

TogoTable (事例紹介)

情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター (DBCLS)
科学技術振興機構 バイオサイエンスデータベースセンター (NBDC)

河野 信

TogoTable

<http://togotable.dbcls.jp/>

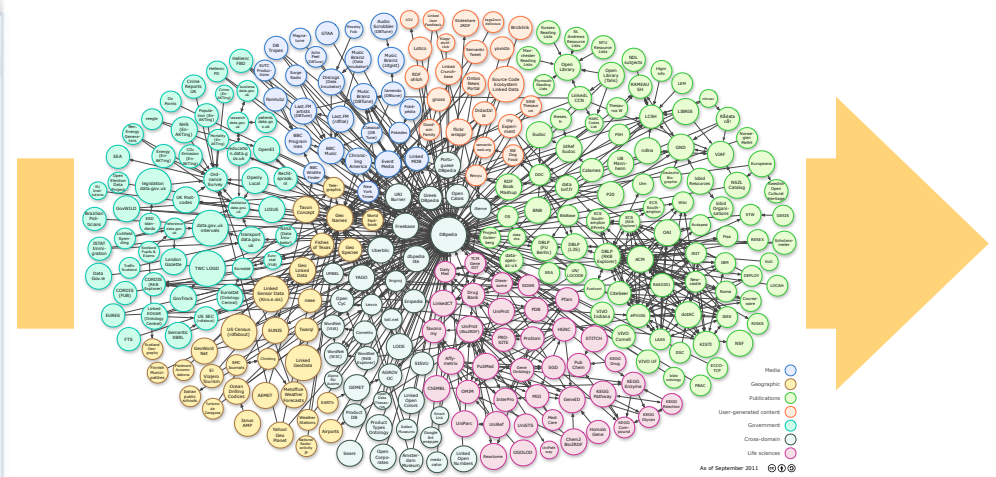
◆ RDFでつながったデータから様々なアノテーションを取得して表に追加

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected 0

Alignments [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	BacName: Full-Vimentin	941	941	100%	0.0	100%	P08670.4
<input type="checkbox"/>	BacName: Full-Vimentin	940	940	100%	0.0	99%	Q5R1W8.4
<input type="checkbox"/>	BacName: Full-Vimentin >sp P94198.3 VIMF_CERAE BacName: Full-Vimentin	937	937	100%	0.0	99%	Q4R4X4.3
<input type="checkbox"/>	BacName: Full-Vimentin	926	926	100%	0.0	98%	P02543.2
<input type="checkbox"/>	BacName: Full-Vimentin	901	901	97%	0.0	98%	P48635.3
<input type="checkbox"/>	BacName: Full-Vimentin	875	875	97%	0.0	97%	P31000.2
<input type="checkbox"/>	BacName: Full-Vimentin	874	874	97%	0.0	97%	P20152.3
<input type="checkbox"/>	BacName: Full-Vimentin	867	867	97%	0.0	97%	P02544.2
<input type="checkbox"/>	BacName: Full-Vimentin	839	839	96%	0.0	96%	P48670.1
<input type="checkbox"/>	BacName: Full-Vimentin	813	813	100%	0.0	86%	P09654.2
<input type="checkbox"/>	BacName: Full-Vimentin-1Q	716	716	99%	0.0	78%	P24789.1
<input type="checkbox"/>	BacName: Full-Vimentin-4	682	682	99%	0.0	76%	P24790.1



input tsv file: ファイルが選択されていません。

first line is header:

filename: sample.txt
 id:529d303cb08c8e133f000001
 total number:5
 status: 0
 Merged:0%

Gene names	Full name	UniProt ID	PDB ID	Pubmed ID	INSDC ID	RefSeq ID	UniGene ID	Ensembl ID	GeneID	KEGG GENES ID	UCSC ID
VIM	Vimentin	P08670									
SEPT7	Septin-7	Q16181									
YBX1	Nuclease-sensitive element-binding protein 1	P67809									
FSCN1	Fascin	Q16658									
CAT	Catalase	P04040									

Select Key: Select DB:

UniProt

Names and origin **Preview Data**

- UniProt ID
- Recommended name(Full name)
- Recommended name(Short name)
- Recommended name(EC)

input tsv file: ファイルが選択されていません。

first line is header:

filename: sample.txt
 id:529d303cb08c8e133f000001
 total number:5
 status: 0
 Merged:0%

Gene names	Full name	UniProt ID	PDB ID	Pubmed ID	INSDC ID	RefSeq ID	UniGene ID	Ensembl ID	GeneID	KEGG GENES ID	UCSC ID
VIM	Vimentin	P08670									
SEPT7	Septin-7	Q16181									

Select Key: : Select DB: :

p.txt
 c8e133f000001

Full name	UniProt ID	PDB ID	Pubmed ID	INSDC ID	RefSeq ID	UniGene ID	Ensembl ID	GeneID	KEGG GENES ID	UCSC ID	UniProt:Recommended name(Full name)	UniProt:Gene names
Vimentin	P08670	1GK6	2251132	AAA61279.1	NP_003371.2	Hs.455493	ENST00000224237	7431	hsa:7431	uc001iou.2	Vimentin	VIM
Septin-7	Q16181	2QAG	8037772	AAB31337.1	NP_001779.3	Hs.191346	ENST00000399034	989	hsa:989	uc010kxc.3	Septin-7	SEPT7
Nuclease-sensitive element-binding protein 1	P67809	1H95	1891370	AAA35750.1	NP_004550.2	Hs.473583	ENST00000321358	4904	hsa:4904	uc001chs.3	Nuclease-sensitive element-binding protein 1	YBX1
Fascin	Q16658	1DFC	3525578	AAA86442.1	NP_003079.1	Hs.118400	ENST00000382361	6624	hsa:6624	uc003sou.3	Fascin	FSCN1
Catalase	P04040	1DGB	2308162	CAA27721.1	NP_001743.1	Hs.502302	ENST00000241052	847	hsa:847	uc001mvm.3	Catalase	CAT

プロテオームデータの標準化と データベースの世界動向

情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター (DBCLS)
科学技術振興機構 バイオサイエンスデータベースセンター (NBDC)

河野 信

NGS vs MS

	NGS	MS
生データ	FASTQ	.raw/.wiff/.baf ...
解析ソフトウェア	マルチプラットフォーム オープンなものが多い	中にはWindows専用のもものも ようやくオープンスースなものがポツ ポツと
	TopHat/Bowtie/Cufflinks/GATK	Mascot/X! tandem/ProteinPilot/ MaxQuant/Skyline...
解析データフォーマット	SAM/BAM/VCF	mzML/mzIdentML/mzQuantML

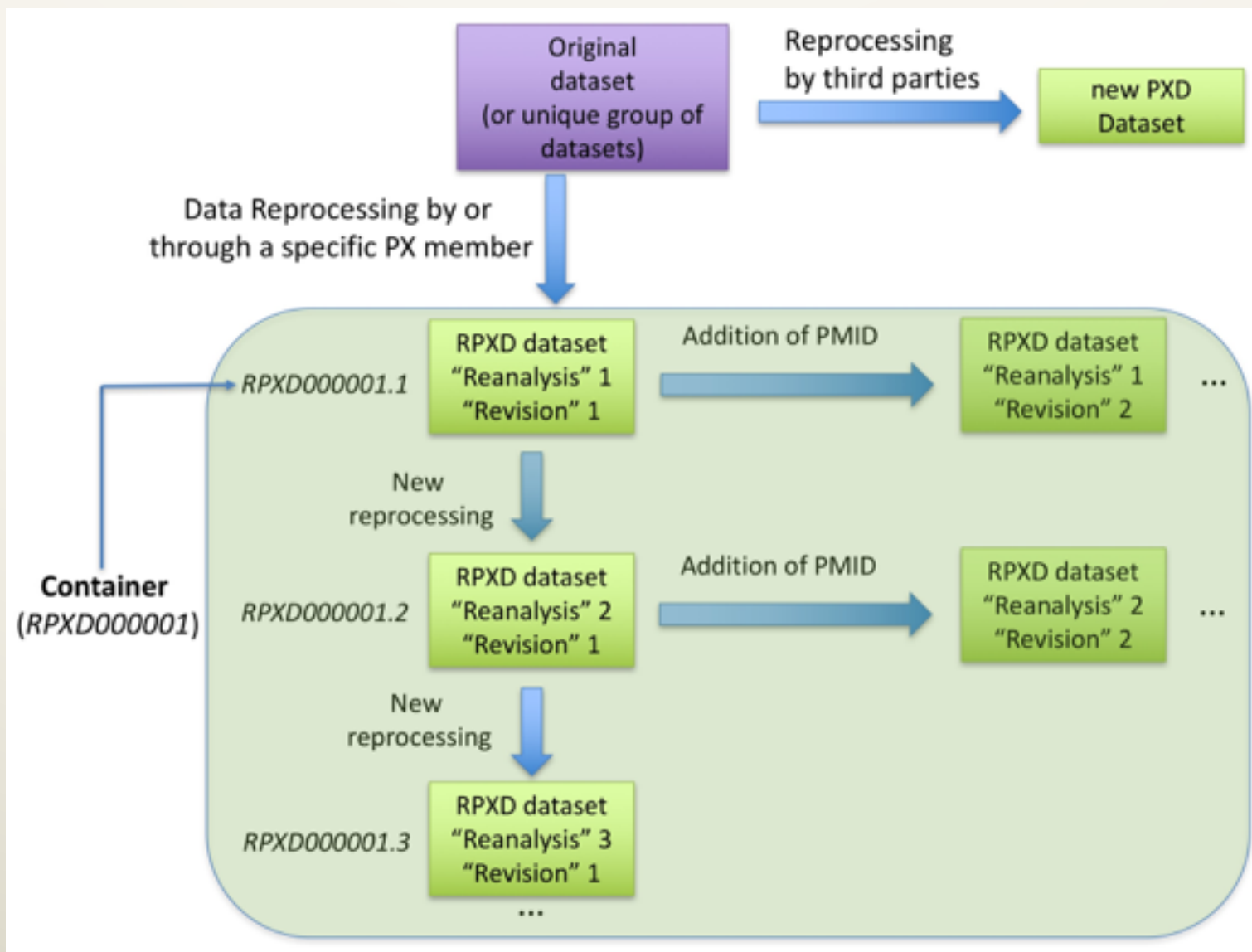
Proteomics Standards Initiative: Fifteen Years of Progress and Future Work

Eric W. Deutsch,^{*,†} Sandra Orchard,[‡] Pierre-Alain Binz,[§] Wout Bittremieux,^{||} Martin Eisenacher,[⊥]
Henning Hermjakob,^{‡,○} Shin Kawano,[◆] Henry Lam,^{□,▼} Gerhard Mayer,[⊥] Gerben Menschaert,[#]
Yasset Perez-Riverol,[‡] Reza M. Salek,[‡] David L. Tabb,⁺ Stefan Tenzer,[¶] Juan Antonio Vizcaíno,[‡]
Mathias Walzer,[‡] and Andrew R. Jones[▽]

HUPO-PSI standards

Working Groups	Guidelines	v.	Formats	v.	Controlled Vocabularies	v.	Software/Tools
Molecular Interactions <i>Group charter</i>	MIMix	1.1.2	PSI-MI XML	2.5.4	PSI-MI CV	2.5.0	
	MIABE	1.0.0	PSI-MI XML	3.0			
	MIAPAR	1.0.0	MITAB	2.7	PAR CV	n/a	
Mass Spectrometry <i>Group charter</i>	Mass spectrometry (MIAPE-MS)	2.98	mzML	1.1.0	PSI-MS	4.0.4	j mzML
			TraML	1.0.0			
	Identification (MIAPE-MSI)	1.1	mzIdentML	1.2.0			
Proteomics Informatics <i>Group charter</i>	Mass spectrometry Quantification (MIAPE-Quant)	1.0	mzQuantML	1.0.1			j mzIdentML
			mzTab	1.0.0			j mzQuantML
			proBed	1.0.0			j mzTab
			proBAM (<i>public review</i>)	1.0.0			
Quality Control <i>Group charter</i>			qcML (PSI spec. under construction)				
Protein separations <i>(Inactive)</i>	Gel electrophoresis (MIAPE-GE)	1.4	GelML	1.1.0	sepCV	1.0.0	
	Gel informatics (MIAPE-GI)	1.0					
	Column chromatography (MIAPE-CC)	1.1					
	Capillary electrophoresis (MIAPE-CE)	0.9.3	spML	1.0.0			
	Phosphoproteomics (MIASSPE)	0.9.					

再解析データ登録ガイドライン



Universal Spectral Identifier



- データセット中の個々のスペクトルデータを指し示すためのIDを設計
 - ペプチド同定結果をスペクトルにさかのぼって確認する
 - Spectral Libraryを作成した際のスペクトルを指定する

mzspec:<collection>:<subfolder>:<msRun>:<indexType>:<scanNumber>

mzspec:PXD000561:Control01:Adult_Frontalcortex_bRP_Elite_85_f09:scan:17555

Controlled Vocabulary (CV)



PSI-MS

- ・ Controlled Vocabularyであってオントロジーではない
 - ・ OBOに登録されてはいるが、Class/Property/Entity等の関係が整理されていない → ただひたすらに語彙が追加されるのみ

PSI-MOD

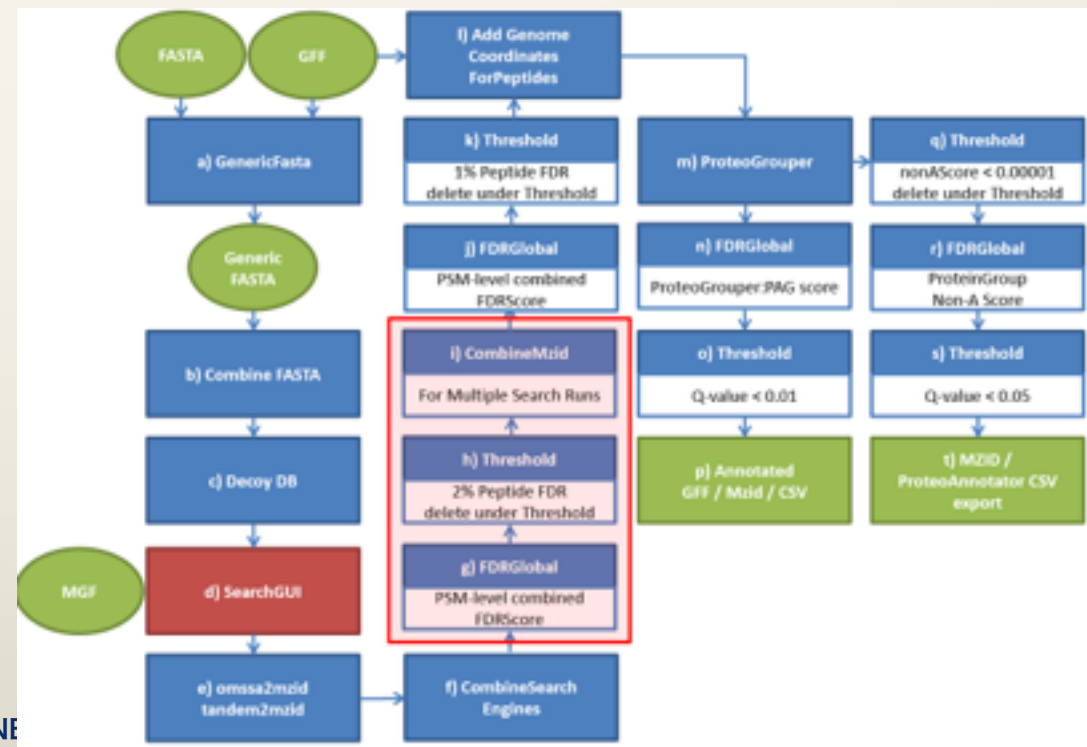
- ・ 更新停止中
 - ・ やる人がいないから
 - ・ UNI-MODの利用を推奨

Proteogenomics file format

- ProBAM/ProBED
 - IGV等ゲノミクス分野で作られているビューワーにプロテオームデータを表示させる
 - 基本的にBAM/BEDの拡張
 - ProVCFは難しそう？

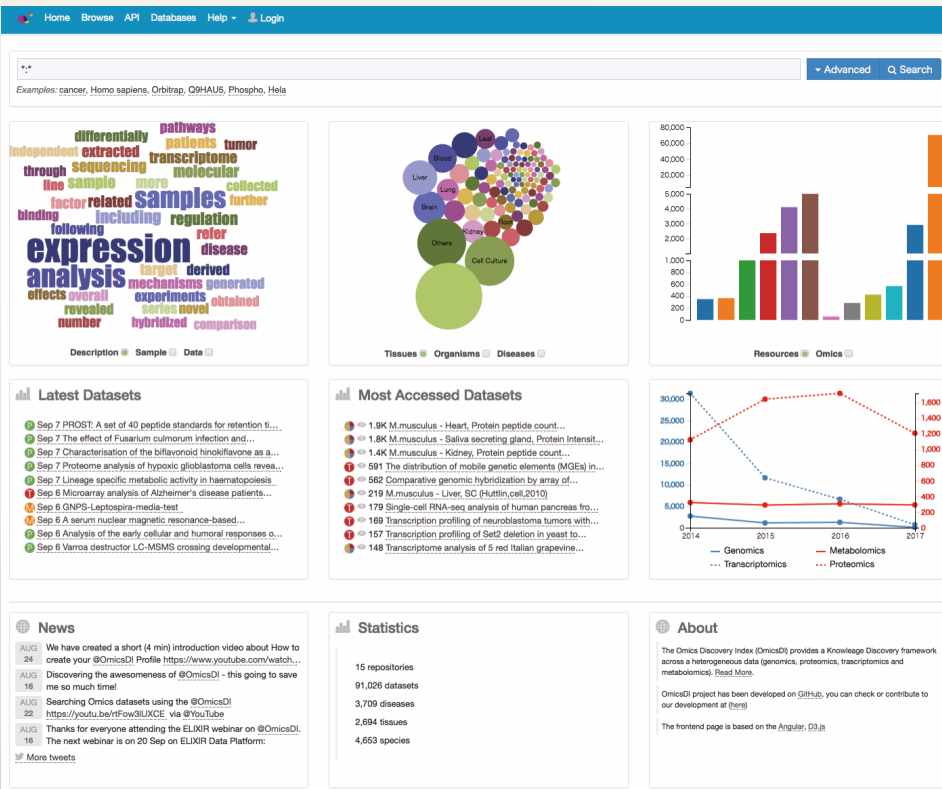
Proteogenomics 解析パイプライン

- ProteoAnnotator、他多数
- 解析対象が多様
 - がんプロテオーム
 - 新規生物のアノテーション/
アノテーションの高度化
 - メタプロテオミクス



Omics Discovery Index

- ゲノム/トランスクリプトーム/プロテオーム/メタボロームレポジトリからメタデータを収集
- オミクスDBをまたいで似たデータセットを見つける
- RDF化



Genome	Ensembl
	ENA
Transcriptome	ArrayExpress
	ExpressionAtlas
Proteome	GPMDB
	MassIVE
	PeptideAtlas
	PRIDE
	jPOST
Metabolome	GNPS
	MetaboLights
	MetabolomeExpress
	Metabolomics Workbench