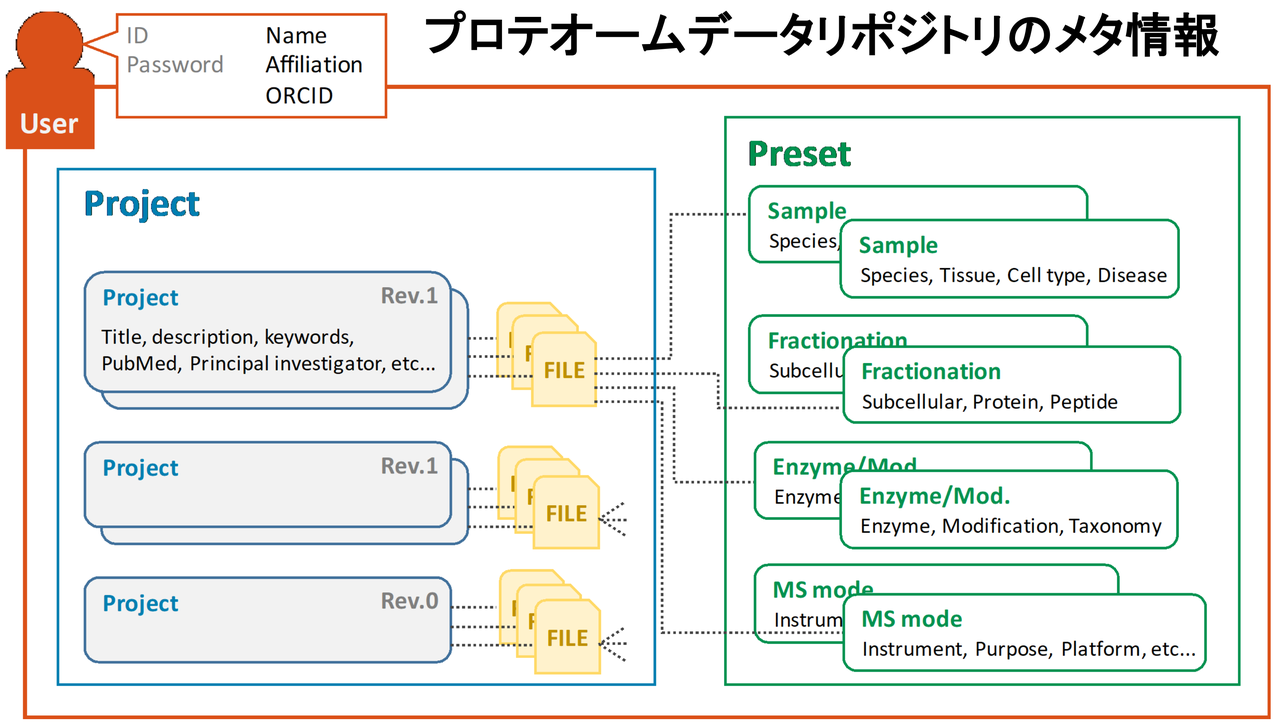
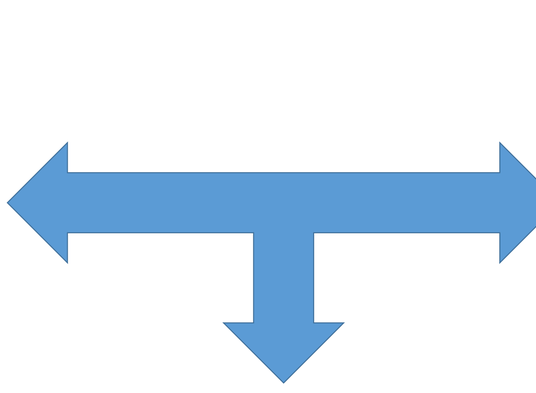
58 jPOST: メタデータのキュレーション

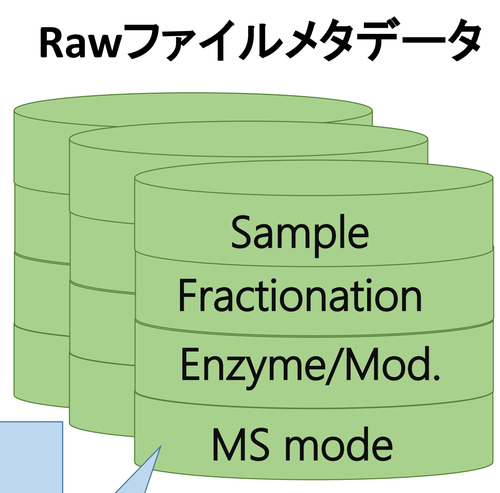
○小林大樹1)、荒木令江1)、奥田修二郎2)、渡辺由2)、守屋勇樹3)、河野信3)、山本格4)、松本雅記5)、高見知代5)、吉沢明康6)、田畑剛6)、杉山直幸6)、田中聡7)、岩崎未央8)、五斗進3)、石濱泰6)
1) 熊本大学大学院生命科学研究部　2) 新潟大学 大学院医歯学総合研究科　3)情報・システム研究機構 ライフサイエンス統合データベースセンター
4) 新潟大学 産学地域連携推進機構生体液バイオマーカーセンター　5) 九州大学 生体防御医学研究所　6) 京都大学 大学院薬学研究科　7)Trans-IT
8)京都大学iPS細胞研究所

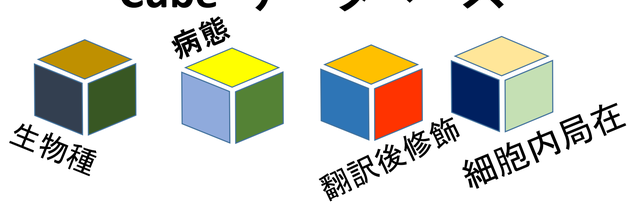
概要

jPOST(Japan ProteOme STandard Repository/Database, http://jpostdb.org/)は、ヒト、および様々な生物種由来のプロテオームデータを統合し、統一された再解析プロトコルによって信頼性の高い結果を集積して、より精度の高いデータベースを構築することを目的としている。プロテオームデータは実験研究者からメタデータが付加された形でリポジトリへと登録され、データベースに利用される。リポジトリに蓄積されたデータを、データベースで利用していくには、生データがどのような実験手順を踏んで得られたものか正確にトレースし、メタデータとして情報を整理していくことが必要となる。我々は、ヒト臓器、疾患、モデル生物サンプル、およびタンパク質定量、翻訳後修飾などのプロテオーム解析にて導き出されるデータに焦点を当て、ユーザーからのサンプルおよび解析手法に関するメタデータをキュレーション(精査、および修正)し、データベースに有益な情報としている。また、メタデータの用語を正確に統制することで、ユーザー側の視点で整備された情報の提供を目指している。









例: jPOSTプロジェクトID:JPST000203のメタデータ

1. “Sample”メタデータ

	A	B	C
1	Sample		
2			
3	Species	Human	
4	Sample Type	Tissues	
5	Organ	Colorectum	
6	Disease	Cancer	
7	Disease name	Colorectal cancer	
8	Note-1	iTRAQ plex 114 or 116 : 12 patients with no metastasis	
9	Note-2	iTRAQ plex 115 or 117 : 12 patients with metastasis	

2. “Fractionation”メタデータ

	A	B	C
1	Fractionation		
2			
3	subcellular fractionation	ex Membrane	
4	subcellular fractions	1	
5	subcellular replicates	1	
6			
7	protein fractionation	ex SDS-PAGE	
8	protein fractions	10	
9	protein replicates	1	
10			
11	peptide fractionation	ex Strong cation exchange column	
12	peptide fractions	30	
13	peptide replicates	1	

3. “Enzyme_Mod.”メタデータ

	A	B	C	D
1	Enzyme_Mod			
2				
3	Fixed Modification	Carbamidomethyl (C)	iTRAQ4plex (K)	iTRAQ4plex (N-term)
4	Variable Modification	Oxidation (M)	iTRAQ4plex (Y)	
5	Enzyme	Trypsin/P		
6	Taxonomy	Homo sapiens		
7				
8				
9				
10				
11				

4. “MS mode”メタデータ

	A	B
1	MS	
2		
3	MS instrument	LTQ Orbitrap Velos
4	Instrument Mode	Data-dependent acquisition (DDA)-high resolution
5	Quantification Platform	Product ion (reporter quantification)
6	Multiplex method	iTRAQ 4-plex
7	Plex number	4
8	Plex name-1	iTRAQ-114 (12 colorectal cancer patients with no metastasis)
9	Plex name-2	iTRAQ-115 (12 colorectal cancer patients with metastasis)
10	Plex name-3	iTRAQ-116 (12 colorectal cancer patients with no metastasis)
11	Plex name-4	iTRAQ-117 (12 colorectal cancer patients with metastasis)
12		
13		

ヒト疾患サンプルメタデータアンケートシート

Sample メタデータ(jPOST ID: JPST00XXXX)

・Species:
☐Human ☐Mouse ☐Others ( )

・Sample Type:
☐Tissue ☐Primary cell ☐Cell Line () ☐Others ( )

・Organ:
☐Brain ☐Lung ☐Stomach ☐Colorectum ☐Small intestine ☐Gallbladder ☐Esophagus
☐Gastrointestinal tract ☐Bile Duct ☐Liver ☐Pancreas ☐Kidney ☐Adrenal gland ☐Thyroid gland
☐Heart ☐Skeletal Muscle ☐Bone ☐Skin ☐Soft tissue ☐Head and Neck ☐Eye ☐Bone marrow
☐Thymus ☐Spleen ☐Blood ☐Lymph node ☐Prostate ☐Testis ☐Breast ☐Ovary ☐Placenta ☐Urine
☐Fece ☐Other ( )

・Disease:
☐Normal ☐Cancer ☐Nervous disease ☐Metabolic disease ☐Cardiovascular disease ☐Infectious disease
☐Immune disease ☐Other disease ( )

・Disease name:

・Note

メタデータキュレーションの内容

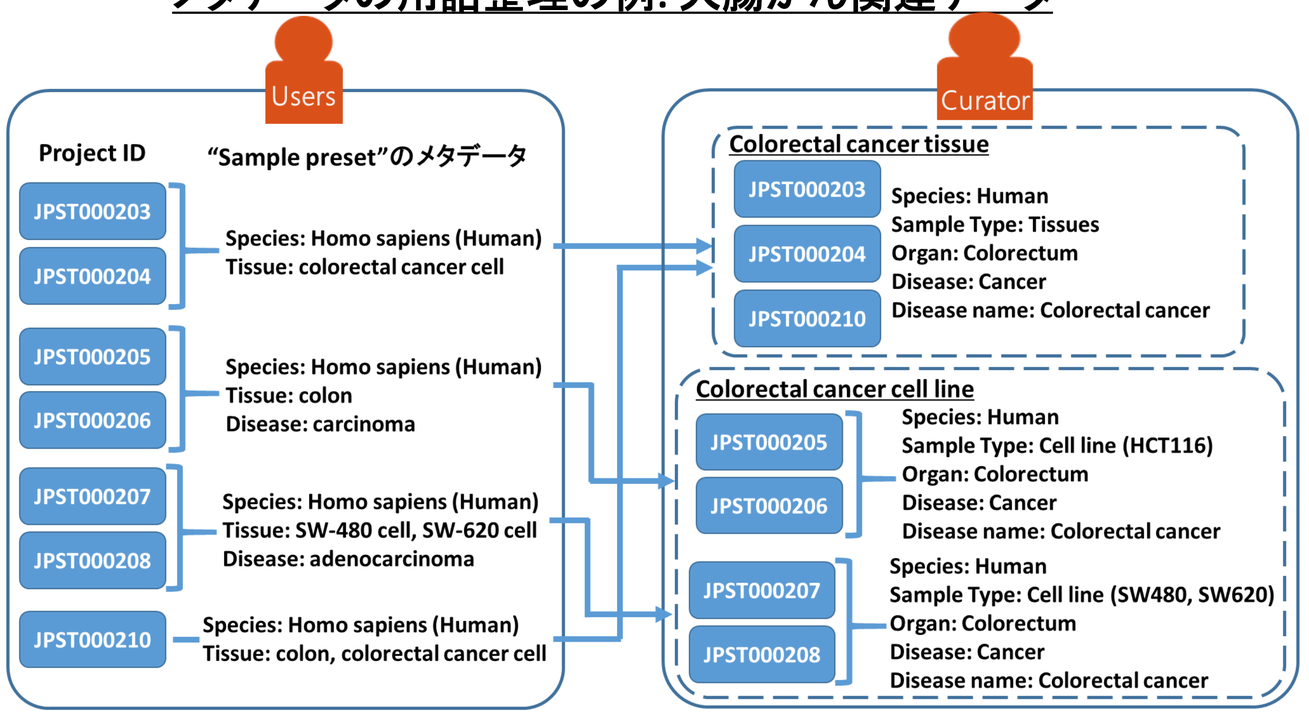
1. メタデータの整理

○統一した言語によるヒト疾患サンプルメタデータの整理
○メタデータの修正、および必要なメタデータの付加

2. 自動メタデータ処理では判断が困難な情報の確認

○サンプル情報
・詳細な疾患組織の情報・特殊な培養細胞株(派生株など)の情報
・Multiplexでの定量解析における各plexのサンプル情報
○プロジェクトID間の関連
・異なるIDだが、解析サンプルが同じである場合のIDの関連性の紐づけ

メタデータの用語整理の例: 大腸がん関連データ



今後の方針

・登録済・新規登録データのメタ情報を随時キュレーションし、データベースの充実化を図る。
・メタデータキュレーションのセミ自動化システムの開発を目指す。

謝辞

本プロジェクトは、科学技術振興機構 (JST)・バイオサイエンスデータベースセンター (NBDC) の「統合化推進プログラム」の助成により実施しました.

Licensed under CC-BY 4.0 ©2017 jPOST project, Daiki Kobayashi (Kumamoto University)