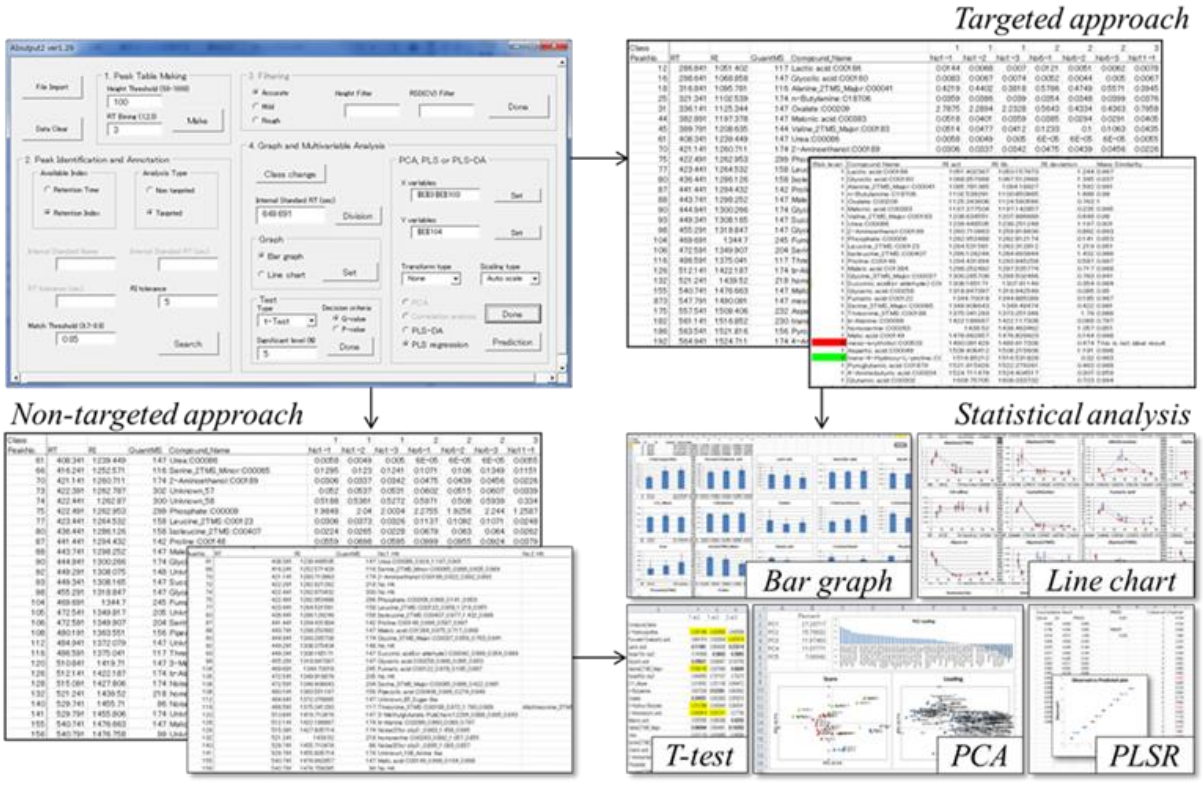
AIoutput tutorial

Objective

GC/MS is one of the most popular platforms for comprehensive analysis of metabolites in living organisms. The crucial process is to construct an organized two-dimensional data matrix containing compound names and their quantitative values. Because this process is the most complicated and knowledge intensive task in GC/MS-based metabolomics, it is essential to develop a tool for accurate, automatic data processing. We used the MetAlign (Lommen, 2009) data pre-processing tool. AIoutput can perform the peak identification, prediction, and data integration from the result exported from MetAlign and user defined retention time and spectra library. AIoutput is a non-targeted and targeted analysis tool for GC/MS based metabolomics written in visual basic for application (VBA, excel macro) available in Microsoft Excel Windows versions 2007 and later.



Data pre-processing

We use MetAlign software for data preprocessing of GC/MS raw data. MetAlign offers the data matrix aligned with the RT and m/z without missing values by performing the noise

estimation, the binomial based digital filter (for smoothing), the peak estimation program (for peak detection), and the peak alignment program based on the amplitude and retention time.

GC/MS library

AIoutput identifies the detected peaks on the basis of the reference retention time and mass spectra included in the user defined GC/MS library. Both the retention time (RT) and retention index (RI) based on n-alkane mixture is available as index.

At 2014/2/1, we offer two type of GC/MS libraries. Both libraries stores the retention time and mass spectra of the derivative metabolites by methoxyamine and MSTFA.

1. Pegasus III TOF-MS system, LECO; GC 6890, Agilent Technologies; Column CP-SIL 8 CB LOW BLEED/MS (Cotributor Osaka Univ., Japan)
2. GCMS-2010 Plus, Shimadzu; Column InertCap 5MS/NP (Contributor GL Science, Japan; Copyright GLS)

The detail of analytical condition is downloadable together with the library.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	CompoundName::ID	RT	RI	QuantMS	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99
2	Propyleneglycol::C00583	245.966	989.303	117	16	3	27	12	17	2	1	0	0	2	10	1	1	0	
3	C10_Alkane::	253.143	1000.000	85	999	67	11	9	9	5	8	6	7	9	11	10	21	254	305
4	n-Propylamine::PubChem:7852	264.863	1017.466	174	29	722	84	32	4	0	0	0	0	1	1	1	2	9	18
5	2-Hydroxypyridine::C02502	275.693	1033.607	152	5	4	0	0	0	0	2	17	16	6	20	13	37	11	11
6	Pyruvate+Oxalacetic acid::C00022+C00036	281.244	1041.879	174	11	20	9	6	680	61	32	2	3	0	0	0	0	12	227
7	Noise1(for silyl)::	286.294	1049.405	174	8	169	42	25	6	0	12	2	0	0	2	0	1	0	6
8	Lactic acid::C00186	286.799	1050.158	117	6	11	46	99	18	1	2	1	1	5	1	0	0	0	4
9	a-MethylBenzylamine_1_Minor::C02455	290.670	1055.927	106	3	3	3	1	13	5	34	6	5	1	0	0	0	0	0
10	Glycolic acid::C00160	298.444	1067.513	147	2	2	16	46	23	4	2	1	1	0	11	1	0	0	3
11	N-Methylethanolamine::PubChem:8016	302.501	1073.558	116	3	31	14	27	17	3	0	1	0	0	1	2	0	4	2
12	Valine_1TMS_Minor::C00183	310.593	1085.618	146	135	203	555	145	61	18	199	19	35	0	5	0	12	8	1
13	Isobutylamine::C02787	313.743	1090.313	174	24	714	71	30	2	0	0	0	0	0	0	1	2	4	14
14	Alanine_2TMS_Major::C00041	316.344	1094.189	116	3	27	9	9	3	0	0	1	1	18	2	0	0	3	2
15	Noise6(for silyl)::	317.244	1095.531	147	0	154	0	0	422	24	0	0	0	0	0	16	0	0	0
16	C11_Alkane::	320.243	1100.000	85	999	69	14	13	12	9	13	17	17	16	17	18	46	263	239
17	n-Butylamine::C18706	320.794	1100.851	174	26	596	64	34	4	0	0	0	0	3	1	1	2	10	22
18	Ketovaline_1_Major::C00141	321.308	1101.645	89	18	36	4	8	999	88	43	2	2	0	1	29	5	10	43
19	Glycine_2TMS_Minor::C00037	329.894	1114.905	102	8	41	35	12	0	0	0	2	0	0	9	0	1	0	0
20	2-HydroxyButyrate::C05984	332.113	1118.332	131	26	4	22	5	9	1	1	1	1	2	27	3	2	1	25
21	Oxalate::C00209	336.159	1124.581	147	2	1	17	3	8	0	1	1	4	0	1	0	0	0	1
22	Ketovaline_2_Minor::C00141	338.031	1127.472	89	40	157	17	18	999	96	45	5	0	2	0	49	10	29	100
23	Sarcosine::C00213	339.929	1130.403	116	1	29	13	9	3	0	0	0	0	2	0	0	0	0	1
24	2-Aminoisobutyrate::C03665	343.808	1136.394	130	5	6	7	2	2	0	0	0	0	2	3	2	2	9	4
25	3-Hydroxybutyrate::C01089	354.454	1152.835	147	9	2	18	147	20	1	2	1	0	0	0	0	1	1	25
26	4-Hydroxypyridine::PubChem:12290	355.993	1155.212	152	26	8	1	0	0	1	5	19	22	14	17	9	8	5	139
27	2-Aminobutyric acid::C02261	362.044	1164.558	130	9	22	10	4	3	1	0	1	1	1	3	3	1	18	5
28	Ketoisoleucine_1_Major::PubChem:2724058	367.303	1172.680	89	169	59	11	11	999	87	51	3	3	9	2	3	7	29	54
29	Ketoisoleucine_2_Minor::PubChem:2724058	379.984	1192.264	89	207	178	31	15	999	89	60	5	4	5	3	7	21	52	80
30	Glyceraldehyde_1_Minor::C00577	382.059	1195.469	103	35	54	59	48	884	96	23	0	0	0	0	0	0	0	21
31	Malonic acid::C00383	383.143	1197.143	147	3	2	18	9	6	1	1	1	3	0	1	0	0	1	47
32	Norleucine_1TMS_Minor::C01933	383.494	1197.685	86	8	999	78	5	3	0	4	2	1	0	0	0	0	0	4
33	C12_Alkane::	384.993	1200.000	85	999	58	9	7	6	5	7	0	5	4	9	16	58	207	202
34	1-Aminocyclopropane-1-carboxylic acid::C01234	386.494	1202.498	147	50	47	23	7	4	2	1	2	2	6	1	11	6	62	21

Data processing

AIoutput can integrate the MetAlign result to the mass spectra tags (MSTs), identify the MSTs on the basis of the retention time similarity (based on gaussian function) and mass spectra (based on correlation coefficient) with the reference, and predict unknown peaks by mulitivariate analysis model (based on soft independent modeling of class analogy).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Class				1	1	1	2	2	2	3	3	3
2	PeakNo.	RT	QuantMS	Compound_Name	No1-1	No1-2	No1-3	No6-1	No6-2	No6-3	No11-1	No11-2	No11-3
3	12	276.291	152	2-Hydroxypyridine	0.096376	0.078623	0.064011	0.098055	0.064505	0.074903	0.159564	0.112311	0.12983
4	14	281.241	174	Oxalacetic acid	0.006475	0.004506	0.004605	0.006674	0.004897	0.005632	0.006578	0.005506	0.005604
5	18	286.991	130	Noise1(for silyl)	0.015323	0.012709	0.013172	0.015287	0.011267	0.01391	0.014747	0.01203	0.014245
6	27	316.841	116	Alanine(2TMS)	0.421901	0.440166	0.381843	0.578632	0.474943	0.557118	0.394482	0.405794	0.329057
7	29	317.841	147	Unknown_OrganicAcid	0.058661	0.14289	0.142693	0.141435	0.1489	0.170282	0.197942	0.104172	0.135581
8	31	317.941	204	Noise6(for silyl)	0.009983	0.024407	0.024603	0.023644	0.025508	0.029718	0.033141	0.017566	0.023772
9	35	321.341	174	n-Butylamine	0.035857	0.038589	0.038979	0.035423	0.034761	0.039894	0.037631	0.029716	0.040229
10	42	336.141	147	Oxalate	2.787463	2.289362	2.232833	0.564343	0.433361	0.436316	0.79576	0.777262	0.795257
11	49	354.941	147	Unknown_OrganicAcid	0.017567	0.015078	0.010735	0.020706	0.016615	0.016368	0.020514	0.013826	0.022002
12	59	371.991	174	Unknown_Amine	0.002863	0.003495	0.004043	0.002767	0.003335	0.004014	0.003242	0.003172	0.003746
13	61	382.891	147	Malonic acid	0.051799	0.04012	0.035901	0.038475	0.029383	0.029096	0.040468	0.032051	0.028078
14	62	389.791	144	Valine(2TMS)	0.051361	0.047746	0.041175	0.123324	0.099958	0.106333	0.043523	0.041477	0.037073
15	63	389.841	147	Unknown_OrganicAcid	0.009467	0.00751	0.006318	0.020877	0.016825	0.017924	0.007077	0.006464	0.005869
16	68	400.091	174	Unknown_Amine	0.024584	0.021865	0.023934	0.025384	0.019409	0.024428	0.02438	0.022265	0.023388
17	84	408.341	147	Urea	0.005804	0.004881	0.004953	5.7E-05	6.01E-05	6.22E-05	0.005487	0.004579	0.004778
18	91	416.291	147	Serine(2TMS)	0.015272	0.014615	0.014939	0.01215	0.012559	0.015871	0.013375	0.008858	0.015484
19	95	421.141	174	2-Aminoethanol	0.03062	0.033708	0.034241	0.047544	0.043895	0.045589	0.022603	0.02412	0.023683
20	98	422.391	302	Unknown	0.05198	0.053667	0.053142	0.060179	0.051496	0.060744	0.03392	0.032559	0.034478
21	99	422.441	300	Unknown	0.518844	0.536091	0.527187	0.597114	0.507992	0.593901	0.333967	0.322899	0.338524
22	100	422.491	299	Pyrophosphate	1.984832	2.039976	2.003427	2.275512	1.92558	2.243971	1.258706	1.222289	1.270188
23	102	423.441	158	Leucine	0.03062	0.03729	0.032554	0.113656	0.109212	0.107142	0.024786	0.027592	0.022769
24	106	436.441	158	Isoleucine	0.022366	0.026516	0.022756	0.067851	0.063003	0.064042	0.026189	0.027592	0.022887
25	108	438.041	117	Threonine(2TMS)	0.01749	0.018226	0.017027	0.028264	0.031997	0.040019	0.021138	0.015531	0.022798
26	114	441.441	142	Proline	0.055875	0.069813	0.05954	0.09988	0.095451	0.09236	0.03788	0.04447	0.033593
27	117	444.941	174	Glycine(3TMS)	0.032736	0.03183	0.02977	0.039587	0.035392	0.042166	0.026843	0.024958	0.024745
28	117	444.941	174	Unknown_Amine	0.032736	0.03183	0.02977	0.039587	0.035392	0.042166	0.026843	0.024958	0.024745
29	119	449.291	148	Unknown	0.075842	0.075734	0.07065	0.055188	0.049273	0.056138	0.060265	0.05665	0.057571
30	120	449.341	147	Succinic acid	0.475222	0.476474	0.444114	0.345131	0.310329	0.346133	0.378769	0.353573	0.358108
31	126	455.291	147	Glyceric acid	0.044344	0.044135	0.043156	0.038332	0.034731	0.03784	0.079657	0.077328	0.076211
32	136	469.691	245	Fumaric acid	0.01878	0.019208	0.017482	0.020079	0.019138	0.021659	0.018426	0.017596	0.017814
33	139	472.591	204	Serine(3TMS)	0.315388	0.408971	0.377962	0.315555	0.311261	0.357554	0.243617	0.287886	0.23574
34	142	475.741	188	Alanine(2TMS)	0.000284	0.001415	0.000222	0.000485	0.001962	0.004543	0.000274	0.001247	0.002274

Statistical analysis

Upon constructing the matrix, AIoutput can normalize the variables by the internal standard and perform further statistical and graphical analyses. The software can visualize the bar or line chart of each metabolite, or perform principal component analysis, projection to latent structure (PLS), PLS discriminant analysis (PLS-DA) and statistical tests.