

2023.8.30

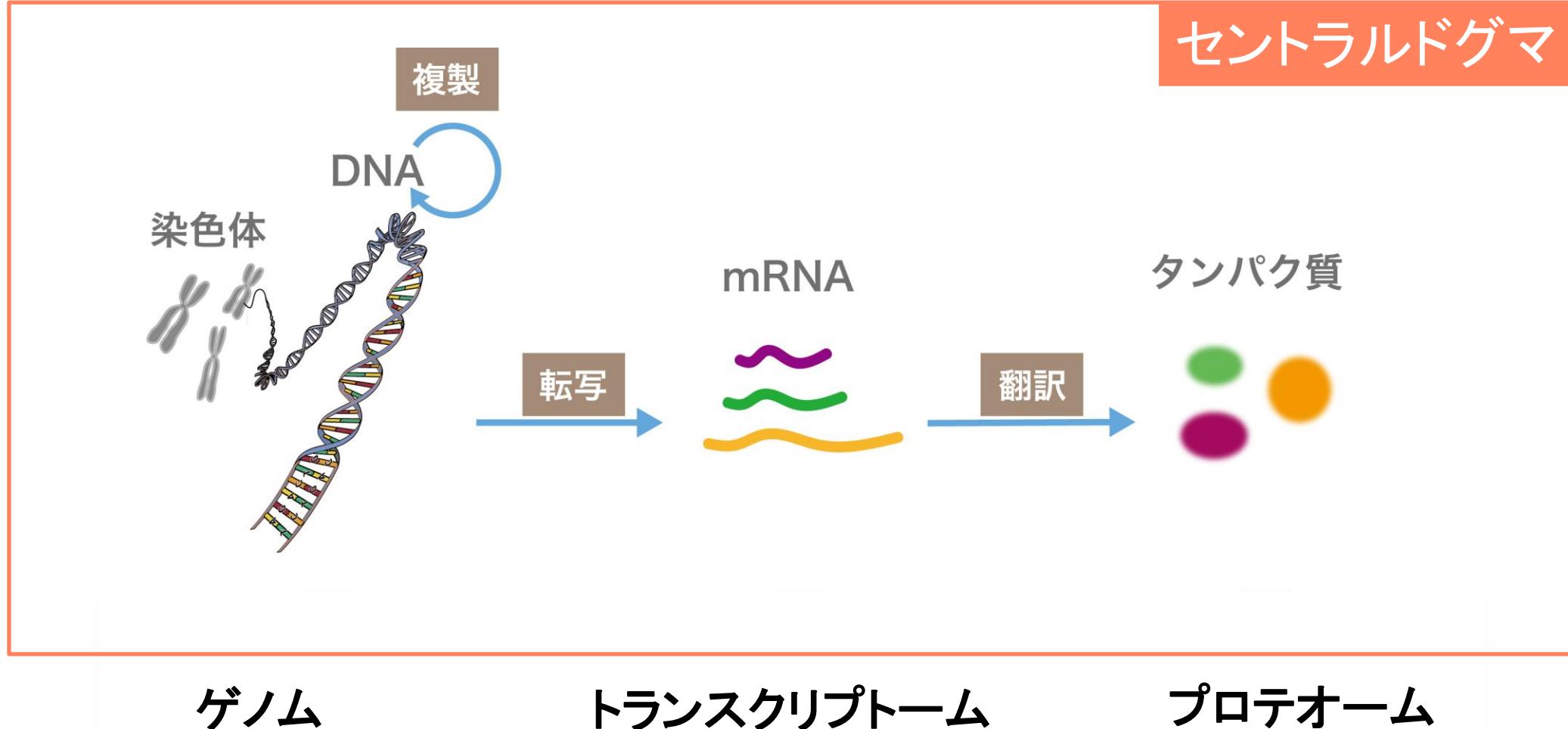
AJACSオンライン17

質量分析データを解析に活用する

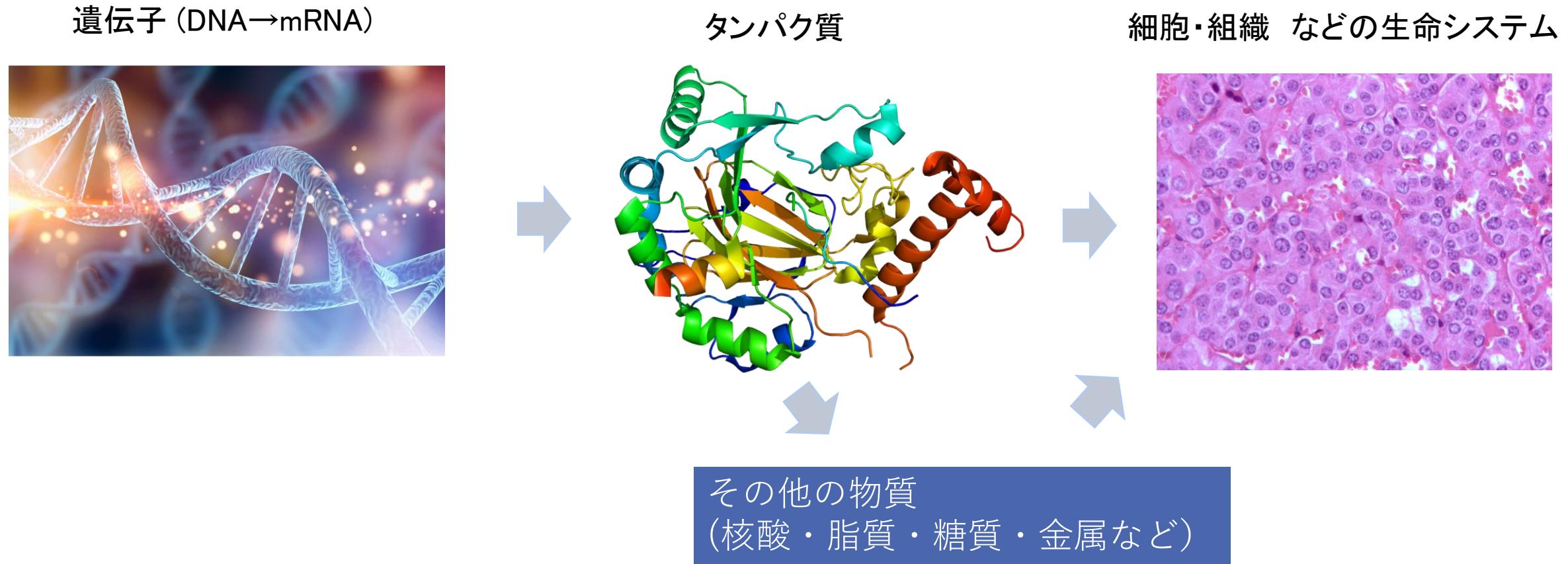
新潟大学大学院 医歯学総合研究科
オミクス生物学分野

松本 雅記

遺伝子はタンパク質の設計図である



ゲノムを基盤とする生命システムの成り立ち

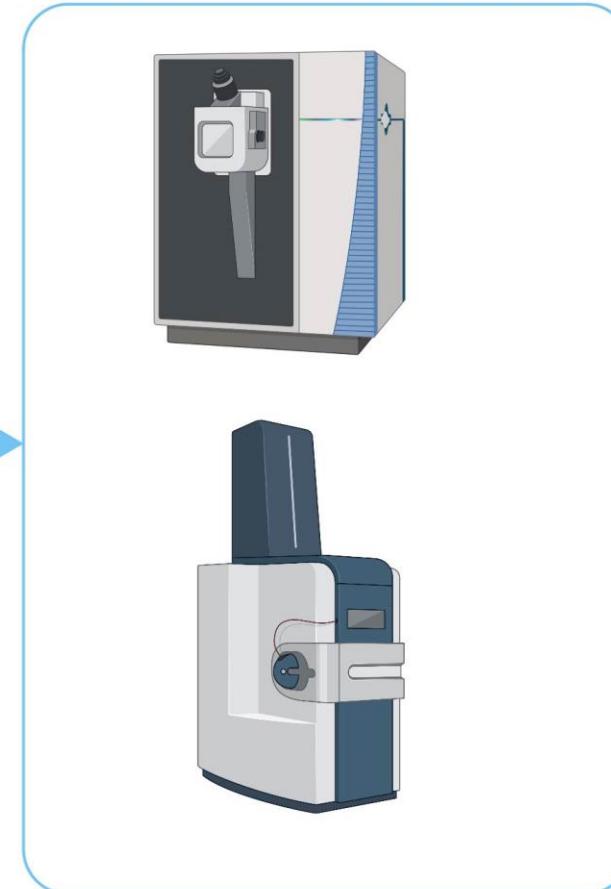


ゲノム情報と分子計測技術の組み合わせが 可能にしたポストゲノム研究

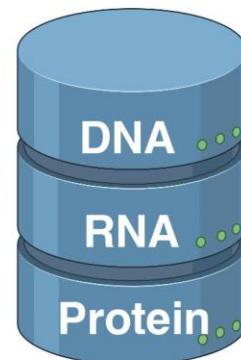
Next-generation sequencer



Mass spectrometry



Sequence DB



核酸

タンパク質

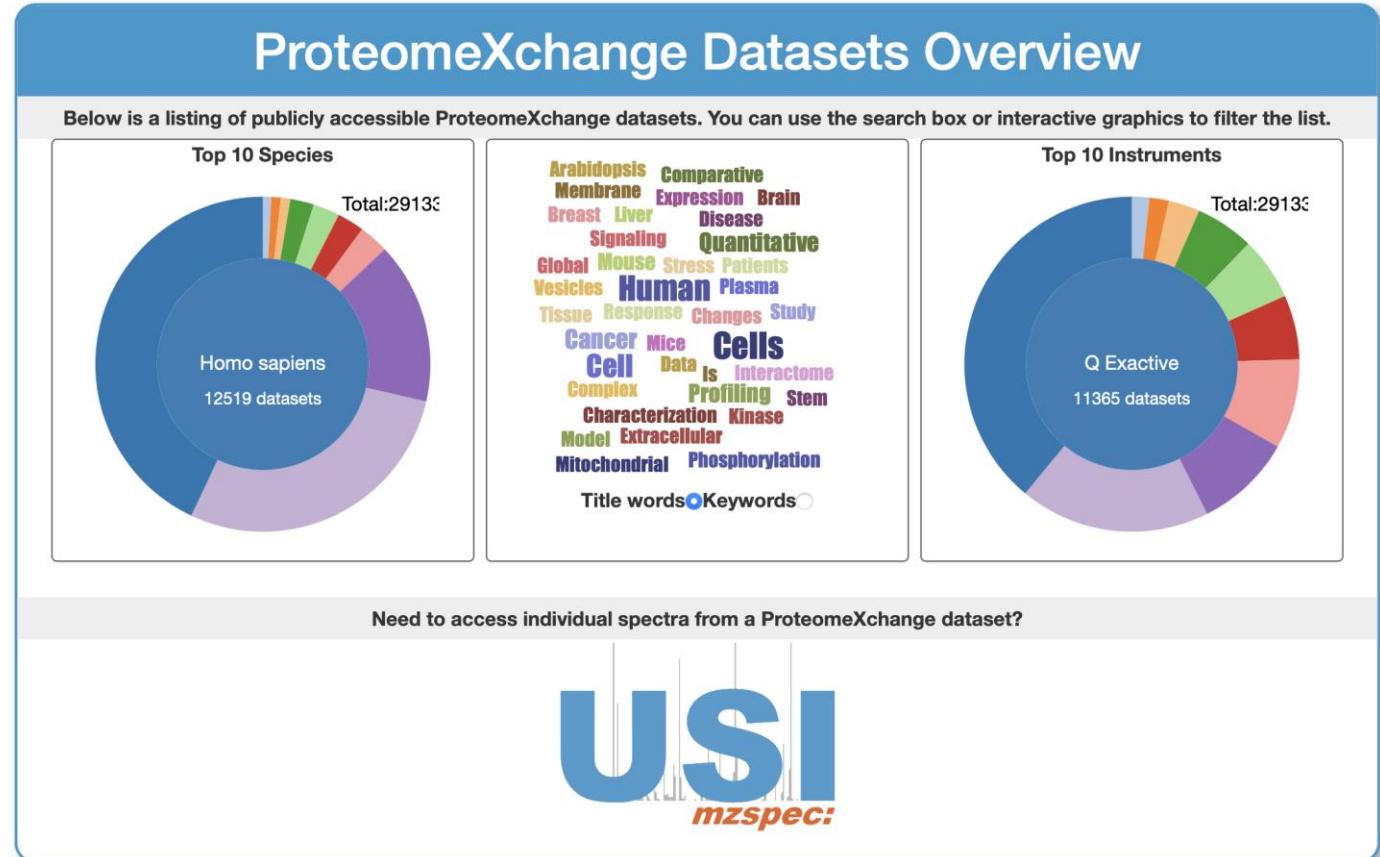
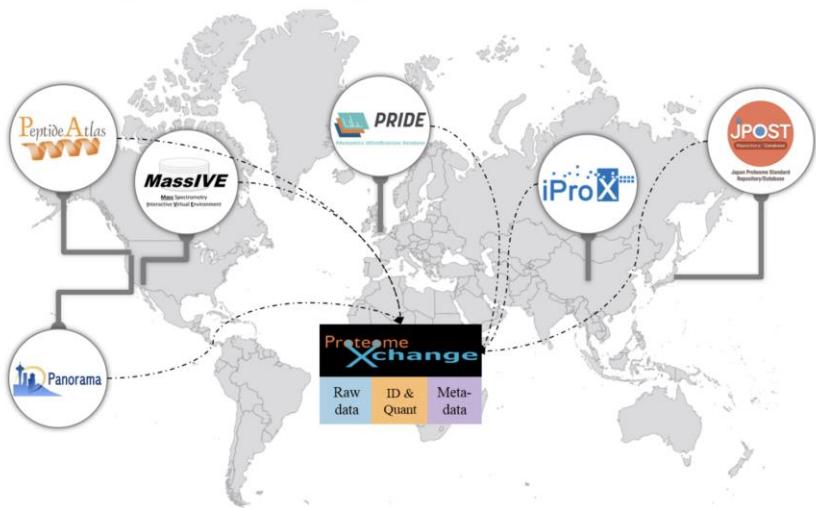
蓄積するプロテオームデータ



Mission

The ProteomeXchange Consortium was established to provide globally coordinated standard data submission and dissemination pipelines involving the main proteomics repositories, and to encourage open data policies in the field. Please review our [Data Submission Guidelines](#), [Guidelines for Reprocessed datasets](#) and [PX Membership Agreement](#).

See also the [original Nature Biotechnology publication](#) and the 2017 and 2020 update papers.





Japan Proteome Standard
Repository/Database

Recent posts

other

JCompMS 8th workshop

⌚ 2023-04-7 🚩 jpost

Workshop details in Japanese

other

PXC has been selected as one of Global Core Biodata Resources.

⌚ 2022-12-19 🚩 jpost

The Global Biodata Coalition (GBC) has announced the first list of Global Core Biodata Resources (GCBRs) – a collection of 37 resources whose long term funding and sustainability is critical to life science and biomedical research worldwide. ProteomeXchange consortium (PXC), of which jPOST is a member, is included in the first list.



Repository [Help ▾](#)

jPOSTRepo is a raw and processed data repository of mass-based proteomics.



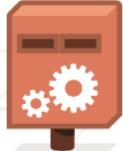
Database [Help ▾](#)

jPOSTdb is a database containing re-analysis results with unified criteria for proteome data from jPOSTRepo.



Workflow [Help ▾](#)

This provides the re-analysis workflow used in jPOST.



Gadgets [Help ▾](#)

jPOST gadget is a collection of tools we have developed as well as other commonly used tools for mass spectrometry-based proteomics.



COVID-19 [Help ▾](#)

This is a special page on coronavirus (SARS-CoV-2) and the COVID-19 disease data, where mass-based proteomics data in ProteomeXchange resources are analyzed by the jPOST re-analysis workflow.



Help

A guide to using the jPOST resources.

jPOSTdb (Japan ProteOme STandard DataBase) is a database containing re-analysis results with unified criteria for proteome data from **jPOSTrepo**. It provides viewers showing the frequency of detected post-translational modifications, the co-occurrence of phosphorylation sites on a peptide and peptide sharing among proteoforms.

Filter ▲

Reset

Species



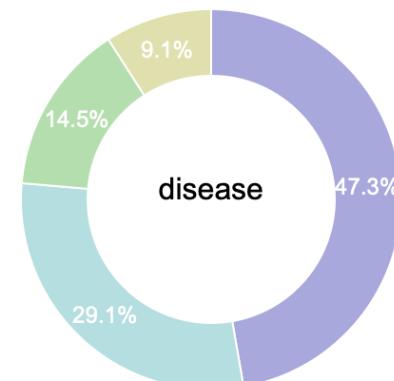
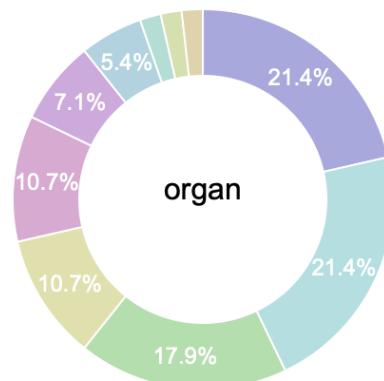
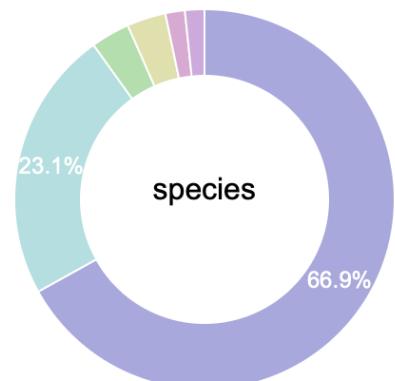
Organ



Disease



Add filter

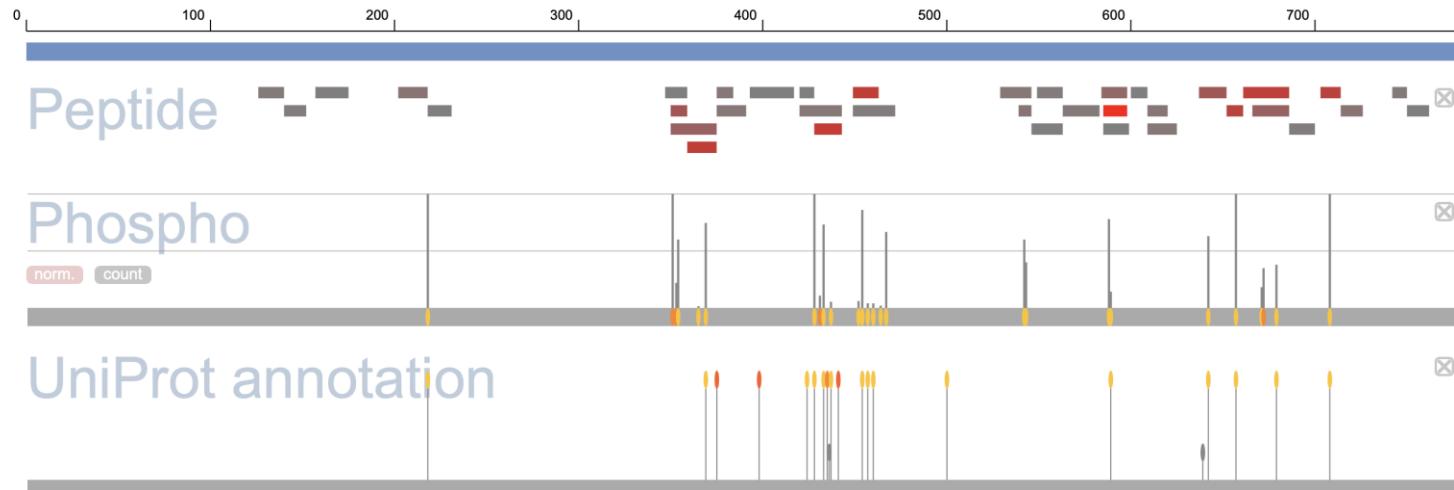


個々のタンパク質情報の検索1

Protein: O14639

Protein Name	Actin-binding LIM protein 1								
Protein ID	ABLM1_HUMAN								
Gene Name	ABLM1								
Accession	O14639								
Length	778 aa								
Sequence	MPAFLGLKCLGKLCSEKSKVTSSESTSARGNSRKRLIVEDRRVSGTSFTAHRRAITIHLYLCPKDYPGRVCNSVDPFAHPODPHHPSEKPVIHCKC6PCKGEVLRVQTKHFHKCFTCKVCGDLAQGGFIKNGEYLCTLDYQRMYGTRCHGEFVEGEVVTALGKTYHPNCFACTICKRPFPGPDRVTNGRDLCLQLCAQPMSSPKETTFSSNCAGCGRDIKNGQALLALDKQWHLGCFCKSCGKVLTGEYISKDGAPYCEKDYQGLFGVKCEACHQFITGKVLEAGDKHYHPSCARCSRCNQMFTEGEEMYLQGSTVWHPDCKQSTKTEEKLRPRTSSEISYSPGSSIPGSPGHTIYAKVDNEILYDYLAAIPKVKAIYDIERPDITYEPFYTSGYDKQERQLGESPRTLSPTPSAEQYDVRDRMIHRSTSQSINSPVSRHSYPTTSSRSPQHFHRPNEPSSGRNSPLPYRPDSRPLPTYAQAPKHFHVPDQGINIYRKPIYKQHAALAAQSKSSEDIKFSKFPAQAQADPSETPKIETDHWPGPSFAVVGPMKRRSSGREEDDEELLRRQLOEEQLMKNGLGQLILKEEMEKESRSSLASRYDSPINSASHIPSSKTAALPGYGRNLHRPVSTDFAQYNSYGDVSGGVRDYQLPDGHMPAMRMMDRGVSMPNMLEPKIFPYEMLMVTNRGRNKILREVDTRLERLHAPEVFREIFGMSIQEFDRPLWRRNDMKKAKLF								
Location	Chromosome 10								
Statistics	<table><tr><td># peptides:</td><td>37</td></tr><tr><td># spectra:</td><td>542</td></tr><tr><td># unique peptides (UniProt entry level):</td><td>36</td></tr><tr><td># unique peptides (gene name level):</td><td>37</td></tr></table>	# peptides:	37	# spectra:	542	# unique peptides (UniProt entry level):	36	# unique peptides (gene name level):	37
# peptides:	37								
# spectra:	542								
# unique peptides (UniProt entry level):	36								
# unique peptides (gene name level):	37								
for Absolute Quantification	GenelD: 3983 search in iMPAQI								

Protein Browser



個々のタンパク質情報の検索2

Protein Name	Tyrosine-protein kinase ABL1
Protein ID	ABL1_HUMAN
Gene Name	ABL1
Accession	P00519
Length	1130 aa
Sequence	MLEICLKLVGCKSKKGSSSSCYLEALQRPVASDFEPQGLSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCEAQTKNGQGVPSNYITPVNSLEKHSWYHPVSRNAEYLSSGINGSFLVRESESSPGQR SISLRYEGRVYINTASDGKLYVSSSERFNTLAEVHHHSTVADGLITLHYAPKRKNKPTVYGVSPNYDKWEMERTDITMKHLGGQYGEVYEGWKKYSLTVAVKTLKEDTMEVEELKEAAVMKEIKHPNLVQLLGVC TREPPFYIITEFMYGN LLDYLRECNRQEVNAVVLVLYMATQISSAMEYLEKKNFIIHDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAKFPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQLVYELLKDYRMRPEGCPEKVYELMRACWQWNPSDR PSFAEIHQAFETMFQESSISDEVEKELGKQGVGRGAVSTLLQAPELPTKTRTSRAAEHRDTTDVPEMHSKGQGESPLDHEPAVSPLLPKERGPPEGGLNEDERLLPKDKTNLFSALIKKKKTAPTPKRSSSFREM DQGPERRGAGEEEGRDISN ALAFATPLTDADPAKSPKPSNGAGVPGNALRESGGSGFRSPHLWKKSSLTSRLATGEEEGGSSSKRFLRSCASCVPHGAKDTEWSVTLPRLQSTGRQFDSSTFGGHKSEKPALPRKRAGENRSQVTRGTVPPTPRLVKKNEEADEVFKDIMESS PGSSPPNLTPKPLRQRVTVAPASGLPHKEAGKGSALGTPAAAEPVPTSKAGSGAPGTSKGPAEESRVRRHKHSESPPGRDGKGLSRLKAPP PPPAASAGKAGGKPSQSPSQEAGEAVLGAKTKATS LVDAVNSDAAKPSQPGEGLKPVLPATPKP QSAKPSGTPISPAPVPSTLPSASSALAGDQPSSTAFIPLISTRVSLRKTRQPPERIASGAIKGVLVLDSTEALCLAISRNEQMASHSAVLEAGKNLYTFCVSYVDSIQQMRNKFAFREAINKLENLRELQICPATAGSGPAATQDFSKLLSSVKEISDI VQR
Location	Chromosome 9
Statistics	# peptides: 28 # spectra: 428 # unique peptides (UniProt entry level): 20 # unique peptides (gene name level): 20

GeneID: 25

Isoform 1

MLEICLKLVGCKSKKGSSSSCYLEALQRPVASDFEPQGLSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCE
1

Isoform 2

MGQQPGKVLGDQRRPSLPALHFIKGAGKKESSRHGGPHCNVFVEHEALQRPVASDFEPQGLSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGEKLRVLGYNHNGEWCE
1 1

[name](#) [color](#) [trypsin](#) [exon](#) [guide](#) [edit](#)

実験グループ(slice)間の比較:sliceの作成

New Slice

Fibroblast

Description

Filter ▲

+

Fibroblast

Page size: 10

Showing 1 to 9 of 9 entries (filtered from 121 entries)

<input checked="" type="checkbox"/>	Dataset ID	Project ID	Project Title
<input checked="" type="checkbox"/>	DS815_1	JPST000815	Human Fibroblast cell_aHDF1388-P9
<input checked="" type="checkbox"/>	DS815_2	JPST000815	Human Fibroblast cell_aHDF1388-P9
<input checked="" type="checkbox"/>	DS815_3	JPST000815	Human Fibroblast cell_aHDF1388-P9
<input checked="" type="checkbox"/>	DS816_1	JPST000816	Human Fibroblast cell_aHDF1419-P10
<input checked="" type="checkbox"/>	DS816_2	JPST000816	Human Fibroblast cell_aHDF1419-P10
<input checked="" type="checkbox"/>	DS816_3	JPST000816	Human Fibroblast cell_aHDF1419-P10
<input checked="" type="checkbox"/>	DS817_1	JPST000817	Human Fibroblast cell_Tig120slc-P8
<input checked="" type="checkbox"/>	DS817_2	JPST000817	Human Fibroblast cell_Tig120slc-P8
<input checked="" type="checkbox"/>	DS817_3	JPST000817	Human Fibroblast cell_Tig120slc-P8

Page: 1

« < < 1 > >>

New Slice

iPS

Description

Filter ▲

+

iPS

Page size: 25

Showing 1 to 25 of 27 entries (filtered from 121 entries)

<input checked="" type="checkbox"/>	Dataset ID	Project ID	Project Title
<input checked="" type="checkbox"/>	DS810_1	JPST000810	Human iPS cell_201B7-P32
<input checked="" type="checkbox"/>	DS810_2	JPST000810	Human iPS cell_201B7-P32
<input checked="" type="checkbox"/>	DS810_3	JPST000810	Human iPS cell_201B7-P32
<input checked="" type="checkbox"/>	DS811_1	JPST000811	Human iPS cell_32R1-P32
<input checked="" type="checkbox"/>	DS811_2	JPST000811	Human iPS cell_32R1-P32
<input checked="" type="checkbox"/>	DS811_3	JPST000811	Human iPS cell_32R1-P32
<input checked="" type="checkbox"/>	DS812_1	JPST000812	Human iPS cell_414C2-P43
<input checked="" type="checkbox"/>	DS812_2	JPST000812	Human iPS cell_414C2-P43
<input checked="" type="checkbox"/>	DS812_3	JPST000812	Human iPS cell_414C2-P43
<input checked="" type="checkbox"/>	DS813_1	JPST000813	Human iPS cell_585A1-P55
<input checked="" type="checkbox"/>	DS813_2	JPST000813	Human iPS cell_585A1-P55
<input checked="" type="checkbox"/>	DS813_3	JPST000813	Human iPS cell_585A1-P55

サンプル群間で発現変化するタンパク質の探索

Spectral counting-based quantification between fibroblasts and iPS cells

Differential Expression Analysis

- The quantification is based on spectral counting.
- Some methods need at least 2 datasets in either slice.

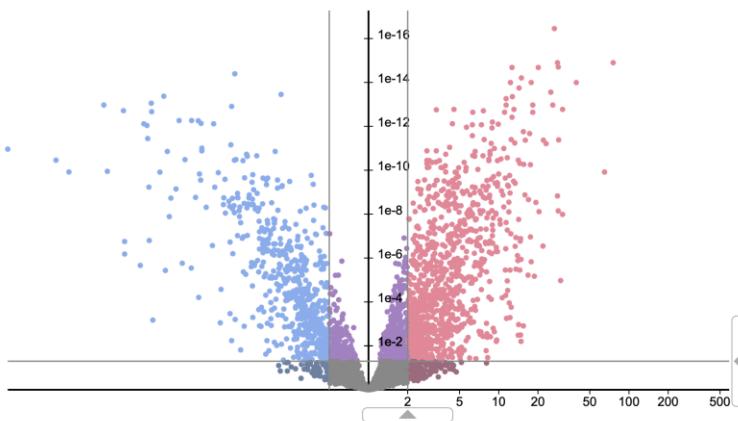
-- Select method --

Fold change >= 2

p-value <= 5e-2

Proteins: 1534

both



Protein name	Accession	ID	Fold change	Showing 1 to 25 of 1534 entries	
				log(fc)	p-value
Fibronectin	P02751	FINC_HUMAN	1 / 590	-9.2	1.09e-11
Collagen alpha-3(VI) chain	P12111	CO6A3_HUMAN	1 / 252	-7.97	3.48e-11
Major vault protein	Q14764	MVP_HUMAN	1 / 200	-7.64	1.20e-10
Basement membrane-specific...	P98160	PGBM_HUMAN	1 / 108	-6.75	1.07e-13
Collagen alpha-2(VI) chain	P12110	CO6A2_HUMAN	1 / 101	-6.66	1.12e-10
Neuroblast differentiation-ass...	Q09666	AHNK_HUMAN	1 / 76.3	-6.25	1.93e-13
Alpha-amino adipic semialdeh...	Q9UDR5	AASS_HUMAN	75.4	6.24	1.26e-15
Nephrilysin	P08473	NEP_HUMAN	1 / 74.4	-6.22	1.76e-07
LIM domain only protein 7	Q8WWI1	LM07_HUMAN	1 / 74.5	-6.22	6.57e-07
Keratin, type I cytoskeletal 18	P05783	K1C18_HUMAN	65.1	6.02	1.21e-10
Collagen alpha-2(I) chain	P08123	CO1A2_HUMAN	1 / 56.5	-5.82	2.17e-06
Vimentin	P08670	VIME_HUMAN	1 / 53.1	-5.73	7.66e-13
Collagen alpha-1(VI) chain	P12109	CO6A1_HUMAN	1 / 50	-5.65	9.13e-13
Unconventional myosin-Ic	O00159	MYO1C_HUMAN	1 / 49.5	-5.63	3.60e-12
EH domain-containing protein 2	Q9ZN4	EHD2_HUMAN	1 / 48.4	-5.6	5.88e-10
Sulfide:quione oxidoreducta...	Q9Y6N5	SQQR_HUMAN	1 / 48.1	-5.59	1.59e-07
Calpain-2 catalytic subunit	P17655	CAN2_HUMAN	1 / 46.5	-5.54	8.84e-14
Annexin A1	P04083	ANXA1_HUMAN	1 / 46.1	-5.53	2.15e-13
Chondroitin sulfate proteoglyc...	Q6UVK1	CSPG4_HUMAN	1 / 45.2	-5.5	6.71e-04
Reticulocalbin-3	Q96D15	RCN3_HUMAN	1 / 39.8	-5.32	1.22e-10
Protein RCC2	Q9P258	RCC2_HUMAN	39.5	5.3	1.01e-14
1,4-alpha-glucan-branching e...	Q04446	GLGB_HUMAN	1 / 37.4	-5.22	4.24e-14
Peptidyl-prolyl cis-trans isome...	O95302	FKBP9_HUMAN	1 / 36.2	-5.18	3.68e-06
Caveolae-associated protein 1	Q6NZI2	CAVN1_HUMAN	1 / 35	-5.13	1.39e-11
Unconventional myosin-IId	O94832	MYO1D_HUMAN	1 / 33.8	-5.08	1.29e-08



Japan Proteome Standard
Repository/Database

Recent posts

other

JCompMS 8th workshop

⌚ 2023-04-7 🚩 jpost

Workshop details in Japanese

other

PXC has been selected as one of Global Core Biodata Resources.

⌚ 2022-12-19 🚩 jpost

The Global Biodata Coalition (GBC) has announced the first list of Global Core Biodata Resources (GCBRs) – a collection of 37 resources whose long term funding and sustainability is critical to life science and biomedical research worldwide. ProteomeXchange consortium (PXC), of which jPOST is a member, is included in the first list.



Repository Help ▾

jPOSTRepo is a raw and processed data repository of mass-based proteomics.



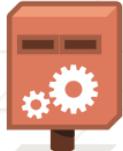
Database Help ▾

jPOSTdb is a database containing re-analysis results with unified criteria for proteome data from jPOSTRepo.



Workflow Help ▾

This provides the re-analysis workflow used in jPOST.



Gadgets Help ▾

jPOST gadget is a collection of tools we have developed as well as other commonly used tools for mass spectrometry-based proteomics.



COVID-19 Help ▾

This is a special page on coronavirus (SARS-CoV-2) and the COVID-19 disease data, where mass-based proteomics data in ProteomeXchange resources are analyzed by the jPOST re-analysis workflow.



Help

A guide to using the jPOST resources.

リポジトリにあるrawデータを活用するには？

こんな場合に有効

- ・検索対象データベースを変えたい (mutation, riboseqから作成したfastaなど)
- ・自分のデータや異なるプロジェクトのデータを同一条件で比較したい
- ・異なる解析プラットフォームを使いたい

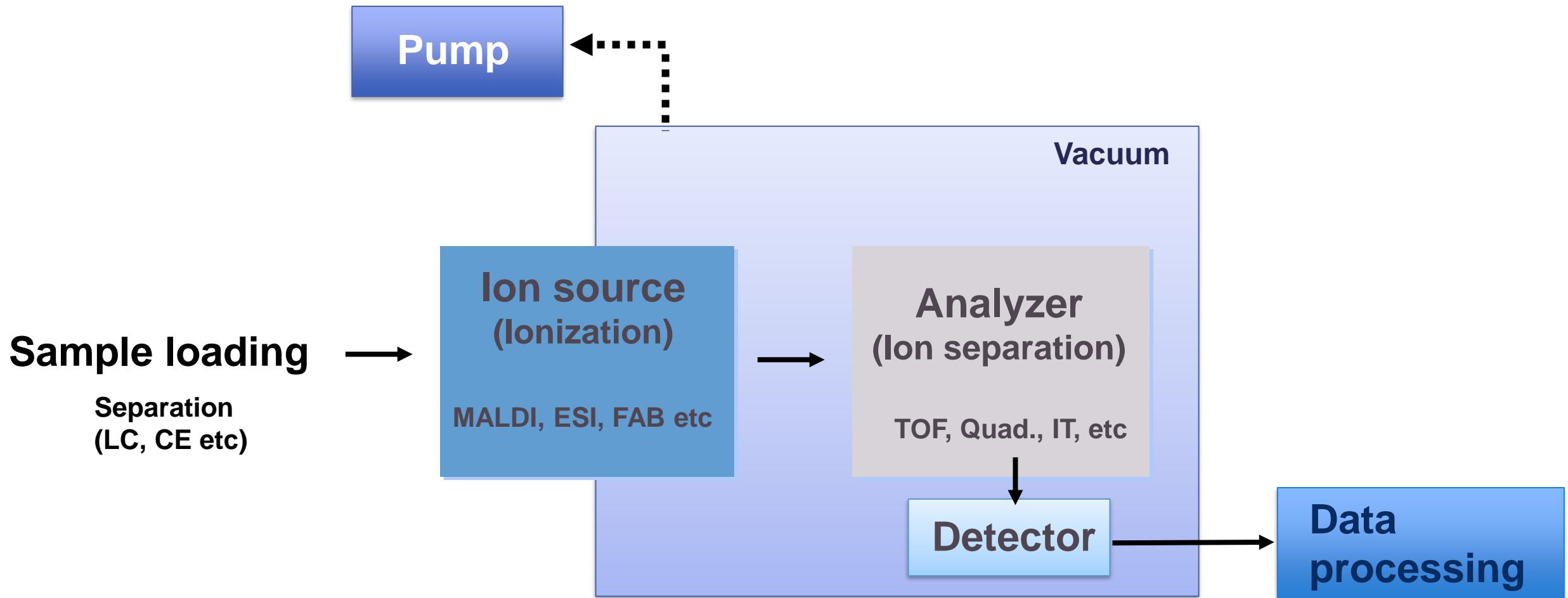
- ・装置や測定手法によって得られるデータは多様。
- ・実験的な前処理法との組み合わせによる複雑化。
- ・リポジトリにおけるメタデータの不足や誤りが多い



- 質量分析計によるプロテオーム解析の原理の理解
- データの種類やデータ解析手法の理解

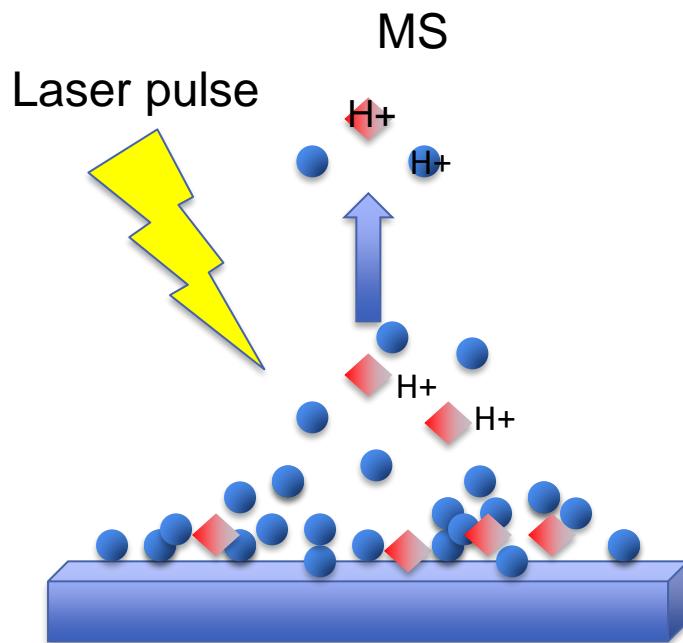
質量分析計によるプロテオーム解析の原理

質量分析計の概要

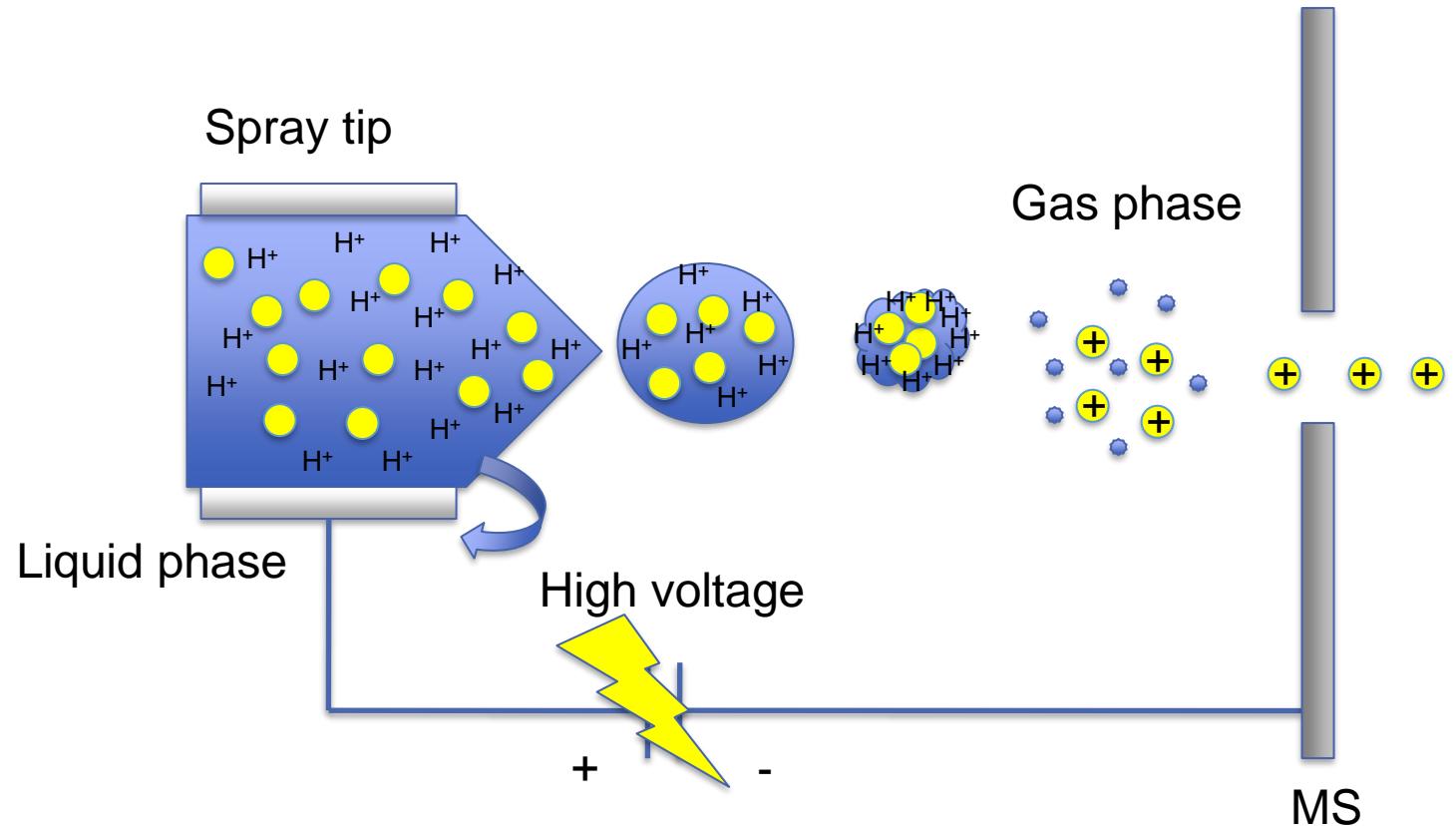


ソフトイオン化法

Matrix Assisted Laser Desorption Ionization (MALDI)



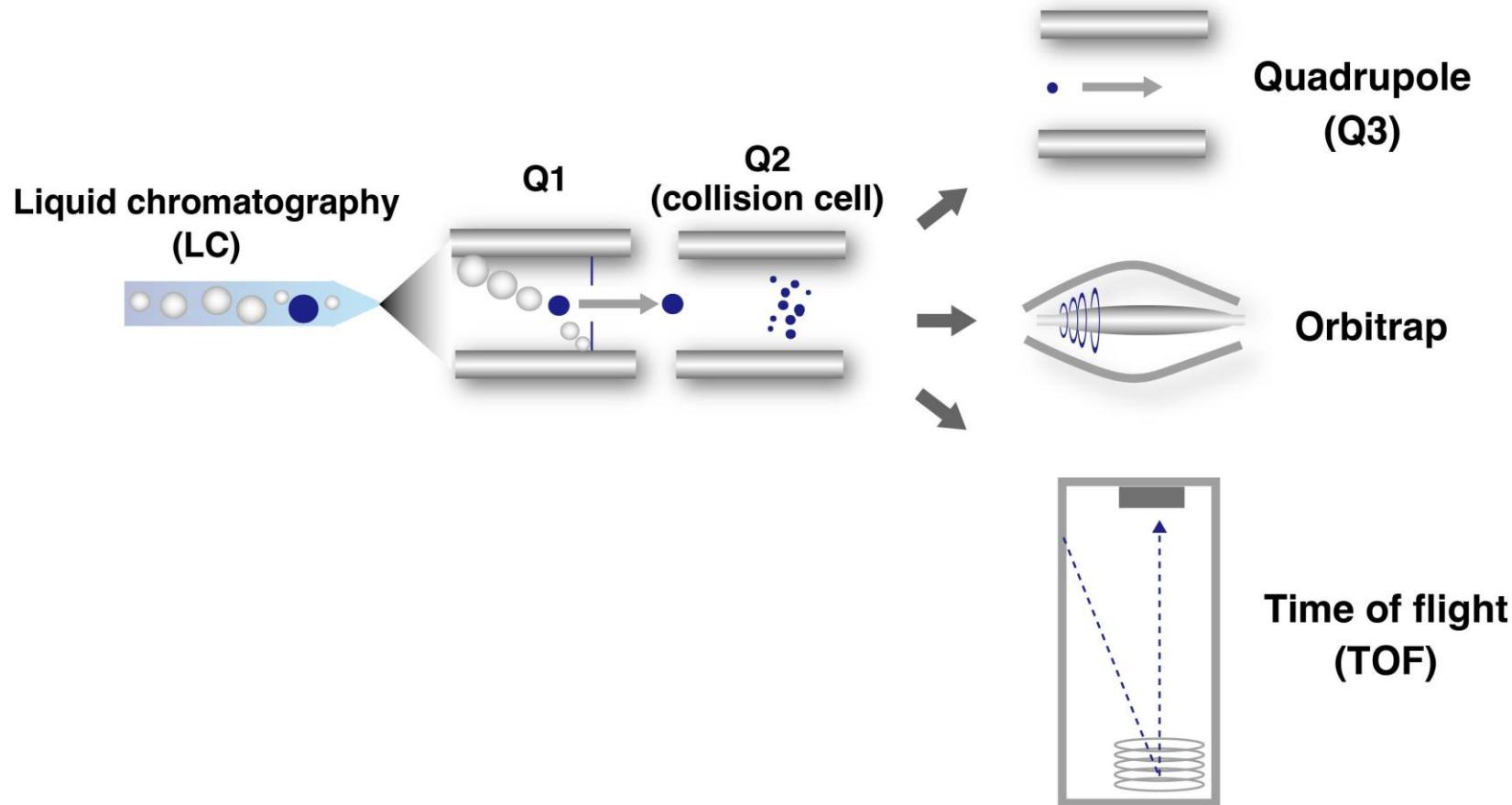
Electro Spray Ionization (ESI)



- 成分数が少ない場合に有効
- 1サンプルあたりの計測時間が短い
- 多検体の解析が可能

- HPLC等とのオンライン化が容易である。
- 装置の中に溶媒が入るため、装置の汚染を招きやすい。

タンデム質量分析計



複数のイオン分離系を組み合わせてMS/MSを取得可能

質量分析計の横軸:m/z (m over z)

ある質量Mを持った物質に水素等が付加してイオン化した場合=ポジティブイオン

	[M+H] ¹⁺	1価イオン
プロトン化イオンの 一般式: [M+nH] ⁿ⁺	[M+2H] ²⁺	2価イオン
	[M+3H] ³⁺	3価イオン

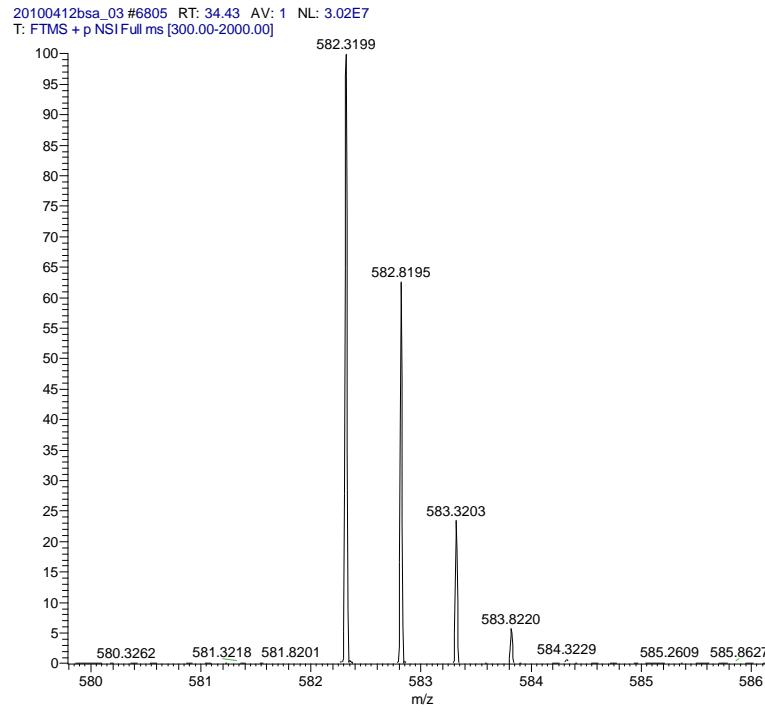
$$m/z = \frac{[M+nH]}{n}$$

¹Hの質量数=1.007825

例えば分子質量数500の分子に2つのプロトンが付加した場合
 $m/z = [500 + 2 * 1.007825] / 2 = 251.0078$

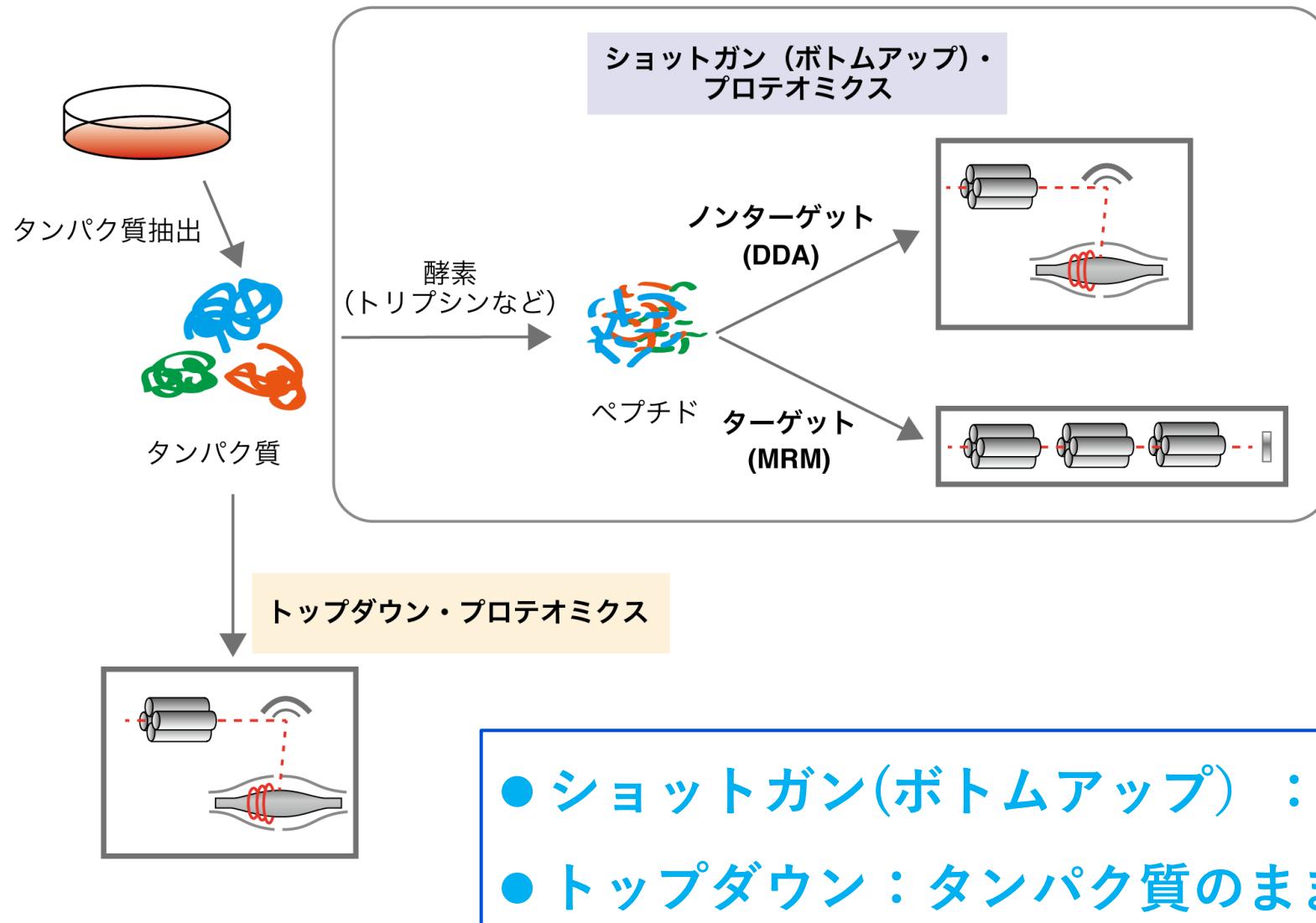
多価イオン

- ESI法では多くの場合、多価イオンを生じ傾向がある。
- プロテオミクスにおいて解析するペプチドはトリプシン消化によって生じているものが多く、N末端のアミノ基とC末端のリジンあるいはアルギニンの側鎖に存在するアミノ基の影響で**2価や3価イオン**を生じやすい。



- 天然の物質は同位体を一定の割合で含んでいる。
- そのため、元素の同位体分布に従って質量が1Daのずれを持つ同位体クラスターが観測される。
- この同位体クラスターの間隔は1価イオンなら 1.0 m/z であるが、2価イオンなら 0.5 m/z であり、3価イオンなら 0.33 m/z である。
- 従って、この同位体ピーク間の幅からイオンの価数可能。逆に言えば価数判定ができないと、実測される m/z から質量を算出できない。

質量分析計を基盤とするプロテオミクスの概要

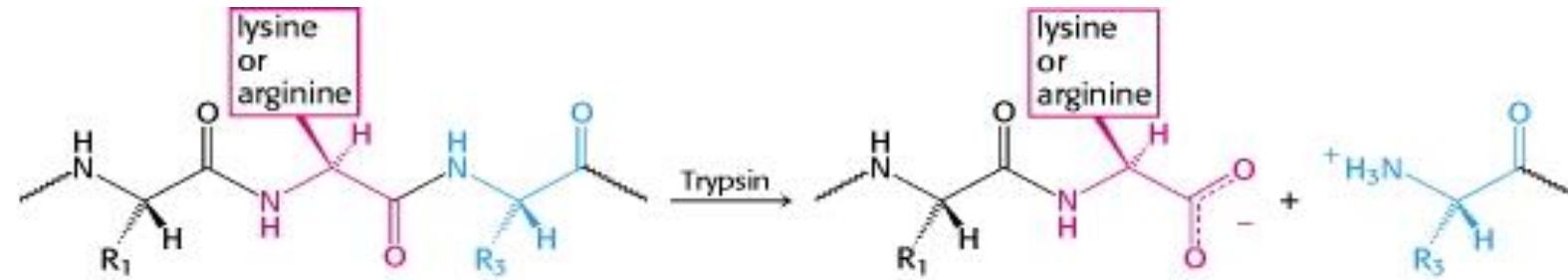


- ショットガン(ボトムアップ)：ペプチドにして解析
- トップダウン：タンパク質のまま解析

プロテオミクスにはボトムアップとトップダウンの二種類の戦略がある

ショットガンプロテオミクスにはタンパク質の断片化 (ペプチド化が必須)

Trypsin



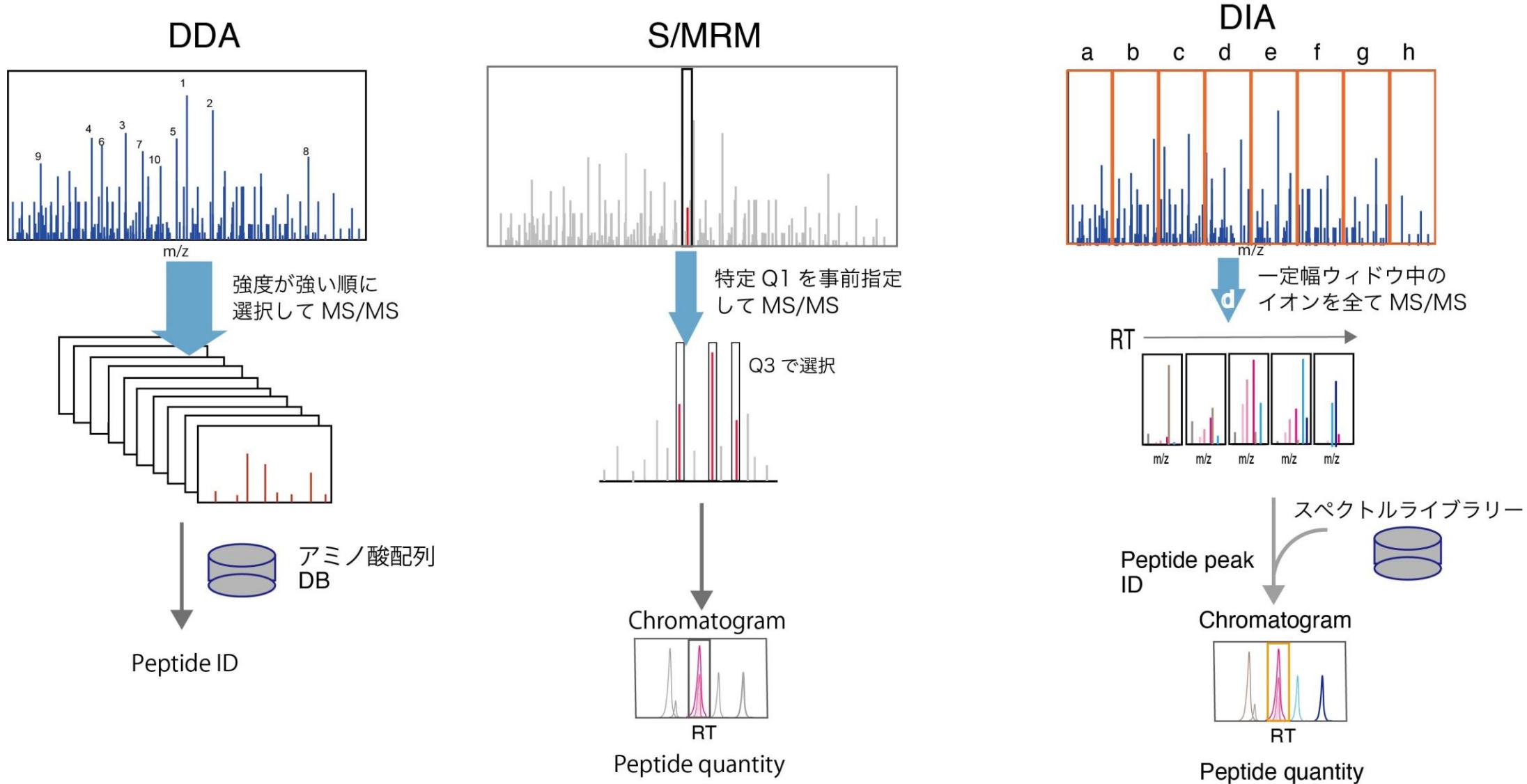
特定のアミノ酸で切断されることが重要！

----KXXXXXXXKXXXXXXXRXXXXXXX----

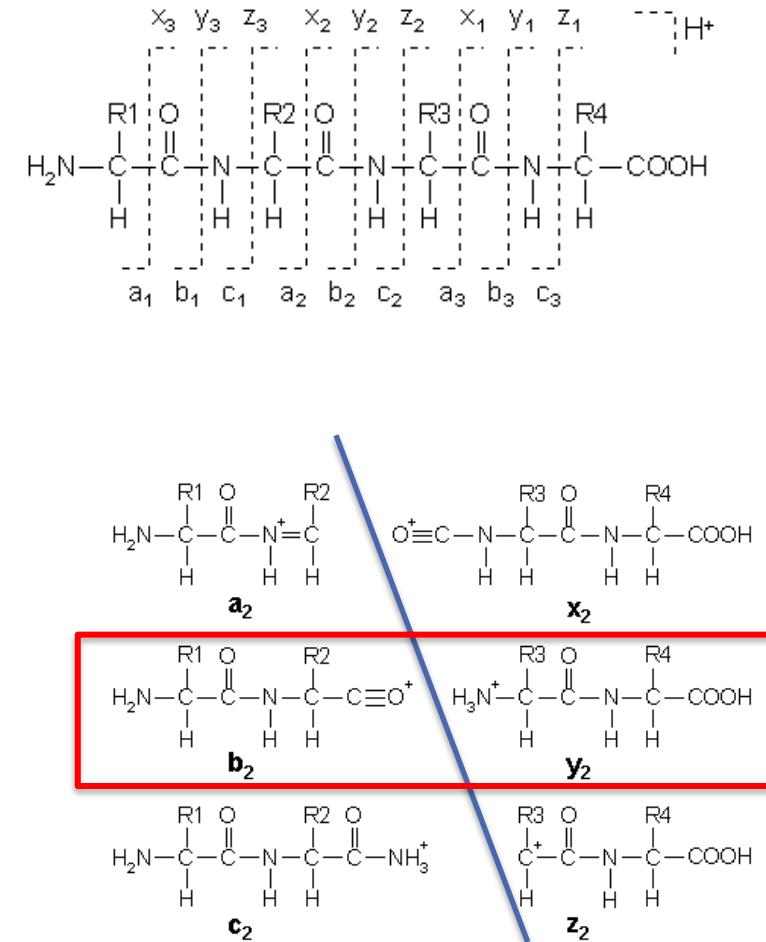
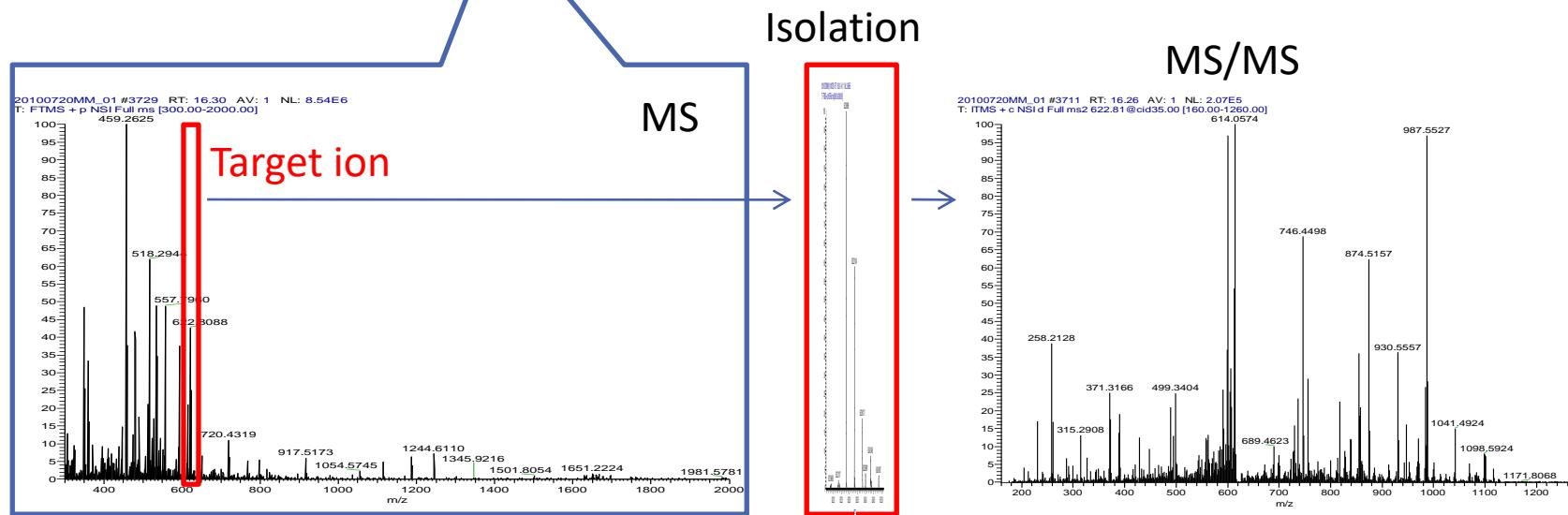
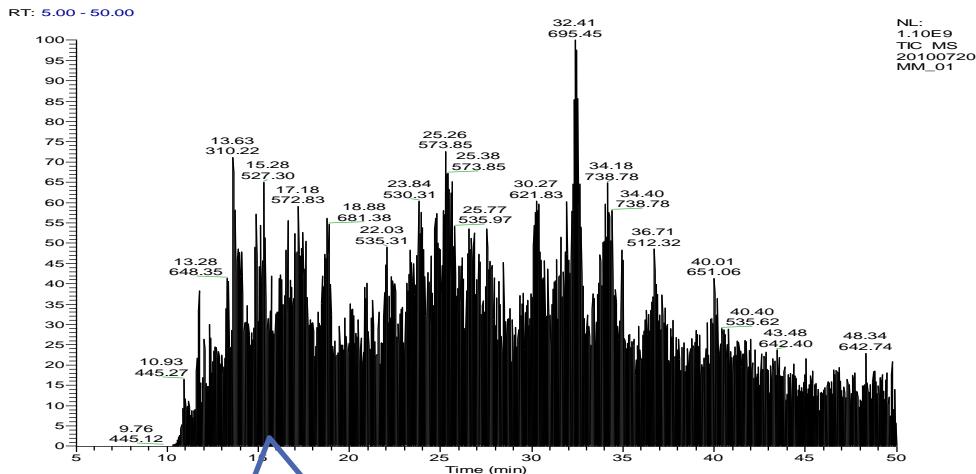
必然的に末端のアミノ酸配列が決められる

Trypsinの場合はC末端がKかR

LC-MS/MSにおける測定モードの多様化

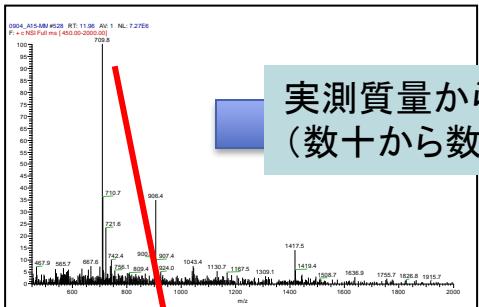


Data-dependent acquisition (DDA)モードによるLC-MS/MS測定



DDAによるタンパク質同定の原理

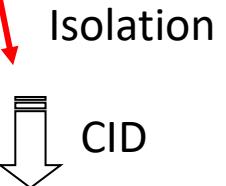
Precursor ion spectra
(MS)



実測質量から候補ペプチドの選定
(数十から数千の候補)

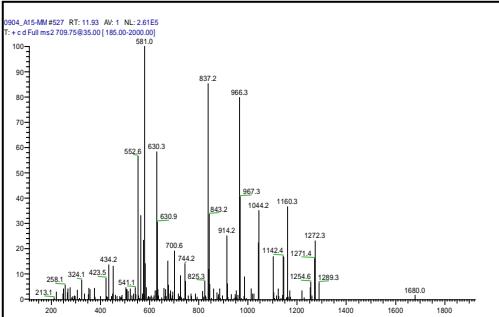
In silico (database)

Precursor ion
SAVGHEYVVAEVEK



CID

Database search
(MASCOT etc)



フラグメントの質量で照合

Identify

Fragment ions

b-ion

S **AVGHEYVVAEVEK**

SA **VGHEYVVAEVEK**

SAV **GHEYVVAEVEK**

⋮

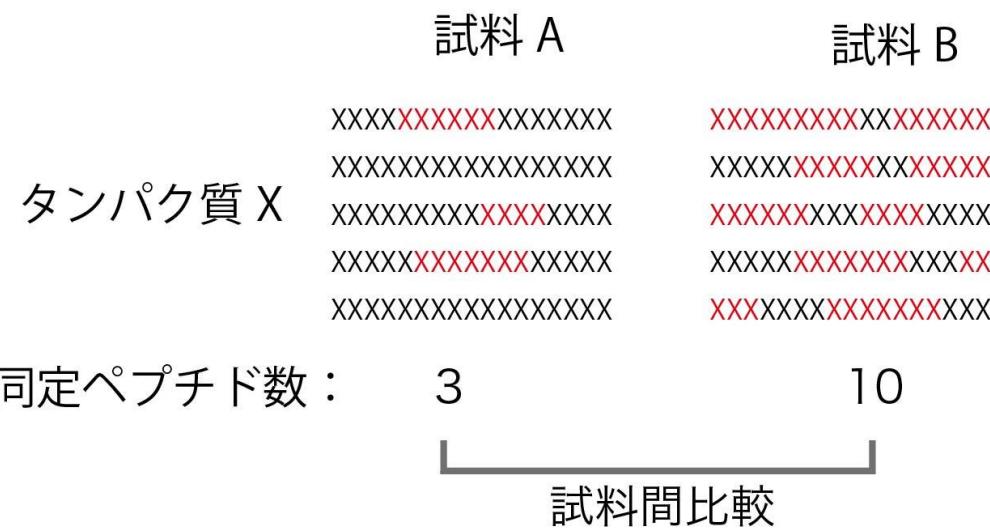
⋮

⋮

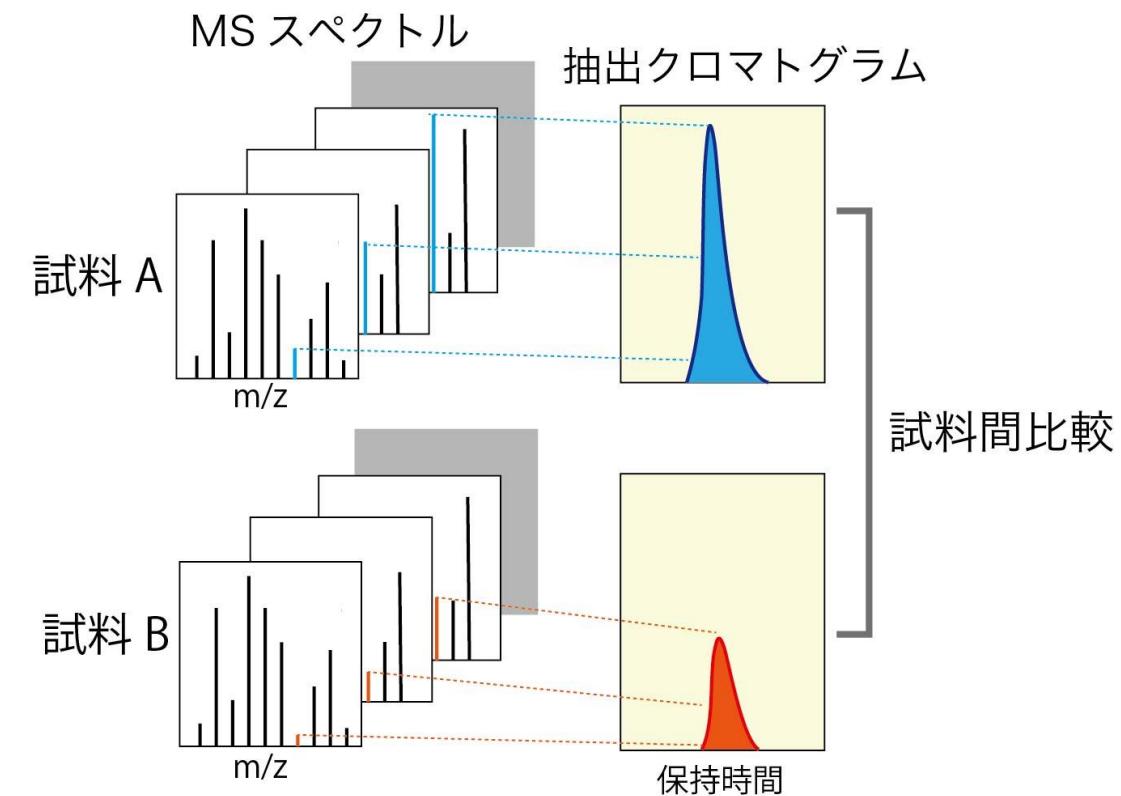
y-ion

ラベル化しない定量法

スペクトルカウント

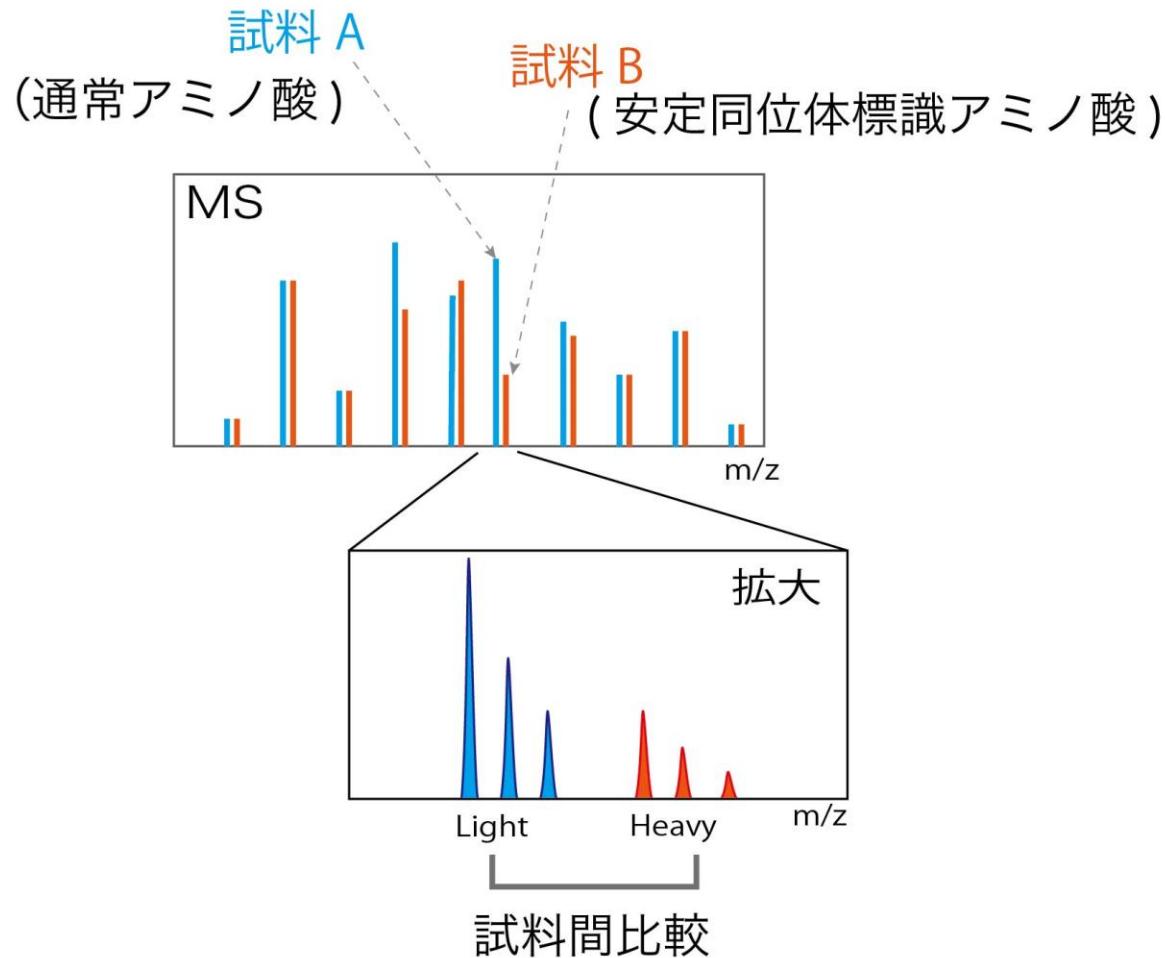


シグナル強度定量

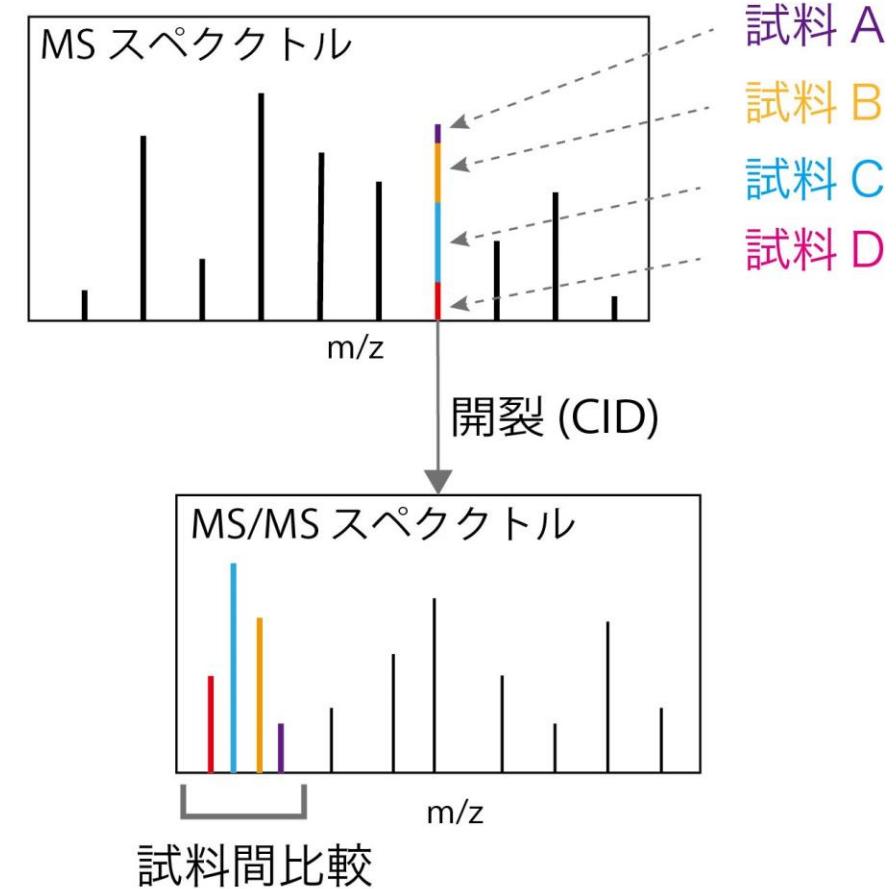


ラベル化による定量法

Isotopic ラベル定量
(SILAC や dimethyl 化など)



Isobaric ラベル定量
(iTRAQ や TMT など)



Public dataの活用：jPOST Repository



About jPOSTrepo

jPOSTrepo (Japan ProteOme STandard Repository) is a new data repository of sharing MS raw/processed data. It consists of a newly-developed, high-speed file upload process, flexible file management system and easy-to-use interfaces. Users can release their "raw/processed" data via this site with a unique identifier number for the paper publication. Users also can suspend (or "embargo") their data until their paper is published. The file transfer from users' computer to our repository server is very fast (roughly ten times faster than usual file transfer) and uses only web browsers – it does not require installing any additional software.

jPOST is a certificated member of ProteomeXchange Consortium and jPOSTrepo provides official ProteomeXchange Identifiers to projects stored in our repository.



Reference

Please cite the following article when using jPOSTrepo:

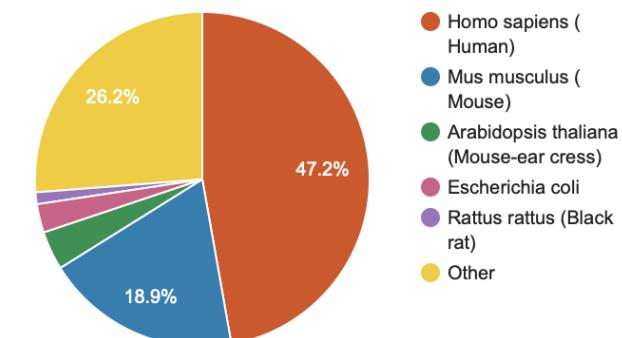
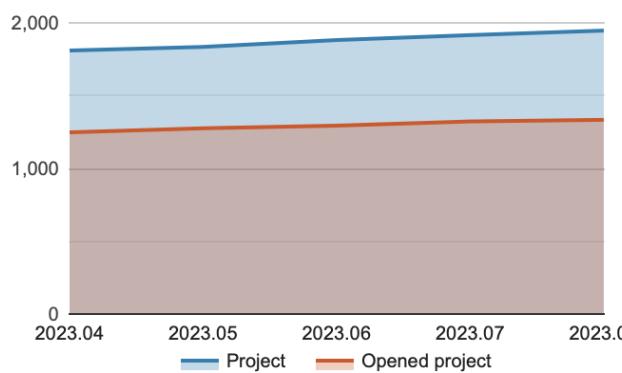
Okuda, S. et al. jPOSTrepo: an international standard data repository for proteomes. Nucl. Acids Res. 45 (D1): D1107-D1111 (2017). doi: 10.1093/nar/gkw1080 [pubmed]

Statistics

1947 projects are registered. **1334** are opened.

129672 files amount to **72.5 TB**.

268 species.



Free word

Ontology keyword

Search by free word

Project type

- All
- Mass spectrometry
- Gel electrophoresis
- Antibody

Search**Reset**1 - 20 / 1334 1 2 3 ... 67

jPOST ID	PXID	Project title	Description	Complete / Partial	Publication	Principal investigator	Announcement date	
JPST002070	PXD040569	lli. Turboid-EV: a proximity-labeling proteomic approach to dissect the uptake mechanism of extracellular vesicles	Extracellular vesicles (EVs), including exosomes, ...	Partial	Pre-publication	Yasushi Ishihama Kyoto university	2023-08-21	Detail page Quick view
JPST001816	PXD036054	lli. Absolute quantification of ASCC3	We measured the ASCC molecules in the complex of s ...	Partial	Pre-publication	Keiichi I Nakayama Kyushu University	2023-08-15	Detail page Quick view
JPST001967	PXD039097	lli. Interactome analysis for Kdm4B	In this project, to identify the proteins interact ...	Complete	Pre-publication	Yukinori Hirano The Hong Kong University of Science and Technology	2023-08-12	Detail page Quick view
JPST001814	PXD035997	lli. Brain phosphoproteomics in aging without AD	(Phospho)proteomic studies in the human aging brai ...	Partial	Pre-publication	isidro ferrer IDIBELL-Barcelona, Spain	2023-08-11	Detail page Quick view
JPST001811	PXD035906	lli. Serum peptidome profiles in metastatic and non-metastatic feline mammary carcinoma	Feline mammary carcinoma (FMC) is a common cancer ...	Partial	Pre-publication	Sittiruk Roytrakul National Center for Genetic Engineering and Biotechnology	2023-08-09	Detail page Quick view
JPST002279	PXD044435	lli. Insulin-administrated phosphorylation of the CNS proteins in Lymnaea	In the gastropod mollusc Lymnaea stagnalis, insuli ...	Partial	Pre-publication	Junko Nakai Department of Biology, Waseda University	2023-08-08	Detail page Quick view
JPST001808	PXD035830	lli. SILAC Bovine Aortic Endothelial Cells	This is a pilot SILAC study wherein bovine aortic ...	Partial	Pre-publication	Luciana Hannibal Medical Center, University of Freiburg	2023-08-06	Detail page Quick view
JPST001807	PXD035832	lli. Proteome analysis of muscle, liver, and blood samples associated with cancer cachexia	We analyzed samples from multiple mouse cancer cac ...	Partial	Pre-publication	Masahiro Aoki, Yasushi Kojima Aichi Cancer Center	2023-08-04	Detail page Quick view
JPST002183	PXD042807	lli. Proteome analysis of liver samples associated with cancer cachexia_sex	We analyzed samples from multiple mouse cancer cac ...	Partial	Pre-publication	Masahiro Aoki, Yasushi Kojima Aichi Cancer Center	2023-08-04	Detail page Quick view

測定方法に応じた検体-raw file対応

Single shot/label-free:

1 検体/1 raw file

Fractionation/label-free:

1 raw fileには単一検体のみ

ただし、複数raw fileが発生 (fractionationの数だけ)

Fractionation/Multiplex (labeled samples) :

1 raw fileに複数検体が含まれる

複数raw fileが存在

残念ながらpublic repositoryのメタデータはあまり信用できない(適当な登録が多い)。
論文からメタデータを引っ張り出した方が正しい。
→ データジャーナルの活用(Journal of Proteome Data and Methods)

プロジェクト検索とメタデータの確認

Data list

Free word Ontology keyword

Project type
 All Mass spectrometry Gel electrophoresis Antibody

Search **Reset**

Search result 016259 1 / 1 [1]

JPOST ID	PXID	Project title	Description	Complete / Partial	Publication	Principal investigator	Announcement date	Detail page	Quick view
JPST000697	PXD016259	Label-free quantification of proteins by LC-MS/MS analysis	The P1 and P2 fractions were subjected to label-fr ...	Partial	Pre-publication	Keiichi I Nakayama Kyushu university	2020-11-13	Detail page	Quick view

JPST000697 summary

Species Homo sapiens (Human) [9606]
Tissue lung [BTO:0000763]
Cell type fibroblast [CL:0000057]
Subcellular Nucleus
Enzyme Trypsin [MS:1001251]
Fixed modification Carbamidomethyl (C) [UNIMOD:4#C]
Variable modification Oxidation (M) [UNIMOD:35#M]
Taxonomy Homo sapiens (Human) [9606]
Instrument Q Exactive [MS:1001911]
Instrument mode DDA-high res.
Purpose Relative quantification
Quantification platform Precursor ion label free
File format other
Software MaxQuant

JPST000697 Mass spectrometry

Project title Label-free quantification of proteins by LC-MS/MS analysis
Keywords label-free, cytosol, nuclear
Project description The P1 and P2 fractions were subjected to label-free proteomics analysis.
PubMed ID(s)
Created date 2019-11-06
Announcement date 2020-11-13
PXID PXD016259
Note
Principal investigator Keiichi I Nakayama
Kyushu university
Submitter Yuki Kito
Kyushu university
XML file JPST000697.0.xml
Revision 0
Submission type Partial submission
Project list

File Name Experimental presets (Sample / Fractionation / Enzyme/Mod. / MS mode) File Size Detail

20170929_TIG3WT_P1_17G05.raw Raw S TIG-3 F P1 fraction [1] E In solution digestion with trypsin M DDA analysis using QE 847.4 MB [Detail](#)

▼ Experimental presets

Sample	Fractionation	Enzyme/Mod.	MS mode
Title: TIG-3 Species: Homo sapiens (Human) Tissue: lung Cell type: fibroblast Disease: Note: TIG-3 infected with retroviruses encoding human telomerase reverse transcriptase (hTERT), the early region (ER) proteins of SV40 and c-Myc.	Title: P1 fraction Subcellular: Nucleus Replicate:1/3 Protein: Peptide: Note:	Title: In solution digestion with trypsin Enzyme: Trypsin Fixed modification: Carbamidomethyl (C) Variable modification: Oxidation (M) Taxonomy: Homo sapiens (Human) Note:	Title: DDA analysis using QE Instrument: Q Exactive Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Precursor ion label free Plex: Label: Note:

Isobaric tag定量のメタデータ

JPST000144 summary

Species	Mus musculus (Mouse) [10090]
Tissue	T-lymphocyte [BTO:0000782]
Cell type	T cell [CL:0000084]
Subcellular	Whole
Peptide	Fe-IMAC
Enzyme	Trypsin [MS:1001251]
Fixed modification	Carbamidomethyl (C) [UNIMOD:4#C], iTRAQ4plex (K) [UNIMOD:214#K], iTRAQ4plex (N-term) [UNIMOD:214#N-term]
Variable modification	Phospho (S) [UNIMOD:21#S], Phospho (T) [UNIMOD:21#T], Phospho (Y) [UNIMOD:21#Y], Gln->pyro-Glu (Q) [UNIMOD:28#Q], Glu->pyro-Glu (E) [UNIMOD:27#E], iTRAQ4plex (Y) [UNIMOD:214#Y]
Taxonomy	Mus musculus (Mouse) [10090]
Instrument	QSTAR [MS:1000190]
Instrument mode	DDA-high res.
Purpose	Relative quantification
Quantification platform	Product ion (reporter quantification)
Label	114, 115, 116, 117
File format	, MGF, mzIdentML
Software	, Mascot

Files

File Name	Experimental presets (Sample / Fractionation / Enzyme/Mod. / MS mode)	File Size	Download all (1.57 GB)	
			Detail	
20131026_CD3_thymocyte.wiff	<input checked="" type="checkbox"/> Mouse thymocyte antibodies to CD3e stimulation 0 min <input checked="" type="checkbox"/> Phosphopeptide Fe-IMAC [1,1] <input checked="" type="checkbox"/> trypsin Fe-IMAC iTRAQ <input checked="" type="checkbox"/> QSTAR elite iTRAQ [114]			
	<input checked="" type="checkbox"/> Mouse thymocyte antibodies to CD3e stimulation 2.5 min <input checked="" type="checkbox"/> Phosphopeptide Fe-IMAC [1,1] <input checked="" type="checkbox"/> trypsin Fe-IMAC iTRAQ <input checked="" type="checkbox"/> QSTAR elite iTRAQ [115]			
	<input checked="" type="checkbox"/> Mouse thymocyte antibodies to CD3e stimulation 5 min <input checked="" type="checkbox"/> Phosphopeptide Fe-IMAC [1,1] <input checked="" type="checkbox"/> trypsin Fe-IMAC iTRAQ <input checked="" type="checkbox"/> QSTAR elite iTRAQ [116]	7.3 MB	Detail	
	<input checked="" type="checkbox"/> Mouse thymocyte antibodies to CD3e stimulation 10 min <input checked="" type="checkbox"/> Phosphopeptide Fe-IMAC [1,1] <input checked="" type="checkbox"/> trypsin Fe-IMAC iTRAQ <input checked="" type="checkbox"/> QSTAR elite iTRAQ [117]			

114	115	116	117
Sample	Fractionation	Enzyme/Mod.	MS mode
Title: Mouse thymocyte antibodies to CD3e stimulation 0 min Species: Mus musculus (Mouse) Tissue: T-lymphocyte Cell type: T cell Disease: Note:	Title: Phosphopeptide Fe-IMAC Subcellular: Whole Replicate:1/1 Protein: Peptide: Fe-IMAC Fraction:1 Replicate:1/1 Note:	Title: trypsin Fe-IMAC iTRAQ Enzyme: Trypsin Fixed modification: Carbamidomethyl (C), iTRAQ4plex (K), iTRAQ4plex (N-term) Variable modification: Phospho (S), Phospho (T), Phospho (Y), Gln->pyro-Glu (Q), Glu->pyro-Glu (E), iTRAQ4plex (Y) Taxonomy: Mus musculus (Mouse) Note:	Title: QSTAR elite iTRAQ Instrument: QSTAR Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Product ion (reporter quantification) Plex: 4 Label: 114, 115, 116, 117 Note:

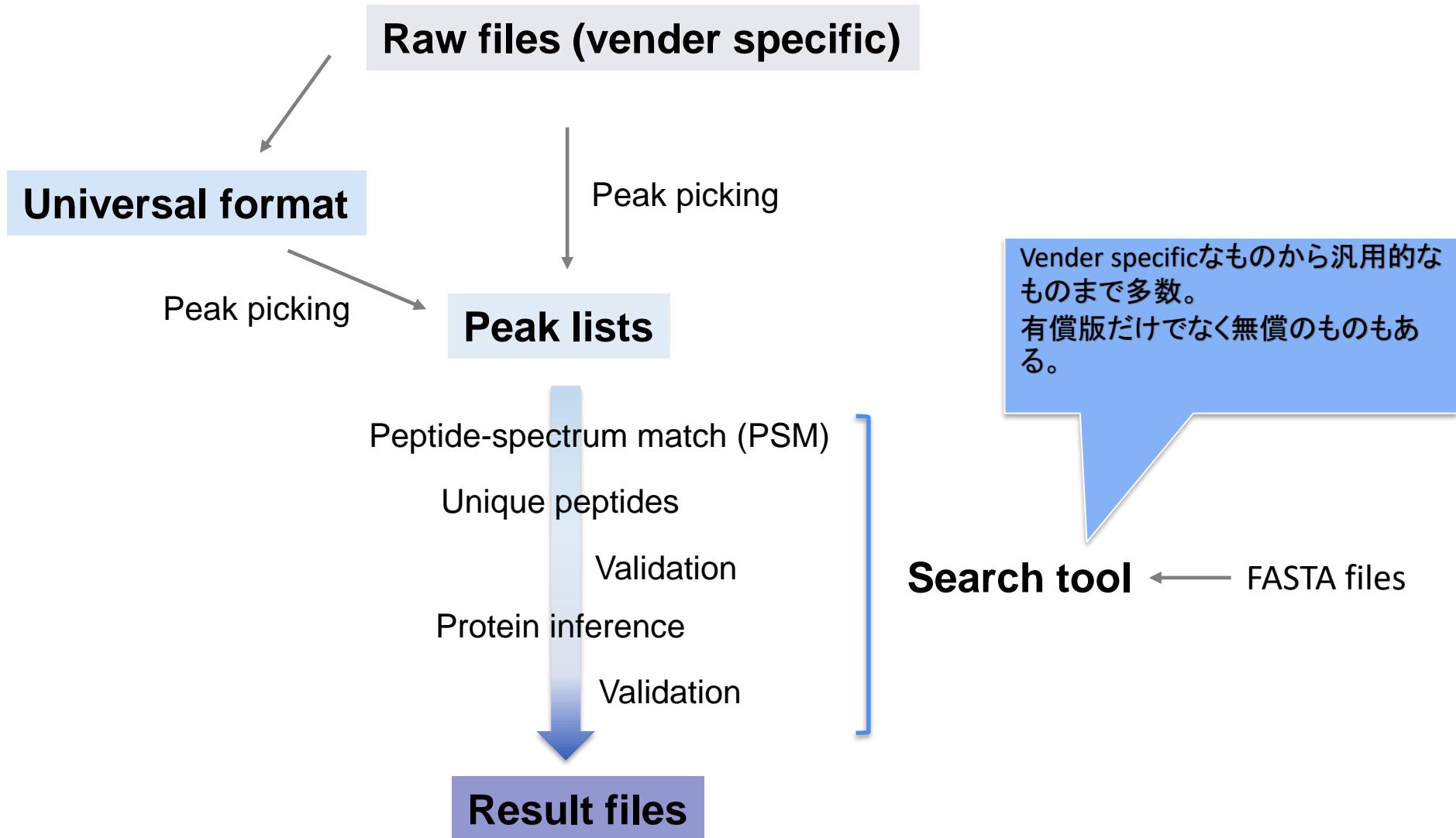
114	115	116	117
Sample	Fractionation	Enzyme/Mod.	MS mode
Title: Mouse thymocyte antibodies to CD3e stimulation 2.5 min Species: Mus musculus (Mouse) Tissue: T-lymphocyte Cell type: T cell Disease: Note:	Title: Phosphopeptide Fe-IMAC Subcellular: Whole Replicate:1/1 Protein: Peptide: Fe-IMAC Fraction:1 Replicate:1/1 Note:	Title: trypsin Fe-IMAC iTRAQ Enzyme: Trypsin Fixed modification: Carbamidomethyl (C), iTRAQ4plex (K), iTRAQ4plex (N-term) Variable modification: Phospho (S), Phospho (T), Phospho (Y), Gln->pyro-Glu (Q), Glu->pyro-Glu (E), iTRAQ4plex (Y) Taxonomy: Mus musculus (Mouse) Note:	Title: QSTAR elite iTRAQ Instrument: QSTAR Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Product ion (reporter quantification) Plex: 4 Label: 114, 115, 116, 117 Note:

114	115	116	117
Sample	Fractionation	Enzyme/Mod.	MS mode
Title: Mouse thymocyte antibodies to CD3e stimulation 5 min Species: Mus musculus (Mouse) Tissue: T-lymphocyte Cell type: T cell Disease: Note:	Title: Phosphopeptide Fe-IMAC Subcellular: Whole Replicate:1/1 Protein: Peptide: Fe-IMAC Fraction:1 Replicate:1/1 Note:	Title: trypsin Fe-IMAC iTRAQ Enzyme: Trypsin Fixed modification: Carbamidomethyl (C), iTRAQ4plex (K), iTRAQ4plex (N-term) Variable modification: Phospho (S), Phospho (T), Phospho (Y), Gln->pyro-Glu (Q), Glu->pyro-Glu (E), iTRAQ4plex (Y) Taxonomy: Mus musculus (Mouse) Note:	Title: QSTAR elite iTRAQ Instrument: QSTAR Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Product ion (reporter quantification) Plex: 4 Label: 114, 115, 116, 117 Note:

114	115	116	117
Sample	Fractionation	Enzyme/Mod.	MS mode
Title: Mouse thymocyte antibodies to CD3e stimulation 0 min Species: Mus musculus (Mouse) Tissue: T-lymphocyte Cell type: T cell Disease: Note:	Title: Phosphopeptide Fe-IMAC Subcellular: Whole Replicate:1/1 Protein: Peptide: Fe-IMAC Fraction:1 Replicate:1/1 Note:	Title: trypsin Fe-IMAC iTRAQ Enzyme: Trypsin Fixed modification: Carbamidomethyl (C), iTRAQ4plex (K), iTRAQ4plex (N-term) Variable modification: Phospho (S), Phospho (T), Phospho (Y), Gln->pyro-Glu (Q), Glu->pyro-Glu (E), iTRAQ4plex (Y) Taxonomy: Mus musculus (Mouse) Note:	Title: QSTAR elite iTRAQ Instrument: QSTAR Instrument mode: DDA-high res. Purpose: Relative quantification Quantification platform: Product ion (reporter quantification) Plex: 4 Label: 114, 115, 116, 117 Note:

DDAのデータ解析

DDA data の解析ワークフロー



広く使われているMS/MSデータ解析ソフト

Mascot

- 最もポピュラーな検索エンジン(有償)
- 操作がシンプルで使いやすい
- スペクトル帰属図などが見やすい
- 結果の統合や定量は別ソフトが必要

MaxQuant

- フリーソフト
- タンパク質同定(Andromeda)から定量まで実施可能
- スペクトルの可視化などが重い

Vender specificなソフトの利用も一般的
(ProteomeDiscover, ProteinPilot など)

MSFagger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics

Andy T Kong^{1,2}, Felipe V Leprevost² , Dmitry M Avtonomov², Dattatreya Mellacheruvu²  & Alexey I Nesvizhskii^{1,2} 

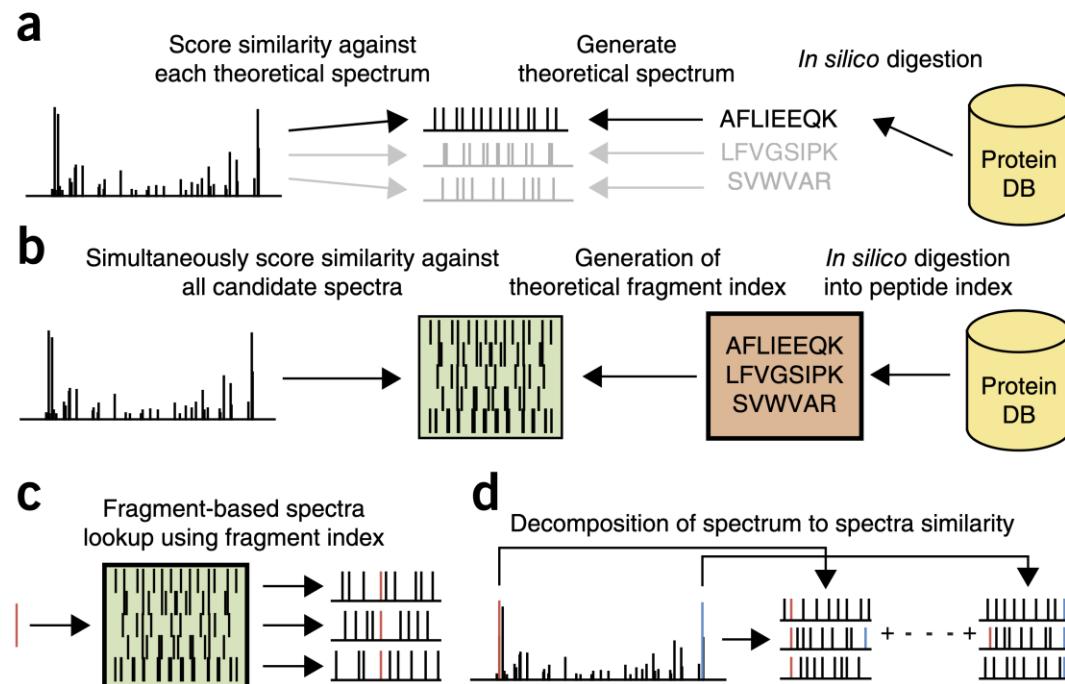
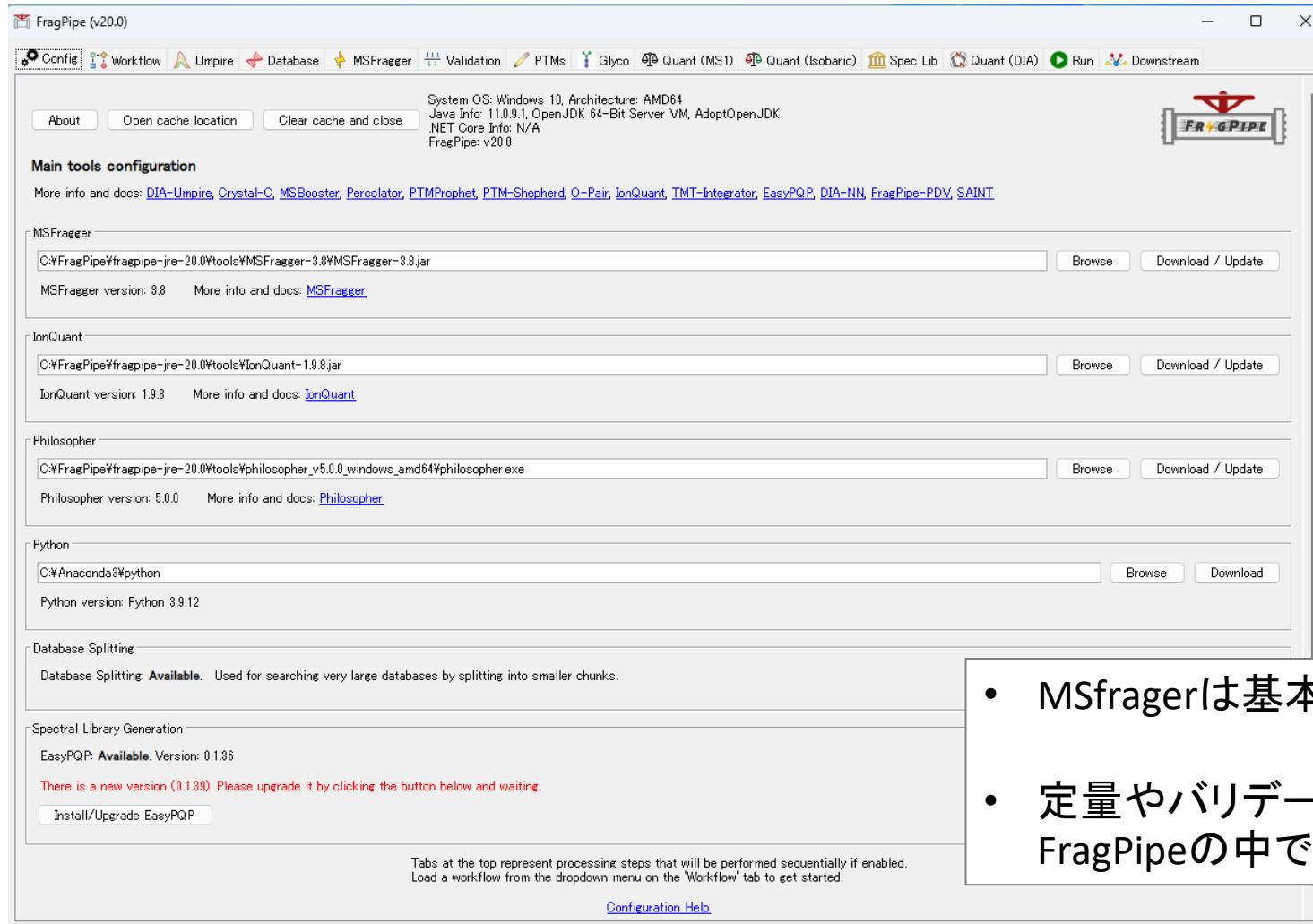


Figure 1 | Database-search strategies and the MSFagger algorithm.

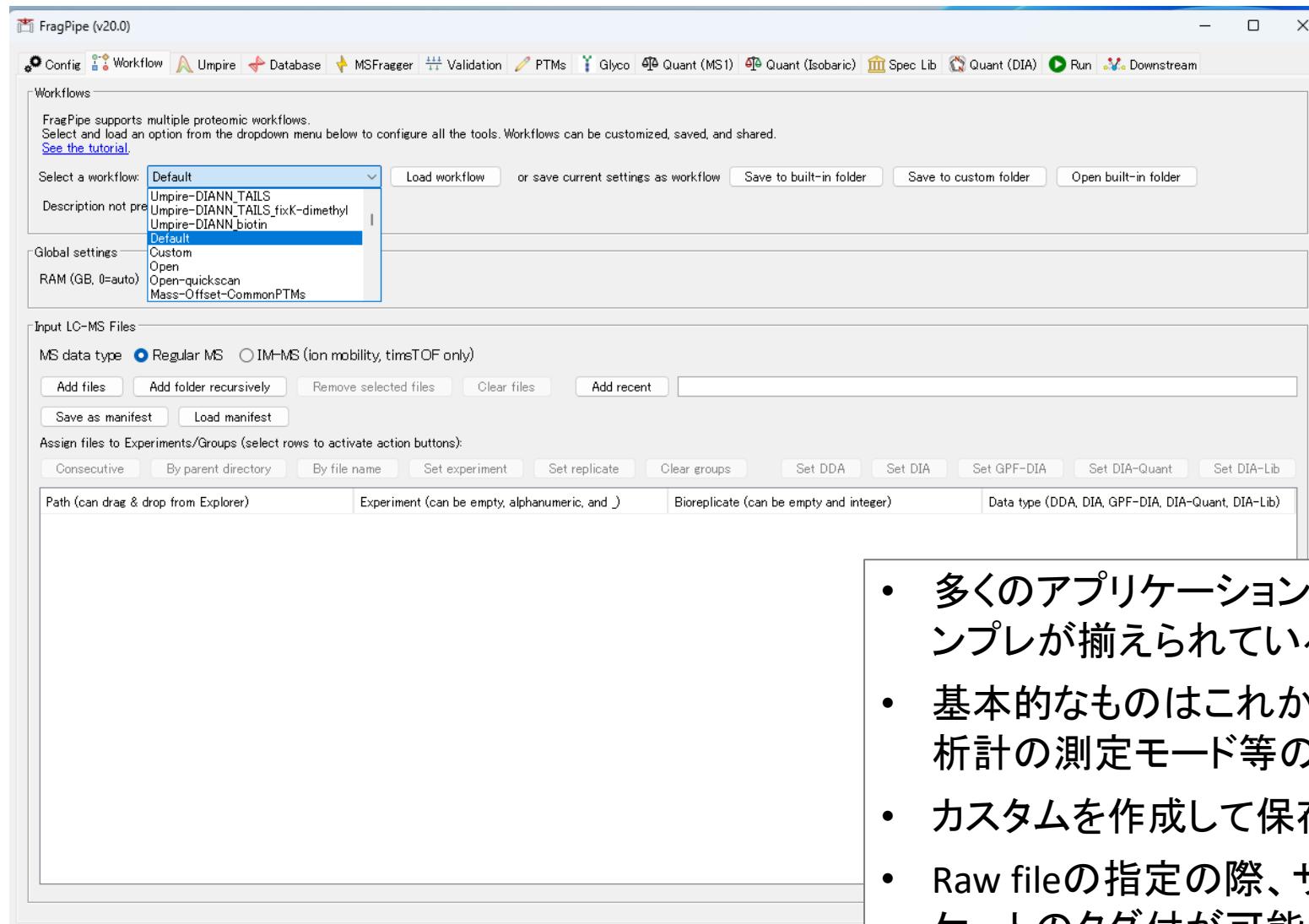
- フリーソフトウェア
- 検索が早い
- Open search が現実的に可能 → 網羅的な翻訳後修飾の解析が可能
- 他のツールとFragPipe のプラットフォーム内で連携

FragPipeの設定



- MSfaggerは基本FragPipeの中で動く。
- 定量やバリデーションなどで働くツールを FragPipeの中で動かす。

Workflowの呼び出しとrawデータの指定



- 多くのアプリケーションに対応したワークフローのテンプレが揃えられている。
- 基本的なものはこれから選べばOK(もちろん質量分析計の測定モード等の条件依存)。
- カスタムを作成して保存も可能。
- Raw fileの指定の際、サンプルグループやレプリケートのタグ付が可能。

検索条件の設定

FragPipe (v20.0)

Config Workflow Umpire Database MSFragger Validation PTMs Glyco Quant (MS1) Quant (Isobaric) Spec Lib Quant (DIA) Run Downstream

Run MSFragger

Save Config Load : Custom MSFragger parameter file from di...

MSFRAGGER

Common Options (Advanced Options are at the end of the page)

Peak Matching

Precursor mass tolerance PP... -20 - 20 Fragment mass tolerance PP... 20

Calibration and Optimization Mass calibration, parameter optimizati... Isotope error 0/1/2

Protein Digestion

Cleavage ENZYMATIC Clip N-term M

Enzyme name 1 trypsin Load rules trypsin Cuts 1 KR No cuts 1 P Missed cleavages 1 2 Sense 1 ... Enzyme name 2 null Load rules null Cuts 2 No cuts 2 Missed cleavages 2 2 Sense 2 ... Peptide length 7 - 50 Peptide mass range 500 - 5,000 Split database 1

Modifications

Variable modifications

Max variable mods on a peptide 3 Max combinations 5,000 Use all mods in first search

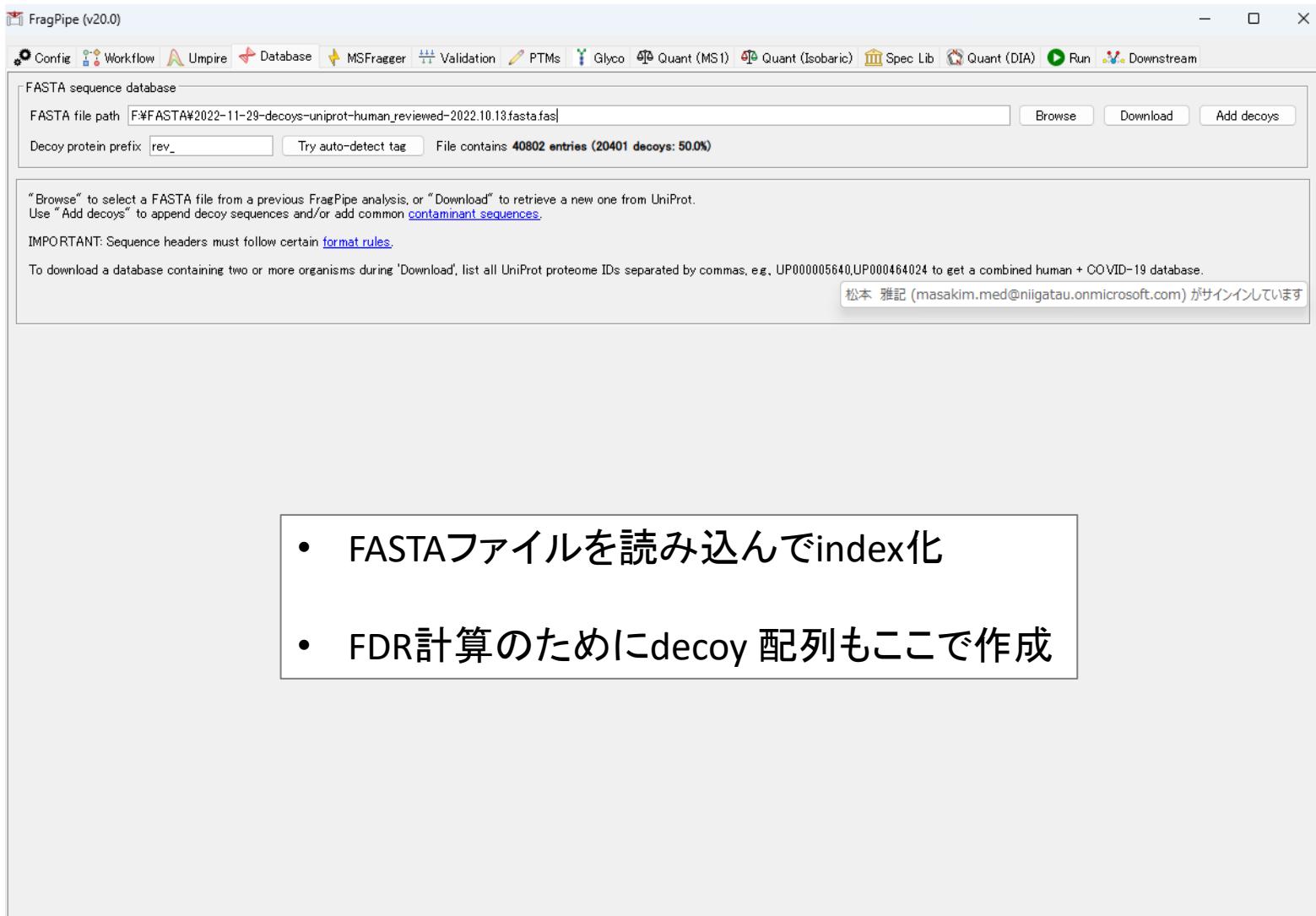
Enabled	Site (editable)	Mass Delta (editable)	Max occurrences ...
<input checked="" type="checkbox"/>	M	15.9949	3
<input checked="" type="checkbox"/>	[42.0106	1
<input type="checkbox"/>	STY	79.96633	3
<input checked="" type="checkbox"/>	nQnG	-17.0265	1
<input checked="" type="checkbox"/>	nE	-18.0106	1
<input type="checkbox"/>	site_06	0.0	1
<input type="checkbox"/>	site_07	0.0	1
<input type="checkbox"/>	site_08	0.0	1

Fixed modifications

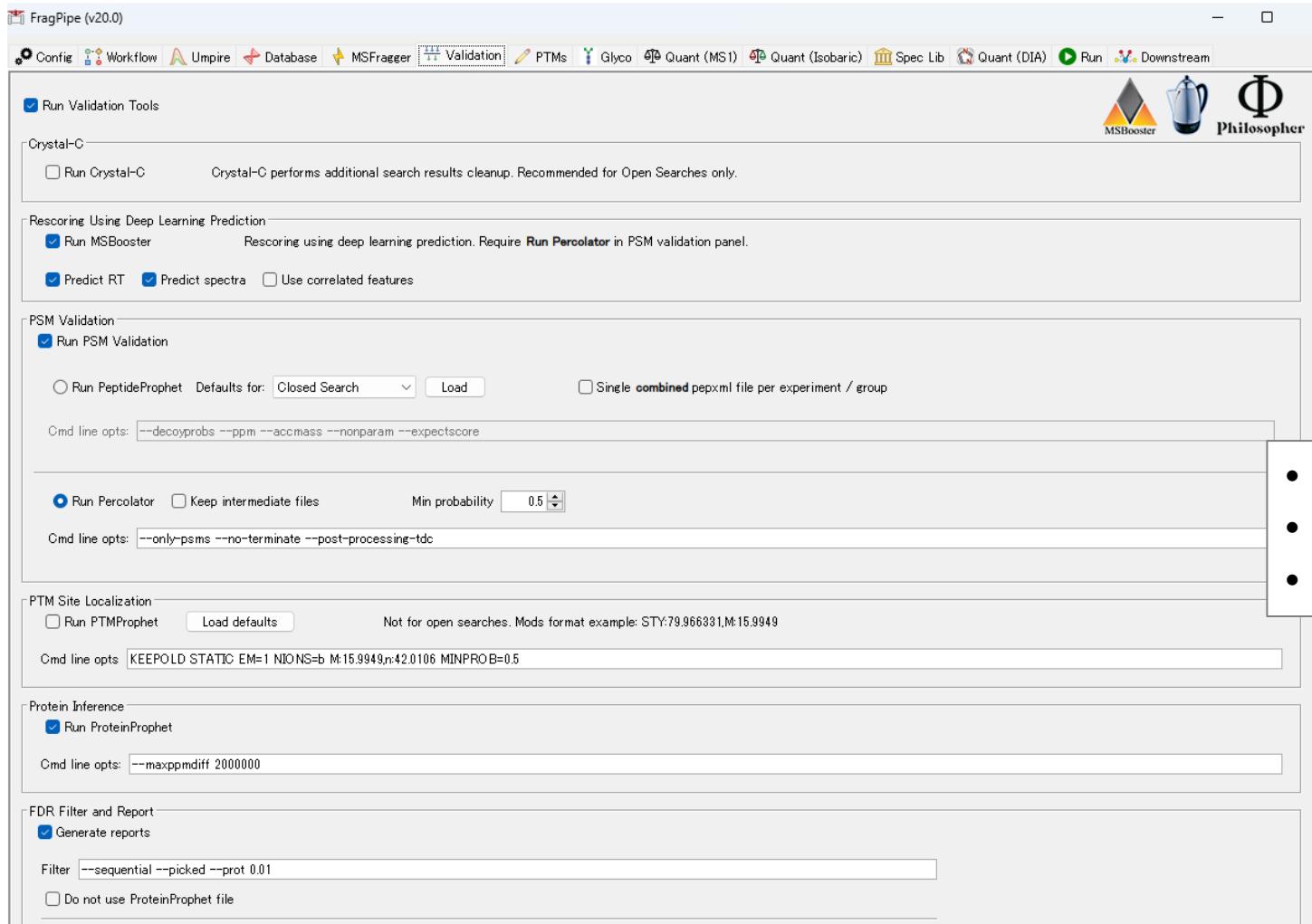
Enabled	Site	Mass Delta (editable)
<input checked="" type="checkbox"/>	C-Term Peptide	0.0
<input checked="" type="checkbox"/>	N-Term Peptide	0.0
<input checked="" type="checkbox"/>	C-Term Protein	0.0
<input checked="" type="checkbox"/>	N-Term Protein	0.0
<input checked="" type="checkbox"/>	C-terminal	0.0

- 酵素の選定
- 翻訳後修飾などを入力
- 装置の質量精度に応じてトレランスを設定

FASTAの準備



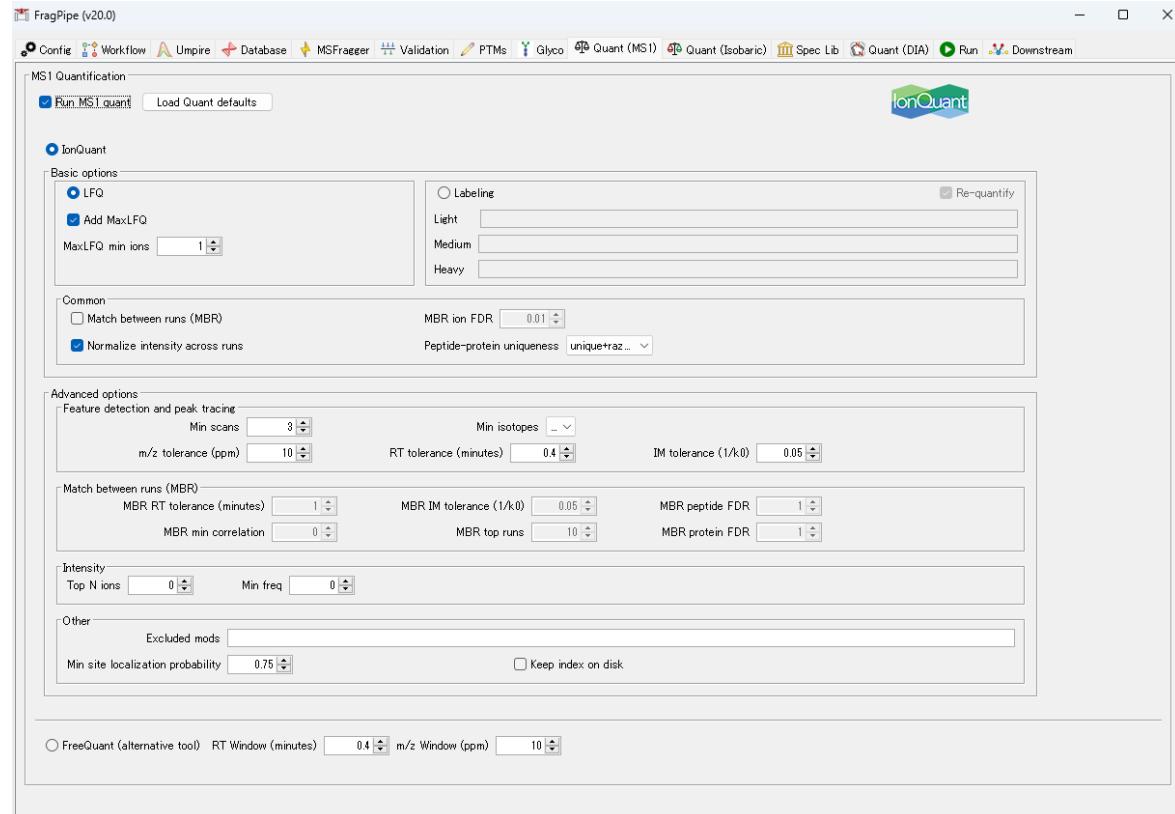
Validation



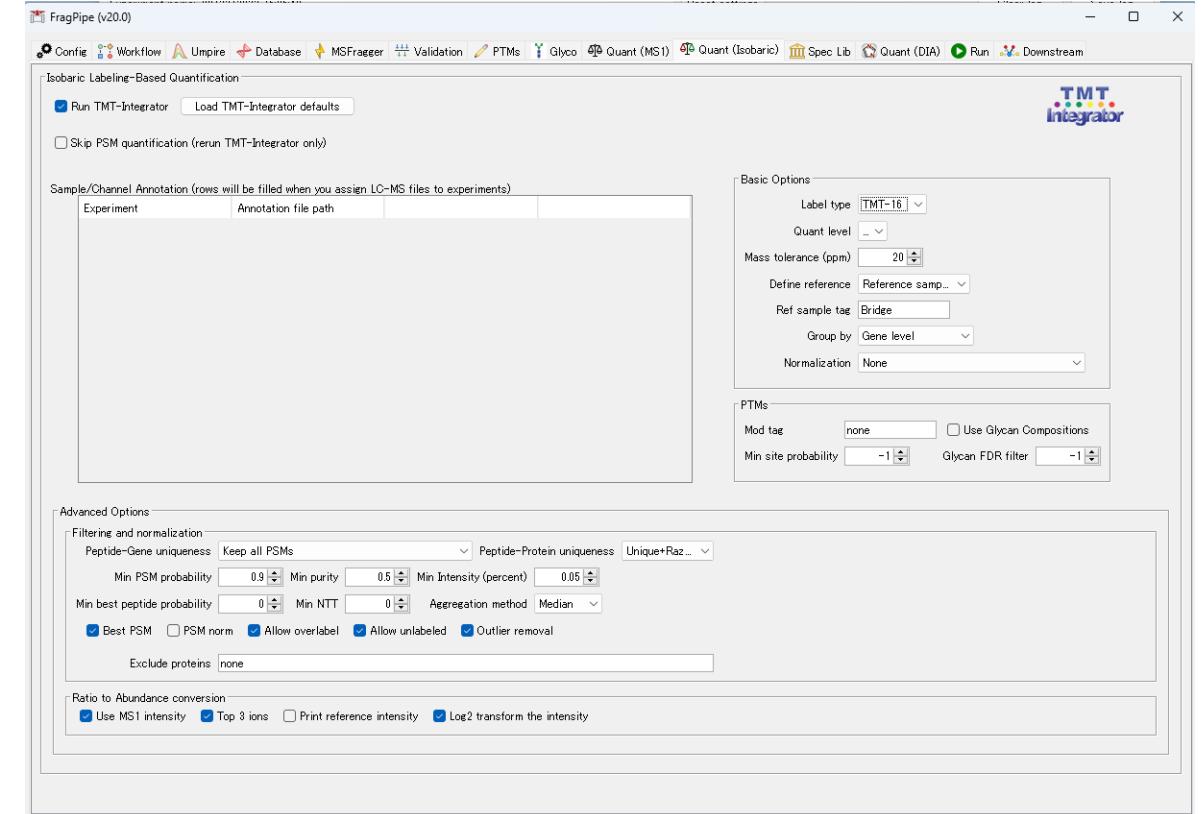
- Validation toolの選択
- FDRフィルターの設定
- PTMのvalidationも可能

定量モード

Label-free or isotopic-label 定量

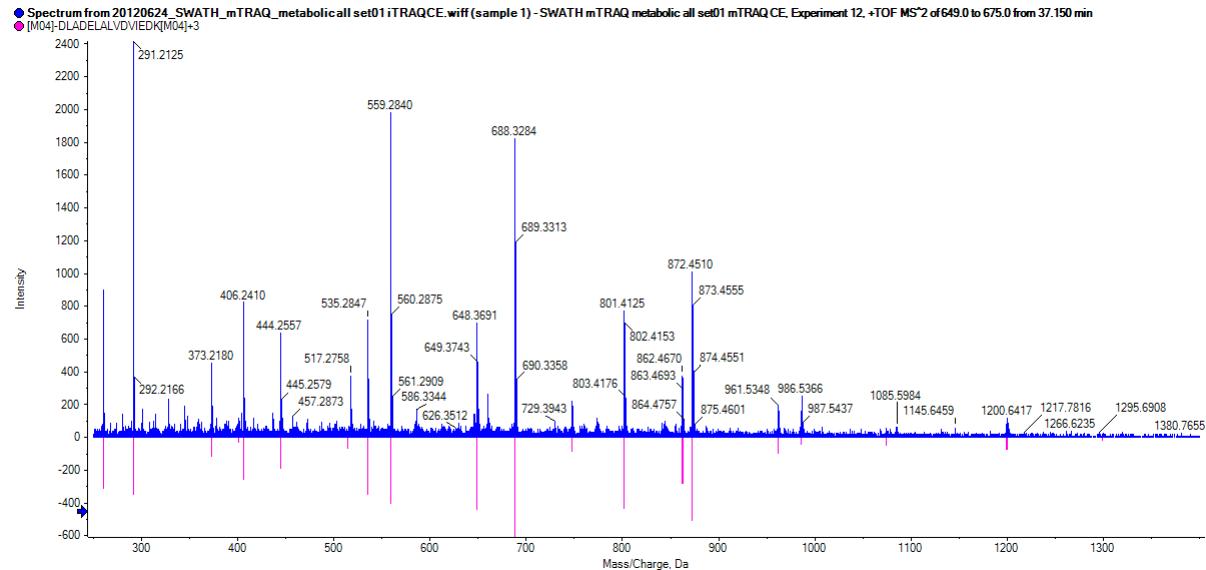


Isobaric-label 定量 (TMT/iTRAQ)

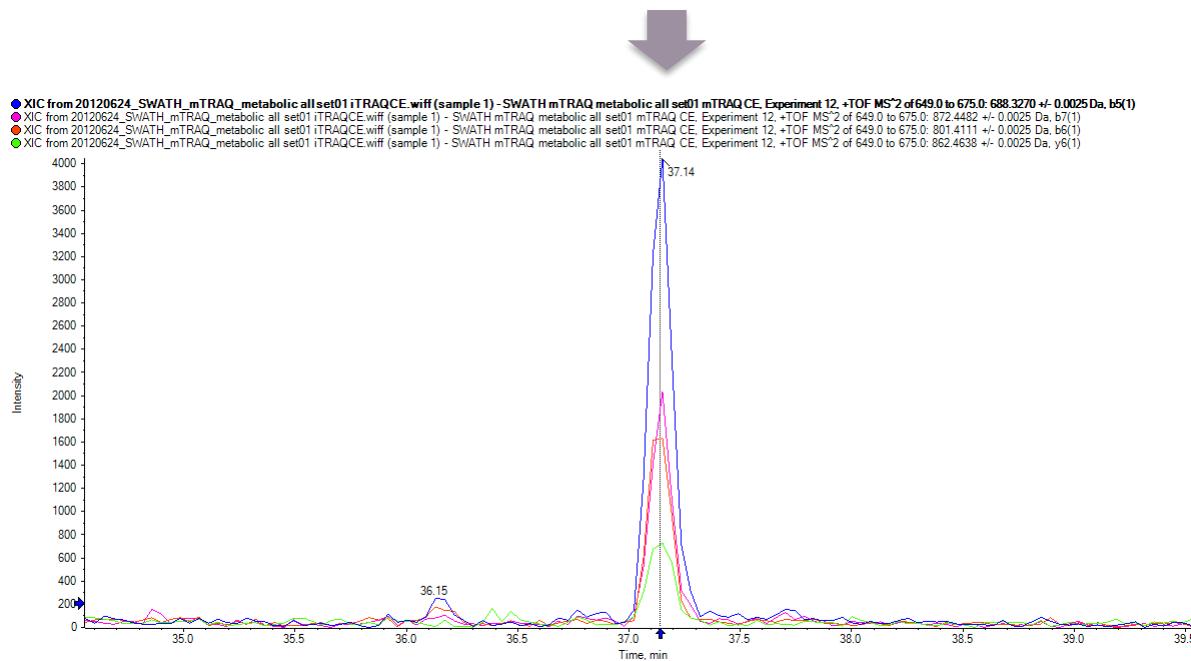


DIAのデータ解析

DIAデータ解析の基本



DIA spectra
(混合MS/MS)

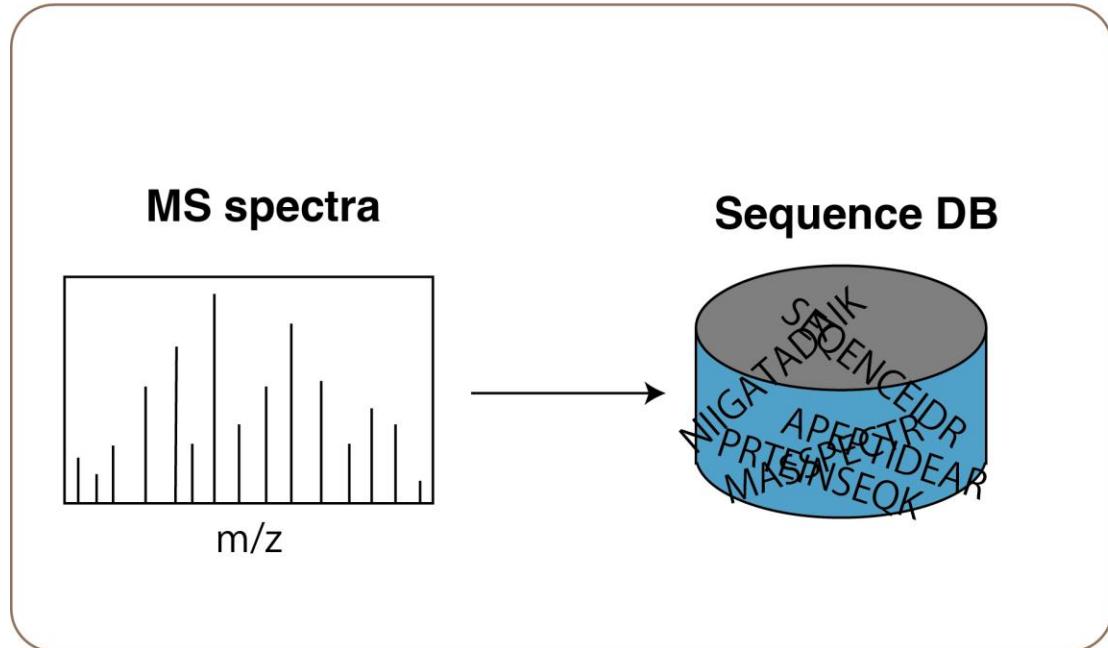


Spectra library

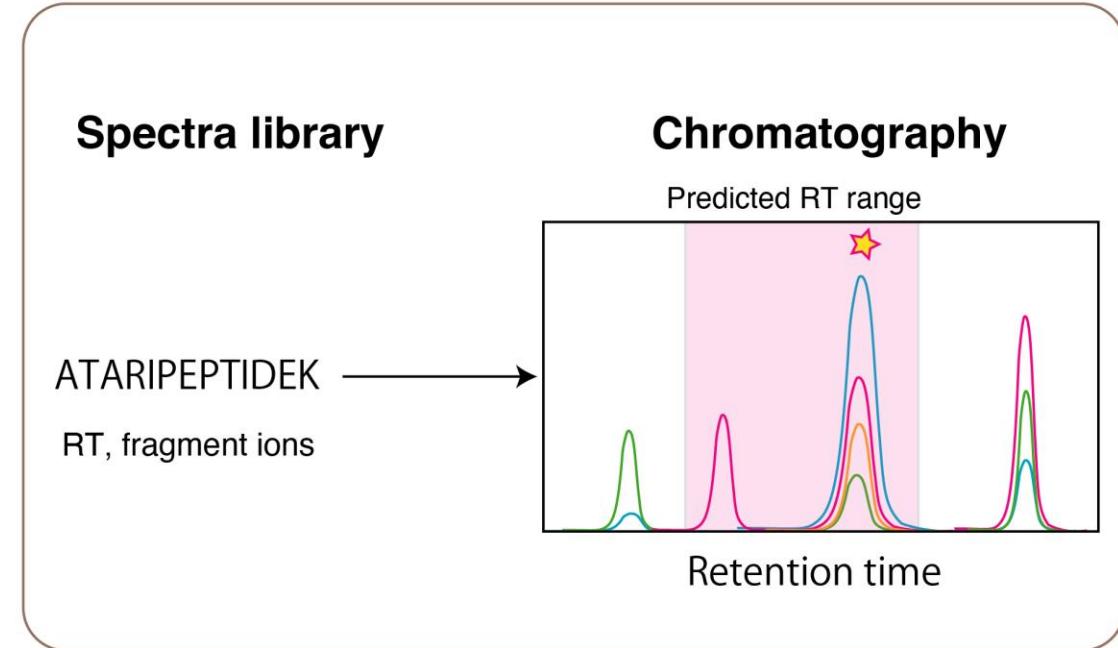
Fragment ion extracted chromatography

質量分析プロテオームデータ解析のアプローチ

Spectra-centric approach



Peptide-centric approach

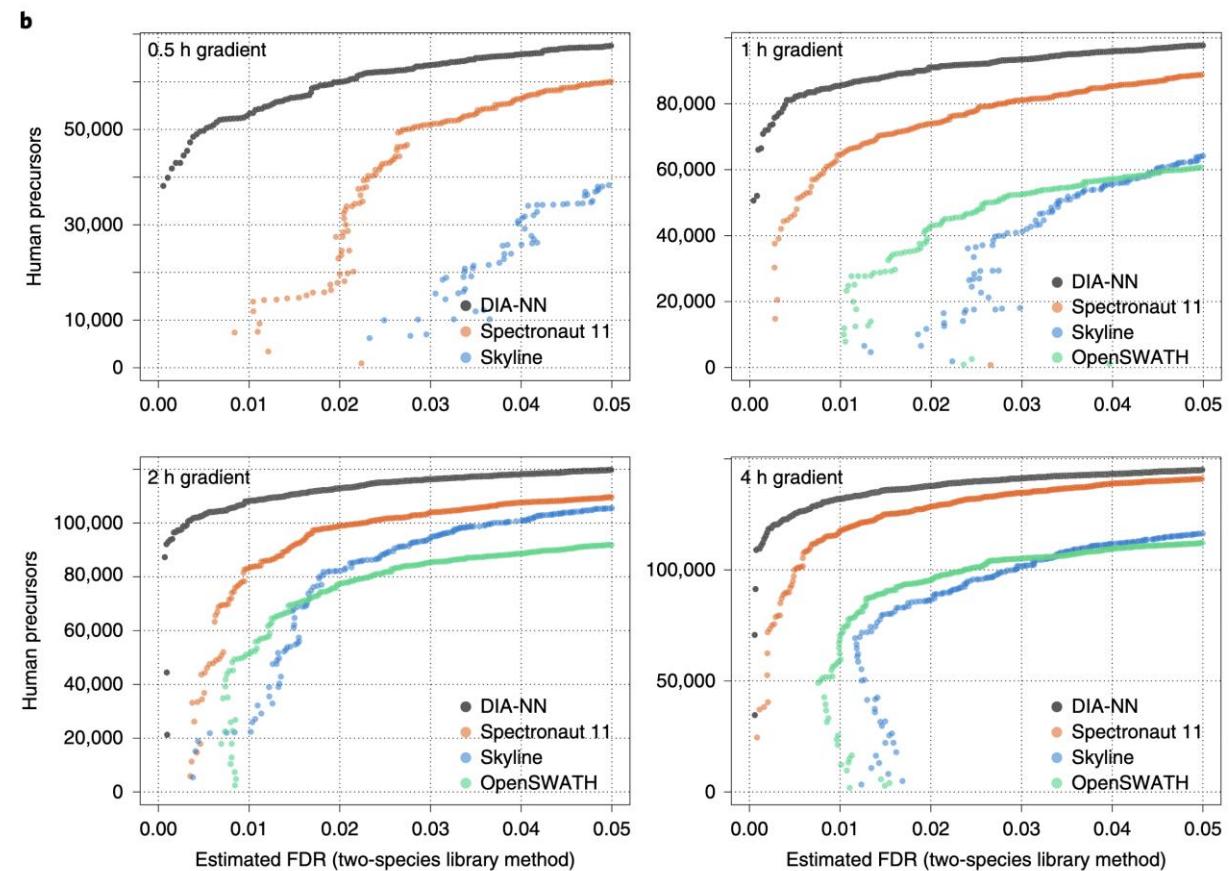


- Suitable for DDA analysis
- Applicable to PTMs search
- Accumulation of data in public repository
- Stochastic identification

- Suitable for DIA analysis
- More efficient peptide identification
- RT constraint enhances reliability
- High data consistency

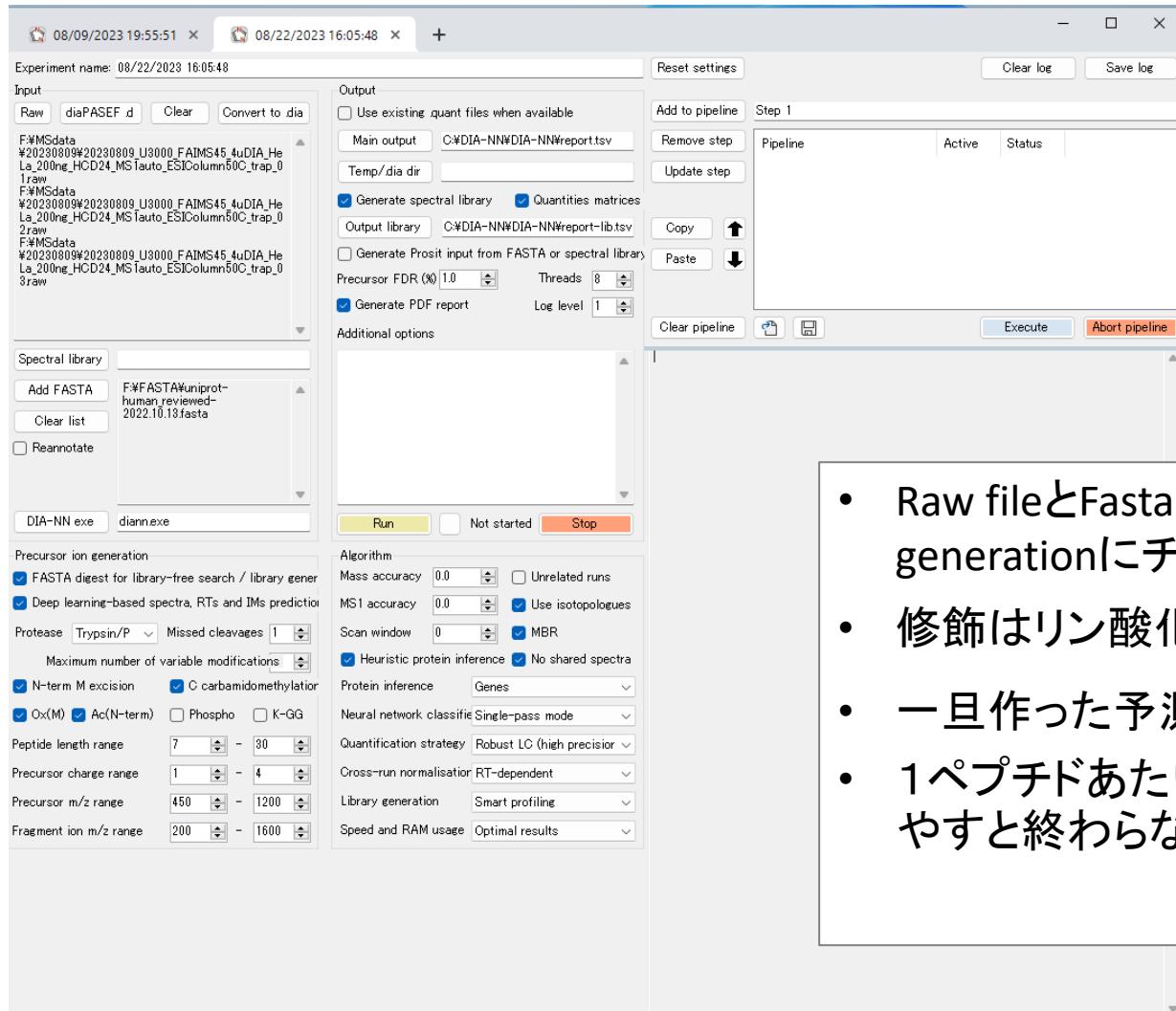
DIA解析ソフトの発展

	Library search	Library free-search	
		Spectra-centric	Peptide-centric
● OpenSWATH	○		
● DIA-Umpire	○		
● PECAN	○		
● Skyline	○		
● EncyclopeDIA	○		○
● DIA-NN	○		○
● DIAmeter	○	○	
● MaxDIA	○	○	
■ Spectronaut	○	○	
■ Scafold-DIA	○	○	
■ PEAKS	○	?	○



Demichev, V et al. *Nat. Methods* 17, 41–44 (2020).

DIA-NNによるライブラリーフリーサーチ

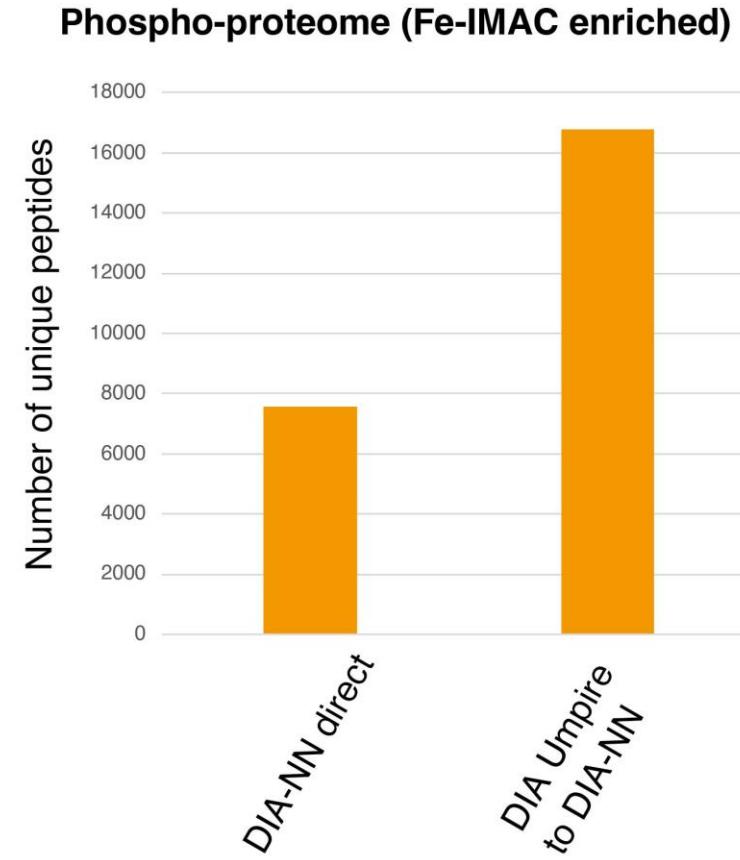
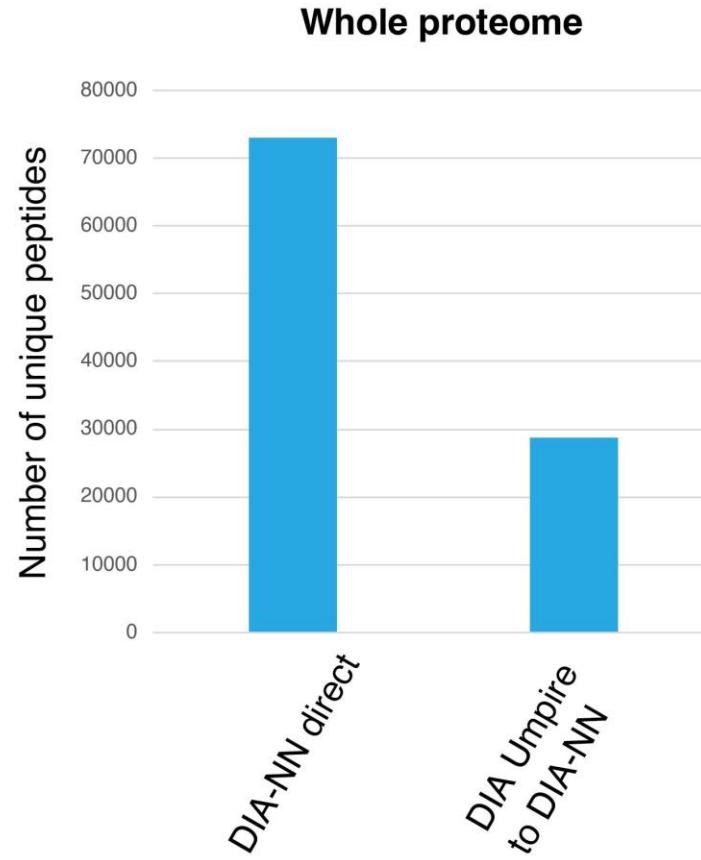
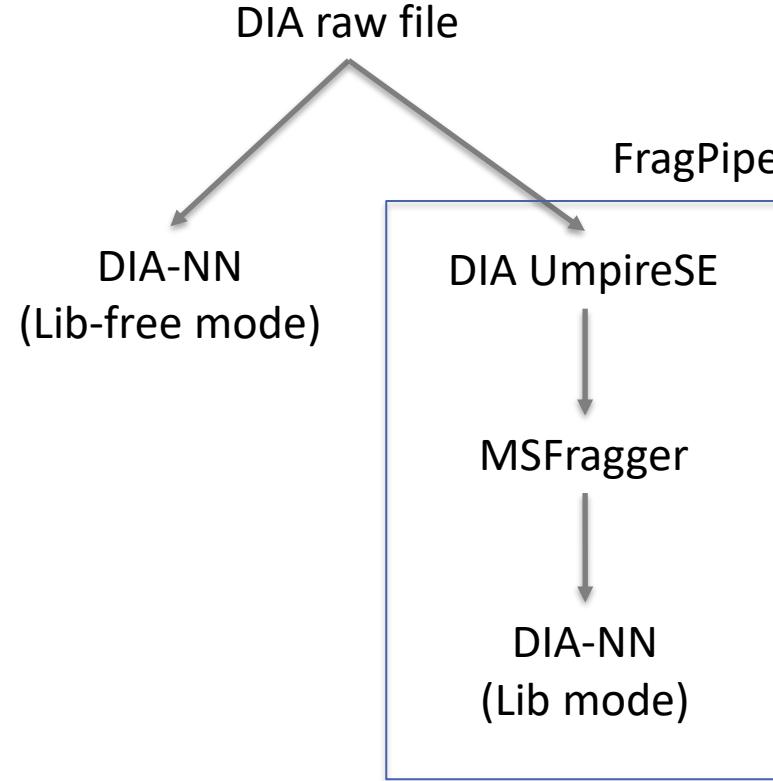


ライブラリーフリーサーチとは？

FASTAファイルから予測スペクトルライブラリーを直接構築して解析

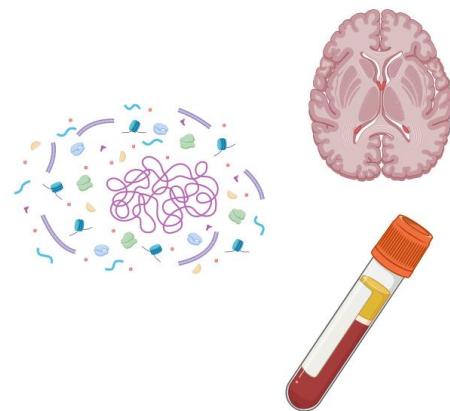
- Raw fileとFastaファイルを指定して、Precursor ion generationにチェックを入れてスタートするだけ。
- 修飾はリン酸化とユビキチン化に対応済み
- 一旦作った予測ライブラリーは使い回し可能
- 1ペプチドあたりの修飾部位数やミス切断数を増やすと終わらないので注意が必要。

DIA Umpire SE effectively enhances identification of phosphopeptides

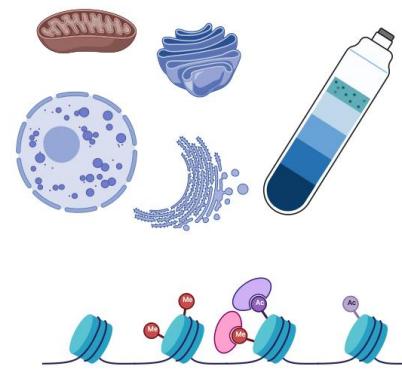


多様化するプロテオーム研究

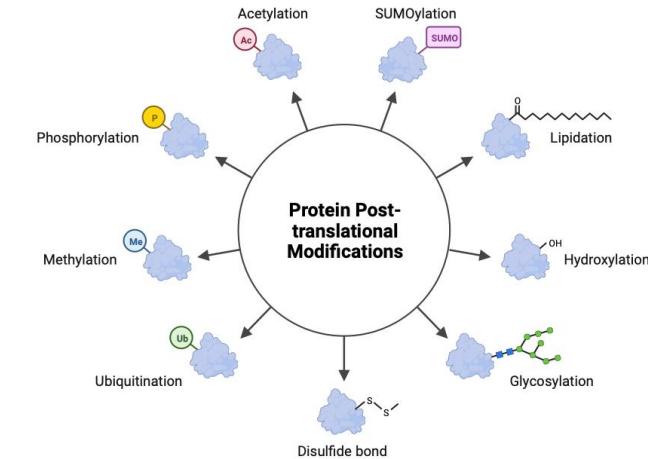
Whole proteome



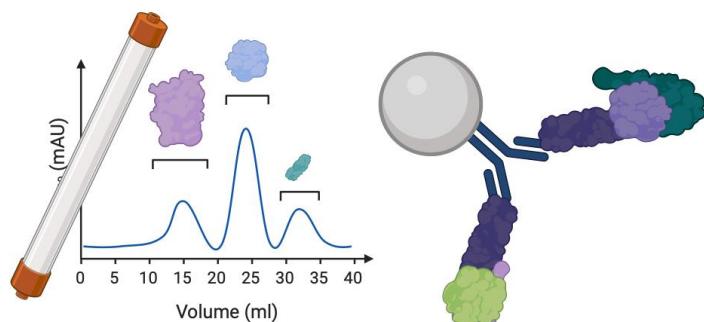
Organelle proteome



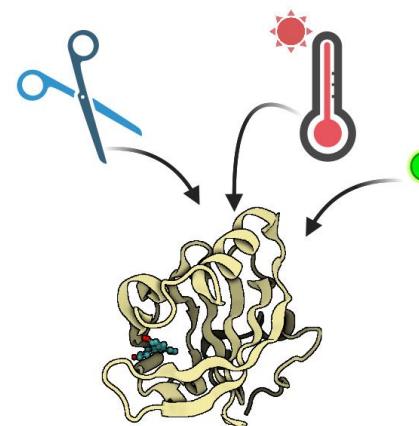
Modificome



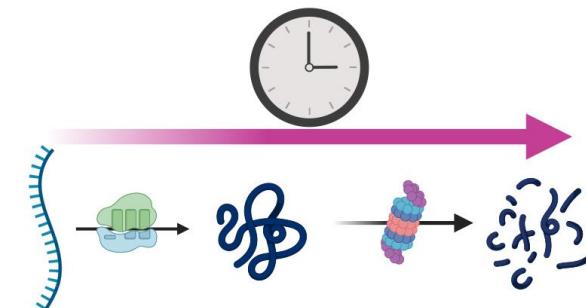
Interactome



Structural proteome



Proteostasis





Take home message

まとめ

- 単純な発現比較や個々のタンパク質情報の収集にはRepositoryと連動したプロテオームデータベースの活用が有効
- Repositoryにあるデータを再解析等をするには実験方法のある程度の理解が必要
- フリーソフトも多数あるので、誰でも再解析は可能

展望

- 今後ますます大規模化が促進（1データセット数百検体時代に）
- 他のオミクスデータも充実し統合が可能
- データサイエンスもChatGPTなどを活用すれば誰でも可能