

2021年10月5日
トーゴーの日
シンポジウム2021

NBDCの10年を振り返って

- できたこと、できなかったこと -

高木利久

NBDC設立の経緯と目的および3つの柱

- JST BIRD(2001～2011)
バイオインフォマティクス推進センター事業
+
= JST NBDC(2011～)
- ROIS DBCLS(2007～) ライフサイエンスデータベース統合推進事業
統合データベース事業

- NBDCの目的

- データやDBの共有・統合による研究者の利便性向上
- データの価値の最大化、データ駆動型の生命科学への貢献
- 我が国の中核センターとして機能する（外部との連携）

- NBDCの3つの柱

DBやデータを統合的に扱える
サービスの提供
(外部連携およびポータルサイトの
構築・運用)

成果
DB

分野別（データ種別や生物種別）に
国内外データを統合的に扱える
統合DB整備
(統合化推進プログラム)

技術

技術

分野別統合DBを含めた国内外主要DBを統合的に扱うための技術開発
(基盤技術開発) with DBCLS

10年間の成果（その一部）

詳しくはポスターを

- 基盤サービス
 - integbioDBカタログ、DB横断検索、DBアーカイブ
 - RDFポータル（with DBCLS）
 - ヒトデータベース（日本人由来データのレポジトリ）
- 実験研究者向けDB開発
 - TogoVar（日本人ゲノム多様性統合データベース）
 - 日本人のホールゲノム約7600人、その他のデータ収録
- 統合化推進プログラム（ファンディング）
 - ChIP-Atlas(エピゲノム)、KEGG MEDICUS(疾患パスウェイ)
GlyCosmos(糖鎖)、PDBj(タンパク質立体構造)
 - MicrobeDB.jp(マイクrobiオーム)、DBKERO(多層オミックス)
PlantGARDEN(植物ゲノム)、MetabaoBank(代謝物)
 - jPOST(プロテオーム)
- 基盤技術開発(with DBCLS)
 - TogoDX、PubCaseFinder, CRISPRdirect、PubAnnotation
 - 統合DB（知識グラフ）の構築・利用のための技術開発
- 普及活動、コミュニティ形成
 - AJACS（講習会）、バイオハッカソン（国際標準化）

データベースの統合とは？

- データベース統合とは？
 - 多種多様なDB群がシームレスに繋がっている状態（＝知識グラフ構築）
 - これまでのDB統合はDBセンターに閉じた形で（例：米国NCBI、欧州EBI）
 - さまざまな機関をまたいだ分散型の本格的なDB統合という営みはなかった
- なぜ統合が必要か？
 - 生命科学の目的は様々な要素が相互作用している複雑なシステムの理解
 - マルチオミックスの解析には多種多様なデータの統合が必須
 - 統合なければ、研究者の手による個別オミックス解析の突合が必要
- NBDC+DBCLSが目指す統合の姿とは？
 - データ駆動型研究を先導する統合データベースの提供
 - 統合DBを生産するためのベストプラクティスの蓄積とリテラシー普及
 - それを容易にする、汎用的な技術基盤の提供
- APIなど他の統合手法との比較、違いは？
 - 生命システムの理解には種々の階層のデータを対象とした統合必要
 - IDなど未整理の小規模な個別データを小刻みに取得するAPIでは実現不能

知識グラフ(RDF)を用いたデータベース統合

- 生命科学のデータや知識は複雑で多様
- 貴重なデータを情報を落とさずにデジタル化したい
- グラフ構造は表などでは表現しきれないものを格納可能
- 例：UniProtは知識グラフ表現の先行成功事例
- NBDC/DBCLS/DDBJはEBI/NCBIとも連携しDBの知識グラフ化推進
- 現状では知識グラフ表現の標準仕様であるRDFを利用
- 前述のデータ統合を目指している組織は世界的にもまだ少ない
- 知識グラフの構築は普及・発展途上
- NBDC/DBCLSで開発してきた技術基盤のさらなる高度化を推進
- 複雑な知識グラフの構築と利活用により
 - 実験研究者に役立つだけでなく **新しいデータサイエンスを開拓**したい

知識グラフ(RDF)化されたデータベース

- 塩基配列とアノテーション
 - INSDC (DDBJ/DBCLS) 
- ゲノム情報
 - Ensembl (EBI) 
 - RefSeq (TogoGenome) 
- アミノ酸配列とアノテーション
 - UniProt (SIB) 
- タンパク質立体構造
 - PDB (PDBj) 
 - BMRB (PDBj) 
 - FAMSBASE (Chuo U) 
- 化合物
 - PubChem (NCBI) 
 - ChEMBL (EBI) 
 - Nikkaji (JST) 
- 遺伝子発現
 - RefEx, GTEx (DBCLS) 
 - ExpressionAtlas (EBI) 
- サンプル
 - BioSamples (EBI/DDBJ) 
 - JCM (RIKEN) 
- 医学 (Med2RDF)
 - ICGC, COSMIC, CIViC 
 - DGIdb, OpenTG-Gates 
 - ClinVar, dbSNP, 
 - dbVar 
 - ExAC, gnomAD 
 - HiNT, INstruct 
- 糖鎖
 - GlyTouCan, 
 - GlycoEpitope, WURCS, 
 - GGDonto, PAConto
- プロテオーム
 - jPOST 
 - The Human Protein 
 - Atlas
- パスウェイ
 - Reactome (EBI) 
- その他
 - MeSH (NCBI) 
 - BioModels (EBI) 
 - MBGD (NIBB/DBCLS) 
 - Quanto (DBCLS) 

TogoDX 開発の目標

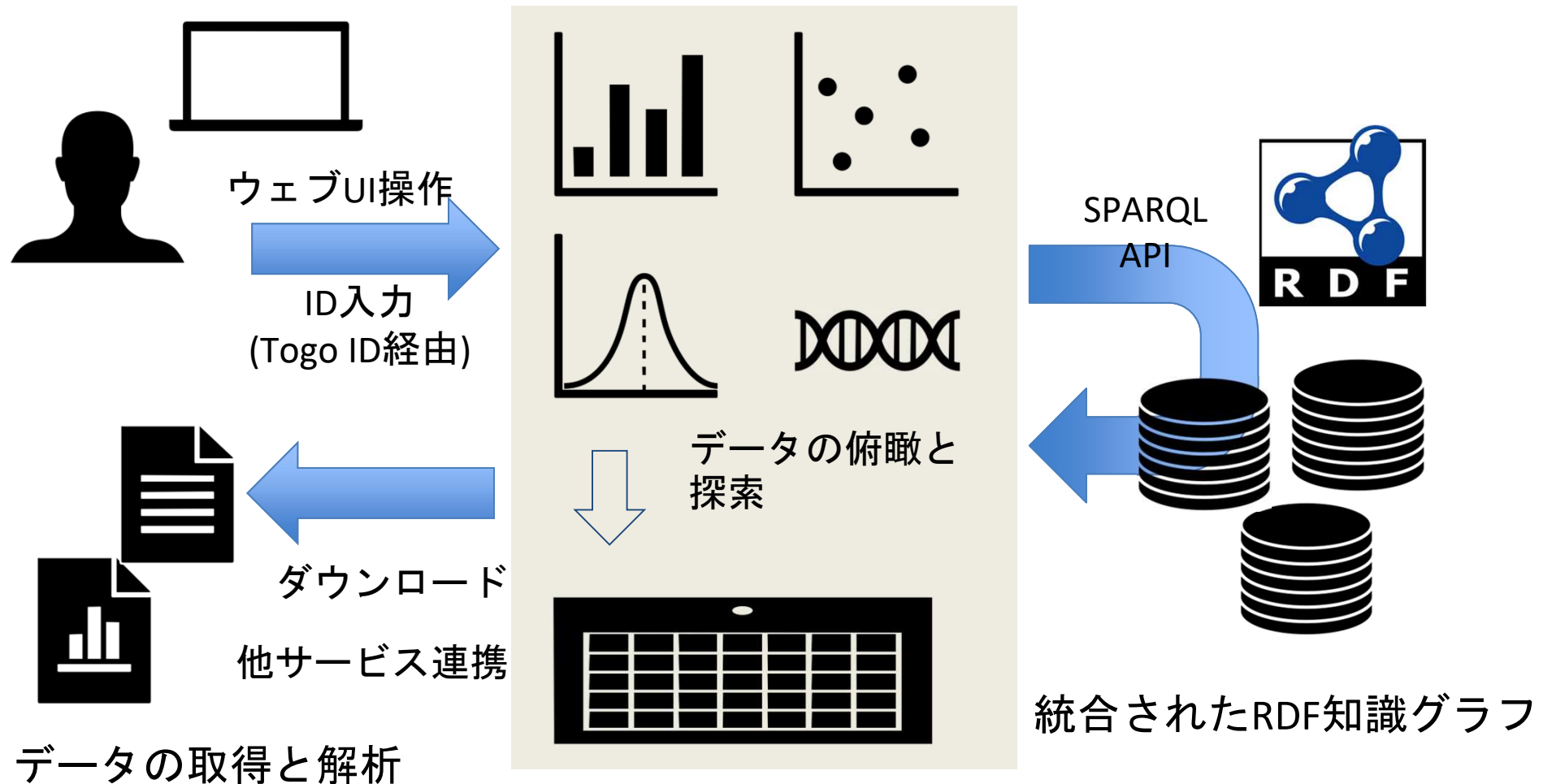
- NBDC, DBCLS, DDBJ, 統合化推進プログラムで作ってきた知識グラフの統合的活用インタフェース
- 統合DBへのニーズ
 - 膨大かつ多様なデータを俯瞰したい
 - 複雑なグラフから必要なデータを柔軟に組み合わせて抽出したい
 - 抽出したデータを機械学習・AIなどデータサイエンスで活用したい
- 統合データはあらゆる生物種の様々な生命現象に広範囲に
- TogoDXでヒト関係のデータの統合的利活用事例を開発
 - ゲノム、遺伝子、発現、タンパク質、立体構造、化合物、医薬品、疾患
- 知識グラフにおけるデータ統合の鍵
 - IDの相互変換による知識グラフ間の連結 (TogoID)
 - 知識グラフからの属性ごとのデータ抽出 (RDF-configによるSPARQL生成)

TogoDX (Togo Data eXproler)

利用者

TogoDX

RDFポータル





TogoDX

遺伝子

タンパク質

タンパク質
立体構造

相互作用

化合物

糖鎖

疾患

バリエーション

TOGO DX		Attribute keys		Attribute values		Log scale	
Gene	> High-level expression	Cere Cere Brai Co Per Spli V Ly Spli Thymus Bone SKI Place ProC Testis 1,955	H-Musc Eso S Ir C Liver Lu Kid Pit T A Pancreas Sa				
	> Low-level expression	Cerebellum	Thymus 176	Bone ma	Placenta 234	Pros Testis 157	Muscle 174 Color Liver/ f Pancreas 347
Protein	> Molecular mass	0-10 kDa 15,159	10-20 kDa 21,139	20-30 kDa 11,380	30-40 kDa	40-50 kDa 50-60	60-70-80 91 11 2
	> PTMs	Phosphoprotein 9,266	Glycoprotein 6,457	Disulfide bond 5,728	Acetylation	Ubl conju	Isopep Lip Me C2
	> # of transmembrane domains	0 65,969					1 6,178 2 34 7
Structure	> # of peptides in a PDB entry	1 peptide 140,890					2 peptides 20 3 p4
	> Other related molecules	Water 42,851	Zinc ic Sulfa Glyc 1,2- Maç Chl N A Ca Sc FC A I I I I	Other molecules 27,510			
Interaction	> Compounds in pathway	Metabolism 1,423		Disease 30	Signaling	Metaboli Transpo Imm	Ni G V C H D N D C I I
	> # of interacting proteins	0 76,991					1 2 3 4 1
Compound	> Action type	Inhibitor 3,048		Antagonist 904	Agonist 824	Bind Bio M P H A f i	
	> Biological role	biochemical role 17,443					inhibitor 2 antimicrob pharma aet ep m x p b p
	> Drug indication	Neoplasms	Immune	Patholo Infectio	Digestiv Respira	Nervou Skin an	Cardio Femal Male I Hemi Endo Nutri Con Mus Ey S O C V
Glycan	> Molecular mass of glycans	20 400-600 600-800	800-1000 I	1000-1200 D	1200-1400 D	1400-1600	1600-18 1800-20 2000-2200 2200-2400 2400-26 2600-2800 2800-3000 3000-3200 3200-3400
	> Tissues	colon 114	lung 82	liver 44	testis 43	lymph nod small inter skin glanc heart	3 blood s pa
Disease	> Diseases in MeSH	Nervous Syst Pathological Infections		Congenital Neoplasm		Cardic Skin	Musc Fem Nutri Dige Hen Eye Wo Re Im Stc Ma En A C
	> Diseases in NANDO	Specific pediatric chronic disease 1,678					Designated intractable disease 1,029
	> Cross referenced disease DBs in Mondo	UMLS 18,101	ICD10 11,583	DOID 9,878	Orphanet	OMIM 9,1	SCTID 9,0 MESH 7 NCIT 6 GARD ICD9 EFIC NG S I I
Variant	> GWAS	Information entity 88,826				Material property 31,883	Process
	> Clinical significance	Uncertain significance 206,658		Likely benign 148,781	Benign 113,835	Pathogenic	Likely Co Nc B I

Condition builder

Clear

Select target dataset

Ensembl gene

Ensembl transcript

NCBI gene

UniProt

PDB

ChEBI compound

ChEMBL compound

PubChem compound

GlyTouCan

Mondo

MeSH

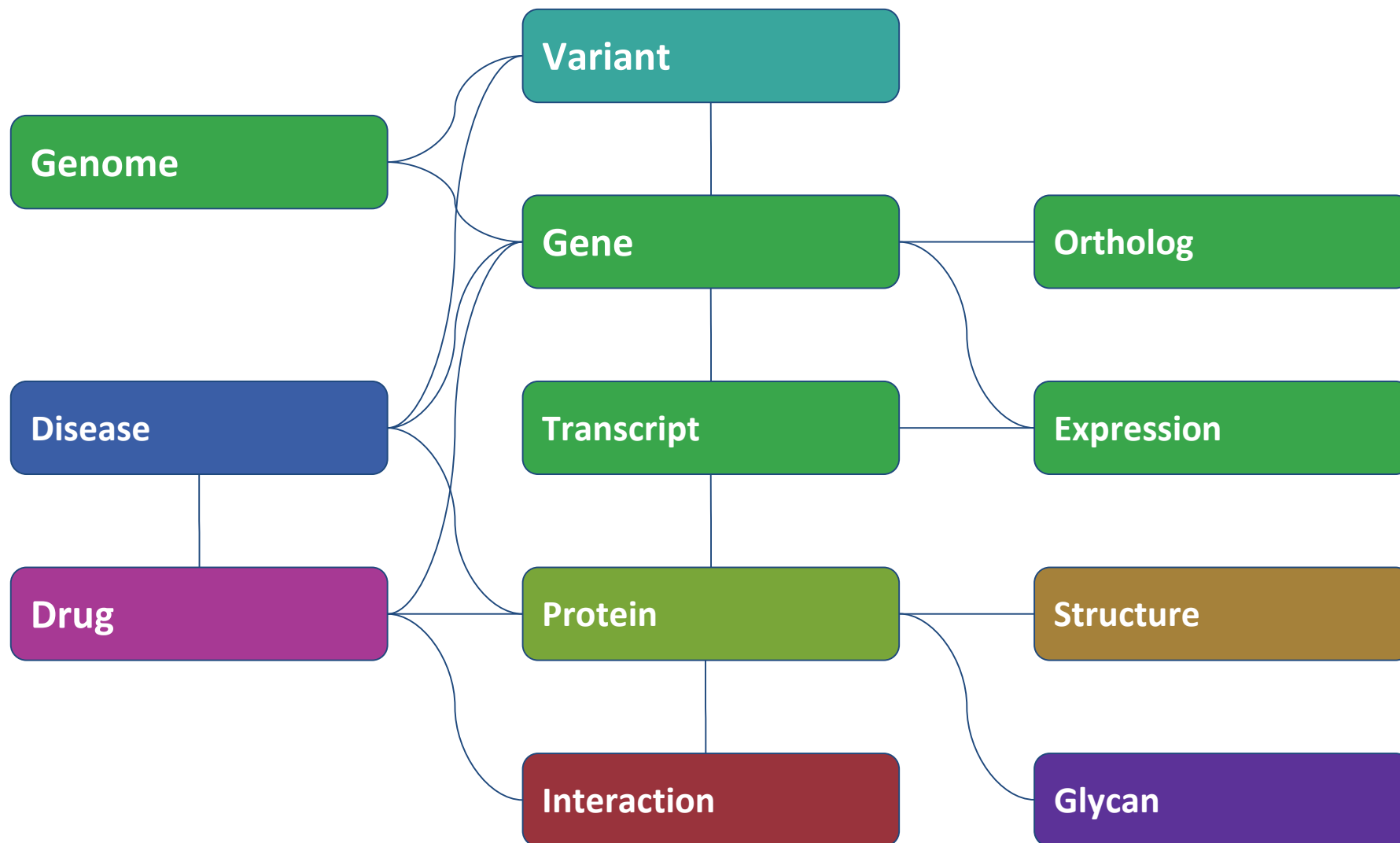
NANDO

Human Phenotype Ontology

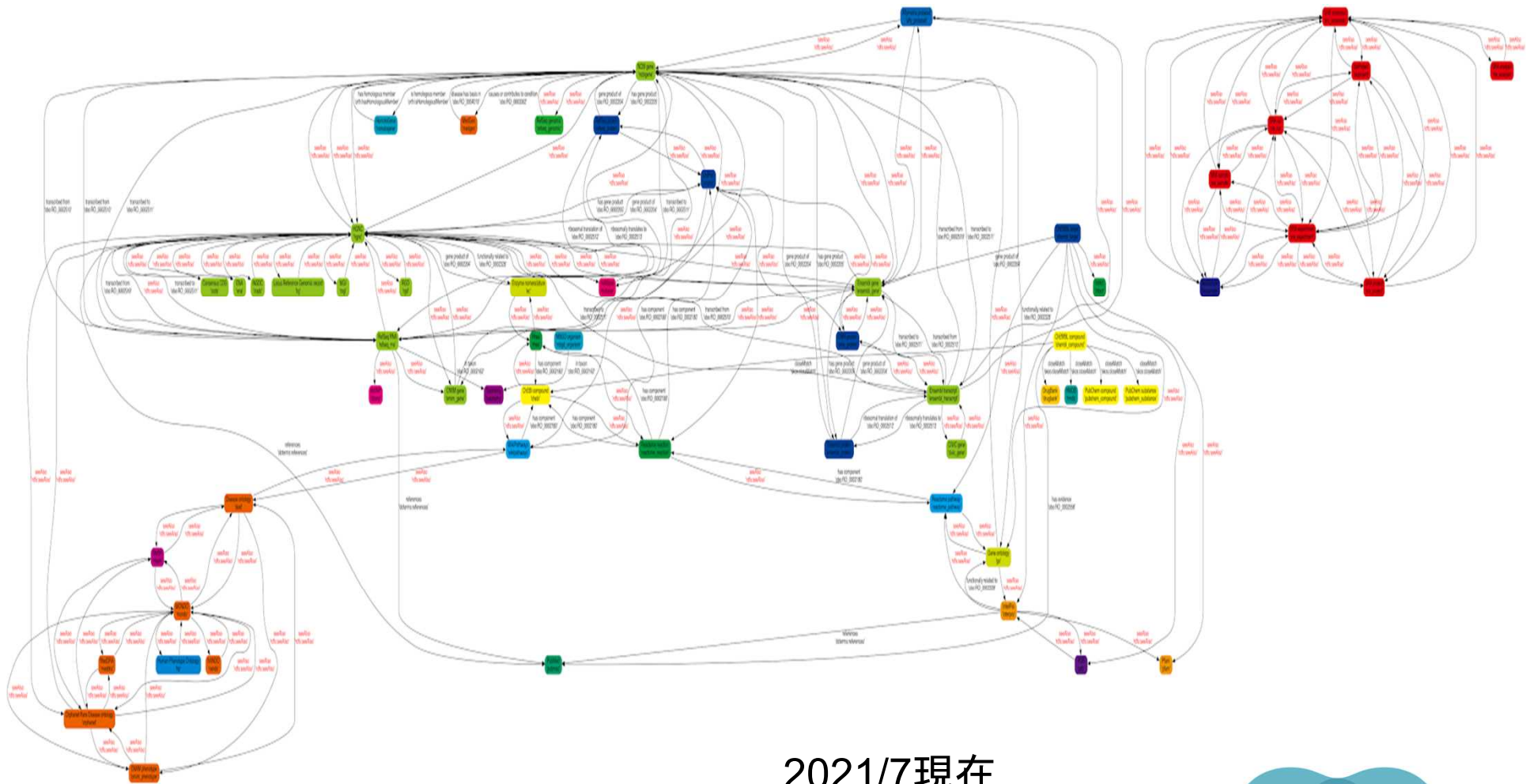
TogoVar variant

選択したデータベースのIDを軸にデータセットを作成可能

TogoDXが扱うオミックスデータ



TogoIDによるリンク情報の集約と管理

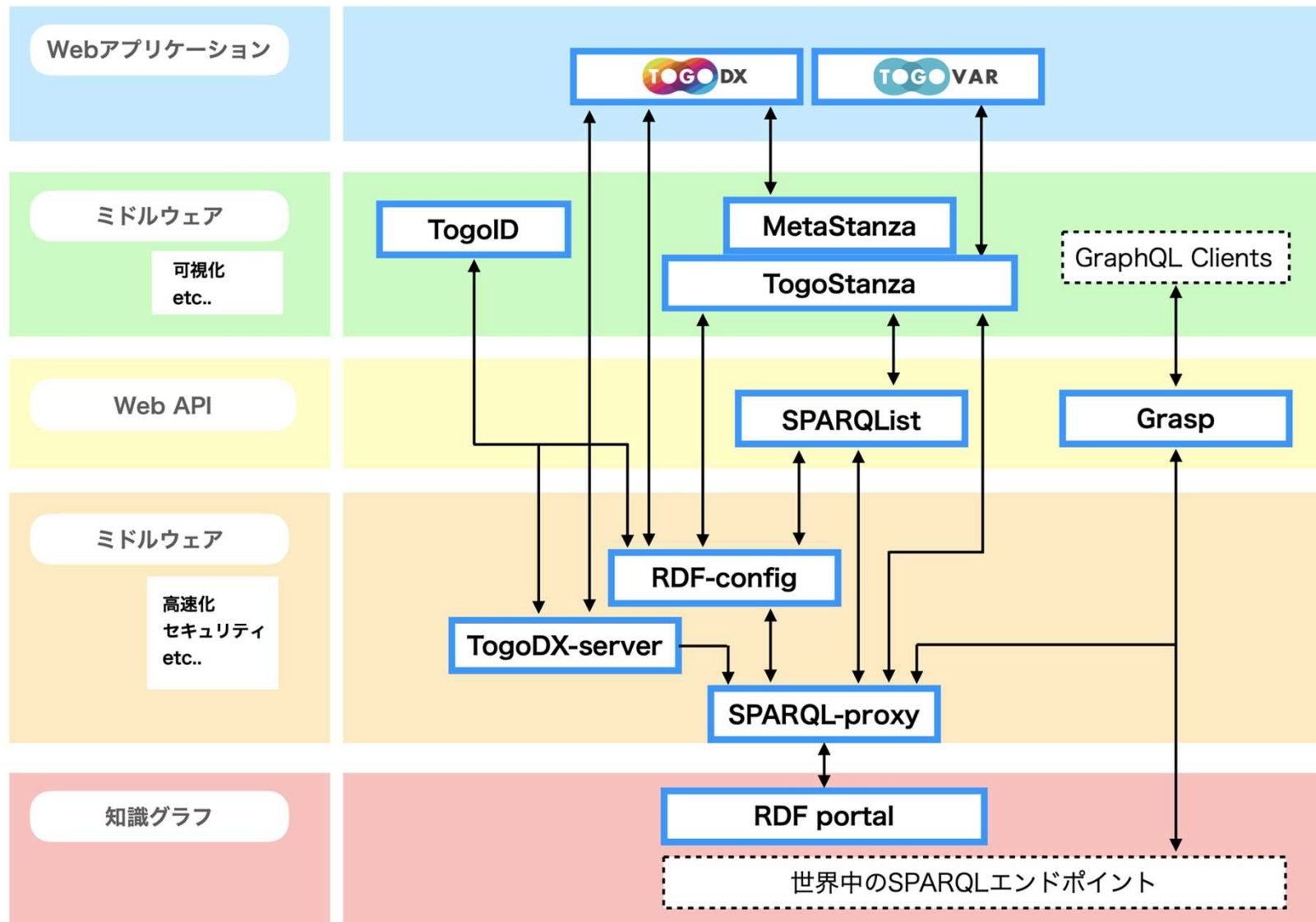


2021/7現在

- ・ 62データベース
- ・ 152データベースペア



統合化実現のために開発した基盤技術



青枠はNBDC/DBCLSが開発したもの



TogoDXで知識グラフを俯瞰する

遺伝子

タンパク質

タンパク質
立体構造

相互作用

化合物

糖鎖

疾患

バリエーション

TOGO DX

About Help Example Contact

> Attribute keys

Attribute values

Log scale

Gene

> High-level expression

Cere Cere Brai Co Per Spl V Ly Spl Thymus Bone SKI Place ProC Testis 1,955 H-Musc Eso Splr Cl Liver Lu Kid Pit T A Pancreas

> Low-level expression

Cerebellum Thymus 178 Bone ma Placenta 234 Pros Testis 157 Muscle 174 Color Liver/ f Pancreas 347

Protein

> Molecular mass

0-10 kDa 15,159 10-20 kDa 21,139 20-30 kDa 11,380 30-40 kDa 40-50 kDa 50-60 60-70 70-80 91 11 2

> PTMs

Phosphoprotein 9,266 Glycoprotein 6,457 Disulfide bond 5,728 Acetylation Ubl conj Isopep Lip Me C2

> # of transmembrane domains

0 65,969 1 6,178 2 34 3 17

Structure

> # of peptides in a PDB entry

1 peptide 140,890 2 peptides 20,364

> Other related molecules

Water 42,851 Zinc ic Sulfa Glyc 1,2- Maç Chl N A Ca Sc FC A I I I I Other molecules 27,510

Interaction

> Compounds in pathway

Metabolism 1,423 Disease 30 Signaling Metaboli Transpo Imm Ni G V C H D N D C I I

> # of interacting proteins

0 76,991 1 2 3 4 1

Compound

> Action type

Inhibitor 3,048 Antagonist 904 Agonist 824 Bind Blo M P H A I I I I

> Biological role

biochemical role 17,443 inhibitor 2 antimicrob pharma aet ep m x p b p

> Drug indication

Neoplasms Immune Patholo Infectio Digestiv Respira Nervou Skin an Cardio Femal Male I Hemi Endo Nutri Con Mus Ey S O C V

Glycan

> Molecular mass of glycans

20 400-600 600-800 800-1000 I 1000-1200 D 1200-1400 D 1400-1600 1600-18 1800-20 2000-2200 2200-2400 2400-2600 2600-2800 2800-3000 3000-3200 3200-3400 3400-3600 3600-3800 3800-4000 4000-4200 4200-4400 4400-4600 4600-4800 4800-5000 5000-5200 5200-5400 5400-5600 5600-5800 5800-6000 6000-6200 6200-6400 6400-6600 6600-6800 6800-7000 7000-7200 7200-7400 7400-7600 7600-7800 7800-8000 8000-8200 8200-8400 8400-8600 8600-8800 8800-9000 9000-9200 9200-9400 9400-9600 9600-9800 9800-10000 10000-10200 10200-10400 10400-10600 10600-10800 10800-11000 11000-11200 11200-11400 11400-11600 11600-11800 11800-12000 12000-12200 12200-12400 12400-12600 12600-12800 12800-13000 13000-13200 13200-13400 13400-13600 13600-13800 13800-14000 14000-14200 14200-14400 14400-14600 14600-14800 14800-15000 15000-15200 15200-15400 15400-15600 15600-15800 15800-16000 16000-16200 16200-16400 16400-16600 16600-16800 16800-17000 17000-17200 17200-17400 17400-17600 17600-17800 17800-18000 18000-18200 18200-18400 18400-18600 18600-18800 18800-19000 19000-19200 19200-19400 19400-19600 19600-19800 19800-20000 20000-20200 20200-20400 20400-20600 20600-20800 20800-21000 21000-21200 21200-21400 21400-21600 21600-21800 21800-22000 22000-22200 22200-22400 22400-22600 22600-22800 22800-23000 23000-23200 23200-23400 23400-23600 23600-23800 23800-24000 24000-24200 24200-24400 24400-24600 24600-24800 24800-25000 25000-25200 25200-25400 25400-25600 25600-25800 25800-26000 26000-26200 26200-26400 26400-26600 26600-26800 26800-27000 27000-27200 27200-27400 27400-27600 27600-27800 27800-28000 28000-28200 28200-28400 28400-28600 28600-28800 28800-29000 29000-29200 29200-29400 29400-29600 29600-29800 29800-30000 30000-30200 30200-30400 30400-30600 30600-30800 30800-31000 31000-31200 31200-31400 31400-31600 31600-31800 31800-32000 32000-32200 32200-32400 32400-32600 32600-32800 32800-33000 33000-33200 33200-33400 33400-33600 33600-33800 33800-34000 34000-34200 34200-34400 34400-34600 34600-34800 34800-35000 35000-35200 35200-35400 35400-35600 35600-35800 35800-36000 36000-36200 36200-36400 36400-36600 36600-36800 36800-37000 37000-37200 37200-37400 37400-37600 37600-37800 37800-38000 38000-38200 38200-38400 38400-38600 38600-38800 38800-39000 39000-39200 39200-39400 39400-39600 39600-39800 39800-40000 40000-40200 40200-40400 40400-40600 40600-40800 40800-41000 41000-41200 41200-41400 41400-41600 41600-41800 41800-42000 42000-42200 42200-42400 42400-42600 42600-42800 42800-43000 43000-43200 43200-43400 43400-43600 43600-43800 43800-44000 44000-44200 44200-44400 44400-44600 44600-44800 44800-45000 45000-45200 45200-45400 45400-45600 45600-45800 45800-46000 46000-46200 46200-46400 46400-46600 46600-46800 46800-47000 47000-47200 47200-47400 47400-47600 47600-47800 47800-48000 48000-48200 48200-48400 48400-48600 48600-48800 48800-49000 49000-49200 49200-49400 49400-49600 49600-49800 49800-50000 50000-50200 50200-50400 50400-50600 50600-50800 50800-51000 51000-51200 51200-51400 51400-51600 51600-51800 51800-52000 52000-52200 52200-52400 52400-52600 52600-52800 52800-53000 53000-53200 53200-53400 53400-53600 53600-53800 53800-54000 54000-54200 54200-54400 54400-54600 54600-54800 54800-55000 55000-55200 55200-55400 55400-55600 55600-55800 55800-56000 56000-56200 56200-56400 56400-56600 56600-56800 56800-57000 57000-57200 57200-57400 57400-57600 57600-57800 57800-58000 58000-58200 58200-58400 58400-58600 58600-58800 58800-59000 59000-59200 59200-59400 59400-59600 59600-59800 59800-60000 60000-60200 60200-60400 60400-60600 60600-60800 60800-61000 61000-61200 61200-61400 61400-61600 61600-61800 61800-62000 62000-62200 62200-62400 62400-62600 62600-62800 62800-63000 63000-63200 63200-63400 63400-63600 63600-63800 63800-64000 64000-64200 64200-64400 64400-64600 64600-64800 64800-65000 65000-65200 65200-65400 65400-65600 65600-65800 65800-66000 66000-66200 66200-66400 66400-66600 66600-66800 66800-67000 67000-67200 67200-67400 67400-67600 67600-67800 67800-68000 68000-68200 68200-68400 68400-68600 68600-68800 68800-69000 69000-69200 69200-69400 69400-69600 69600-69800 69800-70000 70000-70200 70200-70400 70400-70600 70600-70800 70800-71000 71000-71200 71200-71400 71400-71600 71600-71800 71800-72000 72000-72200 72200-72400 72400-72600 72600-72800 72800-73000 73000-73200 73200-73400 73400-73600 73600-73800 73800-74000 74000-74200 74200-74400 74400-74600 74600-74800 74800-75000 75000-75200 75200-75400 75400-75600 75600-75800 75800-76000 76000-76200 76200-76400 76400-76600 76600-76800 76800-77000 77000-77200 77200-77400 77400-77600 77600-77800 77800-78000 78000-78200 78200-78400 78400-78600 78600-78800 78800-79000 79000-79200 79200-79400 79400-79600 79600-79800 79800-80000 80000-80200 80200-80400 80400-80600 80600-80800 80800-81000 81000-81200 81200-81400 81400-81600 81600-81800 81800-82000 82000-82200 82200-82400 82400-82600 82600-82800 82800-83000 83000-83200 83200-83400 83400-83600 83600-83800 83800-84000 84000-84200 84200-84400 84400-84600 84600-84800 84800-85000 85000-85200 85200-85400 85400-85600 85600-85800 85800-86000 86000-86200 86200-86400 86400-86600 86600-86800 86800-87000 87000-87200 87200-87400 87400-87600 87600-87800 87800-88000 88000-88200 88200-88400 88400-88600 88600-88800 88800-89000 89000-89200 89200-89400 89400-89600 89600-89800 89800-90000 90000-90200 90200-90400 90400-90600 90600-90800 90800-91000 91000-91200 91200-91400 91400-91600 91600-91800 91800-92000 92000-92200 92200-92400 92400-92600 92600-92800 92800-93000 93000-93200 93200-93400 93400-93600 93600-93800 93800-94000 94000-94200 94200-94400 94400-94600 94600-94800 94800-95000 95000-95200 95200-95400 95400-95600 95600-95800 95800-96000 96000-96200 96200-96400 96400-96600 96600-96800 96800-97000 97000-97200 97200-97400 97400-97600 97600-97800 97800-98000 98000-98200 98200-98400 98400-98600 98600-98800 98800-99000 99000-99200 99200-99400 99400-99600 99600-99800 99800-100000 100000-100200 100200-100400 100400-100600 100600-100800 100800-101000 101000-101200 101200-101400 101400-101600 101600-101800 101800-102000 102000-102200 102200-102400 102400-102600 102600-102800 102800-103000 103000-103200 103200-103400 103400-103600 103600-103800 103800-104000 104000-104200 104200-104400 104400-104600 104600-104800 104800-105000 105000-105200 105200-105400 105400-105600 105600-105800 105800-106000 106000-106200 106200-106400 106400-106600 106600-106800 106800-107000 107000-107200 107200-107400 107400-107600 107600-107800 107800-108000 108000-108200 108200-108400 108400-108600 108600-108800 108800-109000 109000-109200 109200-109400 109400-109600 109600-109800 109800-110000 110000-110200 110200-110400 110400-110600 110600-110800 110800-111000 111000-111200 111200-111400 111400-111600 111600-111800 111800-112000 112000-112200 112200-112400 112400-112600 112600-112800 112800-113000 113000-113200 113200-113400 113400-113600 113600-113800 113800-114000 114000-114200 114200-114400 114400-114600 114600-114800 114800-115000 115000-115200 115200-115400 115400-115600 115600-115800 115800-116000 116000-116200 116200-116400 116400-116600 116600-116800 116800-117000 117000-117200 117200-117400 117400-117600 117600-117800 117800-118000 118000-118200 118200-118400 118400-118600 118600-118800 118800-119000 119000-119200 119200-119400 119400-119600 119600-119800 119800-120000 120000-120200 120200-120400 120400-120600 120600-120800 120800-121000 121000-121200 121200-121400 121400-121600 121600-121800 121800-122000 122000-122200 122200-122400 122400-122600 122600-122800 122800-123000 123000-123200 123200-123400 123400-123600 123600-123800 123800-124000 124000-124200 124200-124400 124400-124600 124600-124800 124800-125000 125000-125200 125200-125400 125400-125600 125600-125800 125800-126000 126000-126200 126200-126400 126400-126600 126600-126800 126800-127000 127000-127200 127200-127400 127400-127600 127600-127800 127800-128000 128000-128200 128200-128400 128400-128600 128600-128800 128800-129000 129000-129200 129200-129400 129400-129600 129600-129800 129800-130000 130000-130200 130200-130400 130400-130600 130600-130800 130800-131000 131000-131200 131200-131400 131400-131600 131600-131800 131800-132000 132000-132200 132200-132400 132400-132600 132600-132800 132800-133000 133000-133200 133200-133400 133400-133600 133600-133800 133800-134000 134000-134200 134200-134400 134400-134600 134600-134800 134800-135000 135000-135200 135200-135400 135400-135600 135600-135800 135800-136000 136000-136200 136200-136400 136400-136600 136600-136800 136800-137000 137000-137200 137200-137400 137400-137600 137600-137800 137800-138000 138000-138200 138200-138400 138400-138600 138600-138800 138800-139000 139000-139200 139200-139400 139400-139600 139600-139800 139800-140000 140000-140200 140200-140400 140400-140600 140600-140800 140800-141000 141000-141200 141200-141400 141400-141600 141600-141800 141800-142000 142000-142200 142200-142400 142400-142600 142600-142800 142800-143000 143000-143200 143200-143400 143400-143600 143600-143800 143800-144000 144000-144200 144200-144400 144400-144600 144600-144800 144800-145000 145000-145200 145200-145400 145400-145600 145600-145800 145800-146000 146000-146200 146200-146400 146400-146600 146600-146800 146800-147000 147000-147200 147200-147400 147400-147600 147600-147800 147800-148000 148000-148200 148200-148400 148400-148600 148600-148800 148800-149000 149000-149200 149200-149400 149400-149600 149600-149800 149800-150000 150000-150200 150200-150400 150400-150600 150600-150800 150800-151000 151000-151200 151200-151400 151400-151600 151600-151800 151800-152000 152000-152200 152200-152400 152400-152600 152600-152800 152800-153000 153000-153200 153200-153400 153400-153600 153600-153800 153800-154000 154000-154200 154200-154400 154400-154600 154600-154800 154800-155000 155000-155200 155200-155400 155400-155600 155600-155800 155800-156000 156000-156200 156200-156400 156400-156600 156600-156800 156800-157000 157000-157200 157200-157400 157400-157600 157600-157800 157800-158000 158000-158200 158200-158400 158400-158600 158600-158800 158800-159000 159000-159200 159200-159400 159400-159600 159600-159800 159800-160000 160000-160200 160200-160400 160400-160600 160600-160800 160800-161000 161000-161200 161200-161400 161400-161600 161600-161800 161800-162000 162000-162200 162200-162400 162400-162600 162600-162800 162800-163000 163000-163200 163200-163400 163400-163600 163600-163800 163800-164000 164000-164200 164200-164400 164400-164600 164600-164800 164800-165000 165000-165200 165200-165400 165400-165600 165600-165800 165800-166000 166000-166200 166200-166400 166400-166600 166600-166800 166800-167000 167000-167200 167200-167400 167400-167600 167600-167800 167800-168000 168000-168200 168200-168400 168400-168600 168600-168800 168800-169000 169000-169200 169200-169400 169400-169600 169600-169800 169800-170000 170000-170200 170200-170400 170400-170600 170600-170800 170800-171000 171000-171200 171200-171400 171400-171600 171600-171800 171800-172000 172000-172200 172200-172400 172400-172600 172600-172800 172800-173000 173000-173200 173200-173400 173400-173600 173600-173800 173800-174000 174000-174200 174200-174400 174400-174600 174600-174800 174800-175000 175000-175200 175200-175400 175400-175600 175600-175800 175800-176000 176000-176200 176200-176400 176400-176600 176600-176800 176800-177000 177000-177200 177200-177400 177400-177600 177600-177800 177800-178000 178000-178200 178200-178400 178400-178600 178600-178800 178800-179000 179000-179200 179200-179400 179400-179600 179600-17

TogoDXデータ属性ごとの俯瞰

例えば

タンパク質分子の機能

結合	触媒反応	調節	輸送	転写制御	...
----	------	----	----	------	-----

各タンパク質と相互作用するタンパク質の数

0	1	2	3	..
---	---	---	---	----

病気の大分類

臓器系の機能障害	メンデル遺伝病	身体領域による疾患分類	発生過程の疾患	癌		...
----------	---------	-------------	---------	---	--	-----

TogoDXでは、各データ属性の俯瞰から、分類や分布の条件を自由に組み合わせて、データの探索を行うことができる。

TOGO DX		
	> Attribute keys	Attribute va
遺伝子	Gene	
	> High-level expression	Cere Cere Br
	> Low-level expression	Cerebellum
タンパク質	Protein	
	> Molecular mass	0-10 kDa
	> PTMs	Phosp
	> # of transmembrane domains	
タンパク質立体構造	Structure	
	> # of peptides in a PDB entry	
	> Other related molecules	
相互作用	Interaction	
	> Compounds in pathway	
	> # of interacting proteins	
化合物	Compound	
	> Action type	
	> Biological role	
	> Drug indication	Neoplasms
糖鎖	Glycan	
	> Molecular mass of glycans	200-600 6
	> Tissues	col
疾患	Disease	
	> Diseases in MeSH	Nervous Sys
	> Diseases in NANDO	
	> Cross referenced disease DBs in Mondo	UMLS
バリエーション	Variant	
	> GWAS	
	> Clinical significance	Uncert

TogoDX検索条件

- 細胞内局在として、細胞膜を選択

Cellular component	cellular anatomical entity 39,160		protein-containing com		Unclassified 38,572	
Description	Values	Total	Values	Total	Values	Total
Functional annotations (GO term)						
cellular anatomical entity	39,160	0	17,508	0	plasma membrane	6,589
protein-containing complex	9,147	0	13,506	0	organelle membrane	4,908
intrinsic component of membrane			12,534	0	plasma membrane region	1,302
cytosol			7,540	0	membrane microdomain	366
cytoplasm			7,038	0	outer membrane	36
nucleoplasm			6,333	0	phagosome assembly site membrane	20
extracellular region			2,673	0	annulate lamellae	6
cell projection			2,252	0	nuclear outer membrane-endoplasmic reticulum membrane network	2
cell junction			2,236	0		
extracellular space			1,925	0		
supramolecular complex			1,491	0		
membrane-enclosed lumen			1,396	0		
multivesicular body			1,240	0		

- **GTEx の組織別発現データで、肺と小腸を選択**

> Expressed in tissues	Low specificity 21,661	Testis 12,404	Brain - (Brain - Brain Brain Brain Pitui Brain Brain Brai Sple Bra Bra Bra Cel Bra Br Sk Sk Sn Lu Th Ns Es Li P K M V W O C C A A F N A C U S F A I I
------------------------	------------------------	---------------	--

- PDBで、立体構造既知のタンパク質を選択

> Structure data existence	Proteins without structure data 70,821	Proteins with structure data 29,179
----------------------------	--	-------------------------------------

- ChEMBLで、信頼性スコア9のものを選択

ChEMBL assay existence		No ChEMBL assay	
Conf-scr	Conf-	74,473	

Condition builder

Clear

 Select target dataset

UniProt



 Map your IDs optional

 Try

e.g. A0A0A0MQR4, P36888-1, P25054-1,
O75874, P42336, Q9Y6K1-1, P21860-1, 

Submit

Reset

 Add filters

Protein / Cellular component

plasma membrane 

Structure / Structure data existence

Proteins with structure data 

Interaction / ChEMBL assay existence

Conf-score 9: Direct single protein target ass 

Gene / Expressed in tissues

Lung 

Small Intestine - Terminal Ileum 

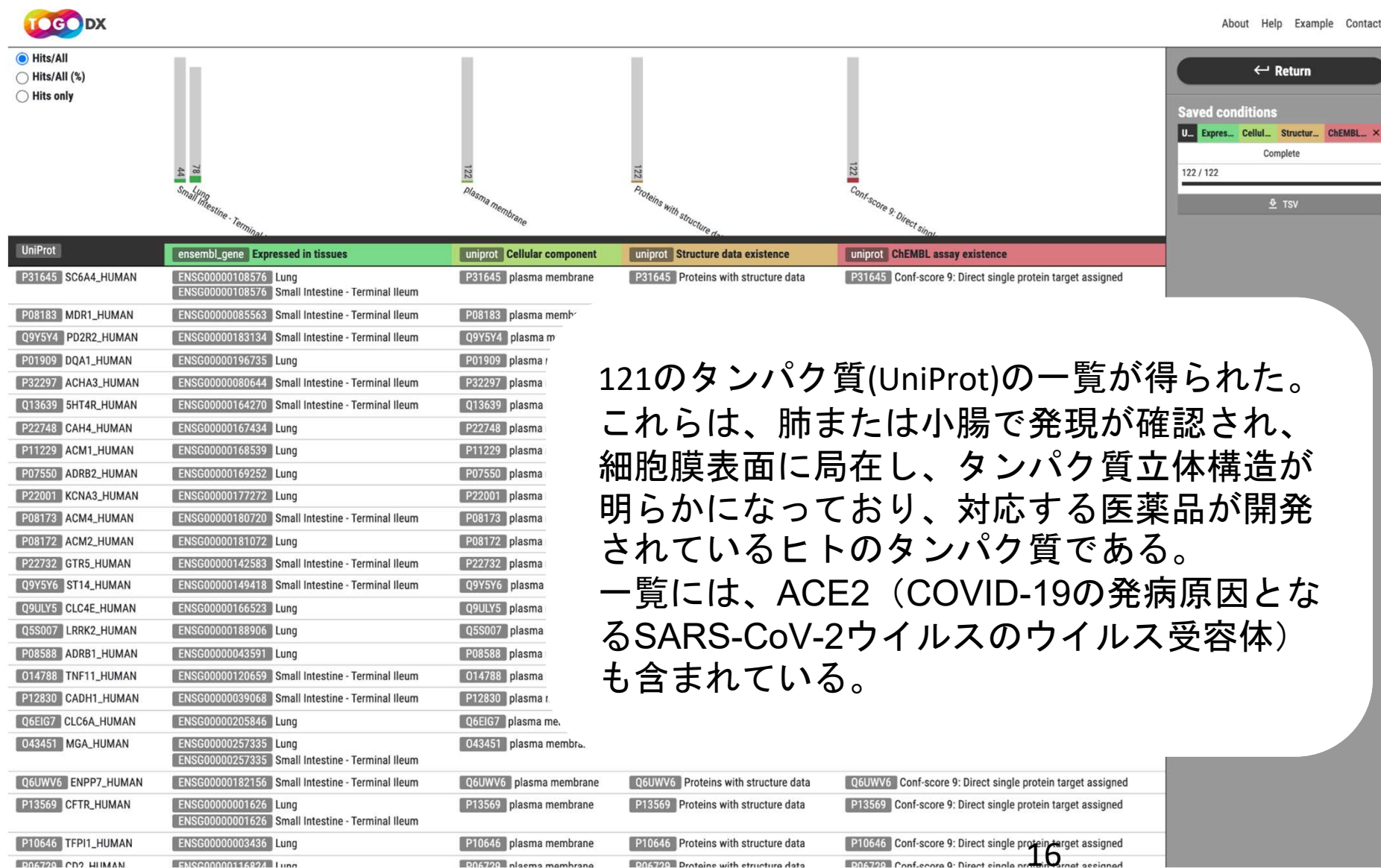
 Map attributes

 View results

これらの選択した条件を
全て満たすIDのリストを
検索する



TogoDX検索結果



121のタンパク質(UniProt)の一覧が得られた。これらは、肺または小腸で発現が確認され、細胞膜表面に局在し、タンパク質立体構造が明らかになっており、対応する医薬品が開発されているヒトのタンパク質である。一覧には、ACE2（COVID-19の発病原因となるSARS-CoV-2ウイルスのウイルス受容体）も含まれている。

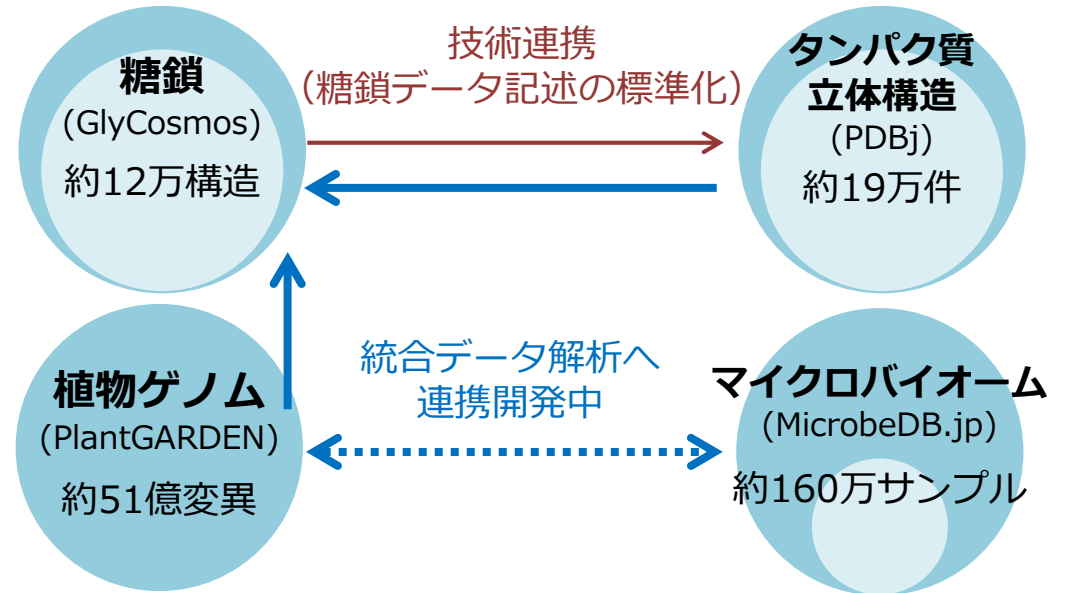
TogoDXとTogoIDの公開

- TogoDX 2021年10月5日公開
 - [https:// togodx.dbcls.jp/human](https://togodx.dbcls.jp/human)
 - 統合された知識グラフを俯瞰的に探索するためのウェブアプリケーション
 - ヒトに関する遺伝子、タンパク質、化合物、疾患などの情報を、国内外の21データベースから収集・統合
 - 現在約50種類の属性で探索可能
- TogoID 2021年7月8日公開
 - <https://togoid.dbcls.jp>
 - 生命科学分野におけるデータベース間で、IDの対応関係を検索および変換するためのウェブアプリケーション

統合化推進プログラム課題間での連携

RDF(知識グラフ)によるDB間連携

- ・ DBどうしの関連データの連携
(タンパク質構造ー糖鎖、
植物ー糖鎖)
- ・ 利用者連携によるホロゲノム
(植物ー微生物ゲノム) 統合
解析ツールを開発中



利用者のデータ解析の流れに沿ったデータ・DB機能の連携の例

- ・ DB検索結果を他
DBでの検索につ
なげる



統合化推進プログラム課題間連携の例

Differential Expression Analysis

① jPOSTでデータセット間を比較し、増減したタンパク質のリストを表示

Enrichment Analysis

別画面が起動

② エンリッチメント解析で「ChIP-Atlas」を選択すると増減した遺伝子リストが自動的にChIP-Atlasに入力

ChIP-Atlas - Enrichment Analysis

③ ChIP-Atlasで更に解析

(jPOST→ChIP-Atlas)
機能開発に加え、受け入れ可能なデータDBを増やす等で連携を実現

NBDC10年:できたこと、できなかったこと

- できたこと

- 欧米の真似ではない統合データベースの実現
- データ統合に関するベストプラクティス、ノウハウの蓄積
 - 複雑で多様でバラバラなデータと長年に渡る試行錯誤、悪戦格闘！
- 我が国の生命科学の基盤構築
 - バイオインフォマティクシヤン、実験研究者、データサイエンティスト
- DB人材の育成、コミュニティ形成
 - 基盤を支える人材、国際標準化

- できなかったこと

- 統合の果実を十分に味わい尽くすまでには至らなかった
- 本事業によるDBや技術を使ったデータ駆動型科学の成功事例の発信
- ヒト関連データ以外のデータベース統合
 - 統合化推進プログラムでの成果は種々あるが

- これからの課題、展開

- ヒト以外の育種、有用物質生産などへの展開
 - そのための種々のDBの知識グラフ化は進行中
- オープンデータとクローズドデータの混合解析
- 生命科学、バイオ産業のより広い分野への適用
- 生命のシステムの理解へ
 - 生命科学のデータ駆動型科学化への貢献だけでなく、その先へ
- 知識グラフ(RDF)は、汎用の仕組み→バイオを超えて

最後に

JST BIRD、ROIS DBCLS時代も含め
約20年の長きにわたり
統合データベースの開発に関わっていただいた
関係者の皆さまに深く感謝申し上げます