LipidSearch 5.2: Improved Software for Parallel Processing of Large Datasets from LC-MS High Resolution Mass Spectrometry Based Lipidomics Workflows

[1]David A. Peake, [2]Yuko Kitahashi, [2]Noritaka Masaki and [2]Yasuto Yokoi, [1]Peake Performance LLC, Indianapolis, IN, USA; [2]Mitsui Knowledge Industry Co., Tokyo, Japan

1. Background

- Thermo Scientific™ LipidSearch™ software version 4.2 introduced in 2018, enabled lipid annotation from LC-HRMS, ddMS^n data
- LipidSearch software version 5.0 introduced in 2021, was a complete software recode
- LipidSearch software version 5.1 introduced in 2023, enabled sequential execution of processes
- *Two part-processing*:
  1. Search lipid database for accurate *m/z* and predict MS2 product ions
  2. Alignment of lipid positive and negative ions, differential analysis of multiple groups
- For more information, please visit **https://thermofisher.com/lipidsearch**

2. Problem

- Lipid annotation is challenging due to the complexity of LC-HRMS[2] data
- Processing of large datasets is time consuming due to thousands of MS and MS2 spectra per file
- Correlation (alignment) of datasets > 100 files is very slow or fails to complete

3. Methods

- LipidSearch software version 5.2: Implemented RAM memory and cache management
- Tested alpha and beta builds on 3 workstations and 5 datasets
- Beta testing with 12 customers from 8 laboratories

Test systems

System	CPU	RAM	SSD Data drive	Windows OS	ID-Align-Files
1. Z840	2X E5-2667 8C 3.5 GHz	128 GB	4 TB 2.5" 870-QVO	10 Pro 64	32-32-20
2. Z8G4	2X Gold 6148 20C 3.7 GHz	384 GB	4 TB 990-PRO NVME M.2	11 Pro 64	79-50-79
3. Lenovo	i9-12900 16C 5.1 GHz *8P+8E*	128 GB	4TB XG8 NVME M.2	11 Pro 64	16-16-12

Test datasets

Dataset	# files	pos	neg	GB	MS	Method	LC 2x150mm	R, MS	R, MS2
FB1 mouse liver	12	6	6	1.1	Q Exactive	30 min ddMS2	Ascentis C18 2.7µm 55°C	70 K	17 K
SRM1950 QC	44	22	22	8.3	Exploris 240	27 min ddMS2	Acclaim C30 3µm 45°C	120 K	30 K
DIO mouse plasma	55	55		18	Exploris 240	28 min AX	Accucore C30 2.6µm 45°C	120 K	30 K
NIST 1, 2, 3	72	36	36	16	Q Exactive HF	60 min ddMS2	Prototype C30 *25cm* 2µm 45°C	120 K	30 K
P250	774	386	388	145	Fusion Lumos	25 min ddMS2	Acclaim C30 3µm 45°C	120 K	15 K

Search conditions

Dataset	mass tol. MS/MS2	Filter	Database selection
FB1 mouse liver	5 ppm / 10 ppm	Rank = 1 Grade ≠ D	mammalian lipids + Hydroxy-CAR + CL
DIO mouse plasma	3 ppm / 5 ppm		mammalian lipids + Hydroxy-CAR + d9-SM
SRM1950 QC	3 ppm / 5 ppm		mammalian lipids + Hydroxy-CAR + d9-SM
NIST 1,2,3	5 ppm / 10 ppm		mammalian lipids + Hydroxy-CAR
P250	5 ppm / 10 ppm		mammalian lipids + Hydroxy-CAR + d9-SM

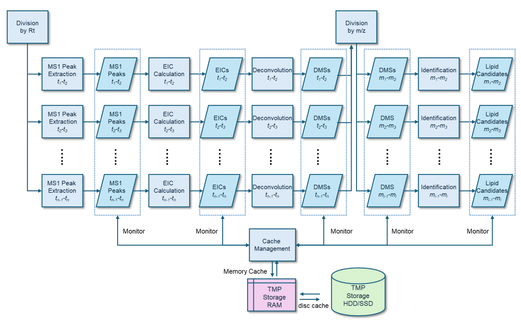
Alignment conditions

Dataset	Rt tol. / Corr. tol.	Merge	Sample groups
FB1 mouse liver	0.10 min / 0.10 min	mean, manual filter	Control, Fumonisin B1
DIO mouse plasma	0.10 min / 0.25 min		M normal, M high-fat, F normal, F high-fat, QC, ID
SRM1950 QC	0.10 min / 0.25 min		QC
NIST 1, 2, 3	0.10 min / 0.25 min		SRM1950, SRM2378-1, SRM2378-2, SRM2378-3
P250	0.15 min / 0.25 min		Control, DCIS, IDC, ILC, QC

4. LipidSearch software version 5.2 workflow strategy

- Increased speed through appropriate parallel processing:
  1. **Peak extraction.** The time axis is divided into the LCMS BPC, MS1 peak extraction, EIC peak extraction, and MS2 deconvolution processes which are executed in parallel for each time region
  2. **Lipid identification process.** Parallel search for lipid candidates matching the deconvolved MS2 spectra (DMS) is performed by *m/z* group
  3. **Alignment process.** Parallel calculation of a series of processes (time-axis correction, normalization, grouping, and statistical calculations) is performed for each lipid class based on the identification results of multiple samples
- Realtime monitoring of parallel processes avoids memory overloads and processing delays caused by frequent memory allocations; caching of large amounts of intermediate data at the appropriate time enables large-scale parallel computing of the results

LipidSearch



5. Results

- Progress on speed of processing large datasets from version 5.1.6 to 5.2.2.1
- Search and alignment times for 5 datasets ranging from 1 GB to 145 GB
- Best overall search performance from System 2 (8 performance cores up to 5GHz)
- System 1 (40C) best performance for alignment of the large P250 dataset (145 GB)

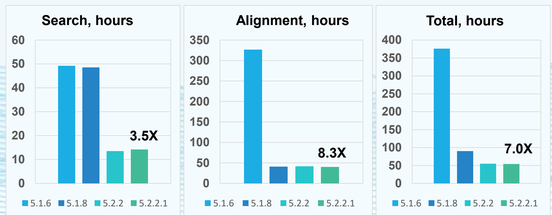
FB1 mouse liver search results

FB1 mouse liver		Alignment result	
File	Group	Groups	Ions
01n	Control	346	701
01p		728	1339
02n		283	603
02p		701	1375
03n		316	612
03p		711	1333
04n	FB1	358	783
04p		820	1553
05n		344	728
05p		765	1496
06n		322	673
06p		771	1448

Alignment results

LC-MS/MS Dataset	Lenovo 8C	
	Molec.	Ions
FB1 mouse liver	2174	2959
SRM1950 QC	1160	1689
DIO mouse plasma	1445	1709
NIST 1	4439	5837
NIST 2	4941	6380
NIST 3	2504	3204
P250	5944	8570

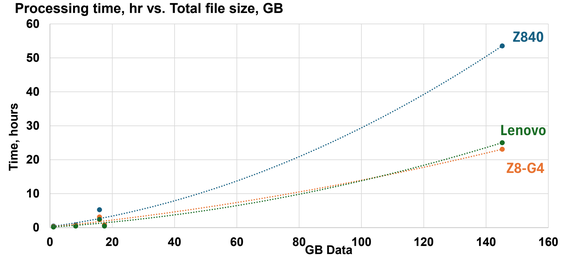
Large dataset performance vs LipidSearch software build



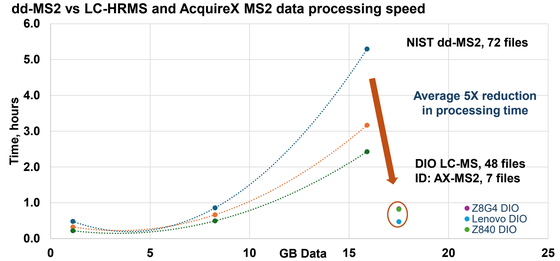
LipidSearch software version 5.2.2.1 processing time in hours

LC-MS2 datasets	Z840 16C			Z8G4, 40C			Lenovo 8pC		
Dataset	Search	Align	Total	Search	Align	Total	Search	Align	Total
FB1 mouse liver	0.43	0.05	0.48	0.28	0.05	0.33	0.20	0.03	0.23
SRM1950 QC	0.72	0.14	0.86	0.53	0.14	0.67	0.42	0.08	0.50
DIO mouse plasma	0.25	0.57	0.82	0.25	0.58	0.83	0.15	0.32	0.47
NIST 1-3	4.42	0.88	5.30	2.70	0.45	3.15	2.01	0.42	2.43
P250	14.1	39.5	53.6	10.7	12.4	23.1	9.87	15.1	25.0

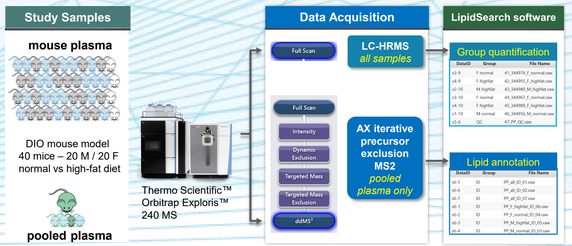
Processing time, hr vs. Total file size, GB



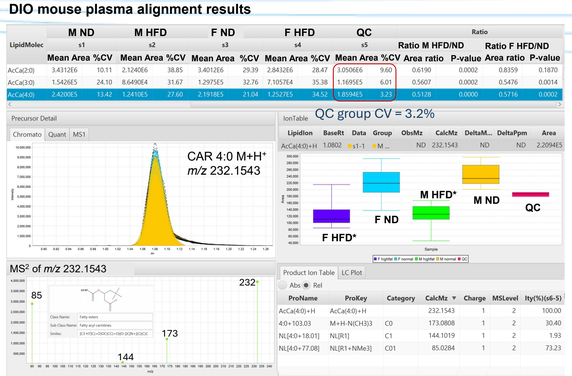
dd-MS2 vs LC-HRMS and AcquireX MS2 data processing speed



LC-MS and AcquireX MS2 workflow for acquisition and data processing



DIO mouse plasma alignment results



6. Conclusions

- Performance was greatly improved by parallel processing of peak detection, lipid annotation and alignment processes in addition to memory allocation management and caching of intermediate data
- Sufficient (minimum 128GB) RAM, 4TB NVME M.2 SSDs and 4-5 GHz processor speed contributed to the best data processing performance
- Using the LC-HRMS and AcquireX MS2 workflow greatly reduces by ~5 fold the processing overhead associated with searching redundant data acquired using LC-ddMS2 for every sample injection

7. Acknowledgements

- The authors are grateful for the participation of the beta software testers including Ifat Abramovich, Ben Crossett, Mark David, Anthony Don, Daniel Gachotte, Beatriz Abad Garcia, Michael Isay-Del Viscio, Desmond Li, Phil Lorenzei, Henrik Molina, Russell Pickford and Bo Wei
- The authors thank BCAL Diagnostics for the use of the Lenovo PC and for kindly providing the use of the P250 dataset
- The Whitehead Institute, Fikadu Tafesse, Hidde Ploegh and Lisa Freinkman kindly provided the FB1 mouse liver data
- COI. The authors declare no competing financial interest
- The authors are affiliated with a third-party organization, and this work represents independent third-party content related to LipidSearch software