

創薬標的候補遺伝子選定・検証のためのツール 「遺伝子情報検索システム(GIS2)」の構築

43

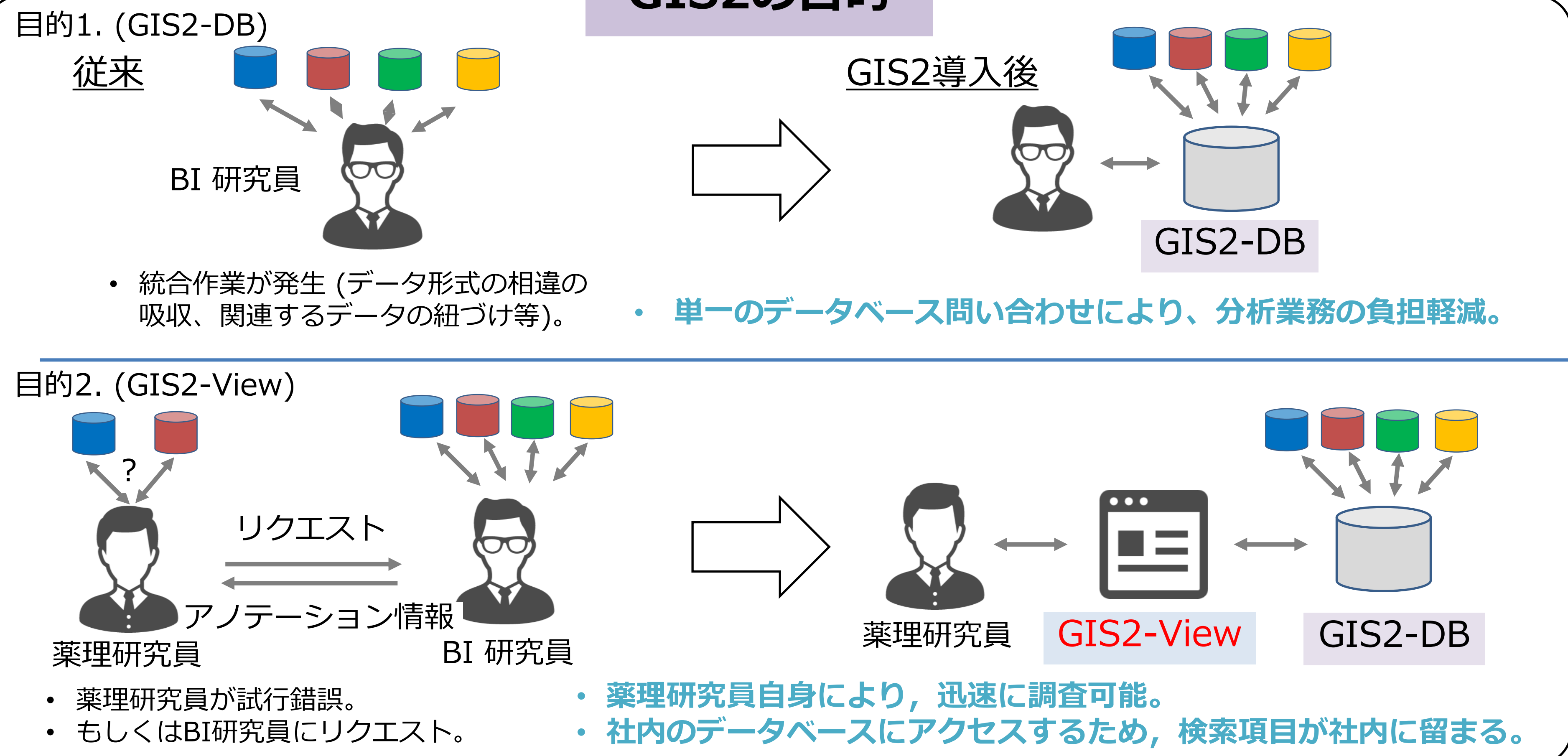
○池田 大祐、北内 信次、國吉 勇輝、伊波 大志、小櫻 裕司、沢田 啓、植松 直也 (大塚製薬)

要旨

製薬会社の創薬において、コンセプトに合致した標的遺伝子を選定することは、最初に直面する重要な課題である。これを目的として、各種オミクス解析や文献検索等により、数十・数百の候補遺伝子で構成されたリストを得ることがある。このようなリスト上の遺伝子を一つ一つ検証する過程は膨大な時間がかかる。本プロセスを簡便に行えるよう支援するため、「遺伝子情報検索システム(GIS: Gene Information Search system)」と命名したシステムを構築した。社内的には開発第2期にあたり、GIS2とした。

GIS2はそのIT基盤として、公共データベースの集合体であるGIS2-DBを有する。遺伝子に紐づく各種公共データベースの情報が、社内AWS環境S3バケット内にParquet形式として格納、カタログ化されている。コードを書ける解析専門家には、遺伝子情報や文献情報解析のための豊富なリソースを、Amazon Athena等を介して提供している。一方で一般的な薬理研究員に対しては、可視化ツールGIS2-Viewを提供する。これまではTIBCO Spotfireで実装してきた。しかしコンテンツの増大とともに構成が複雑化してきた。そこで最近では、PythonのWebフレームワークの1つであるDashを活用して、可視化を再構築している。

GIS2の目的



GIS2-Viewの基本コンセプト

- public** (1) 公共データベース(*)上の
- gene** (2) 遺伝子に紐づく情報を,
- broad** (3) 多様な疾患領域において
- easy** (4) 簡便に調査できる。

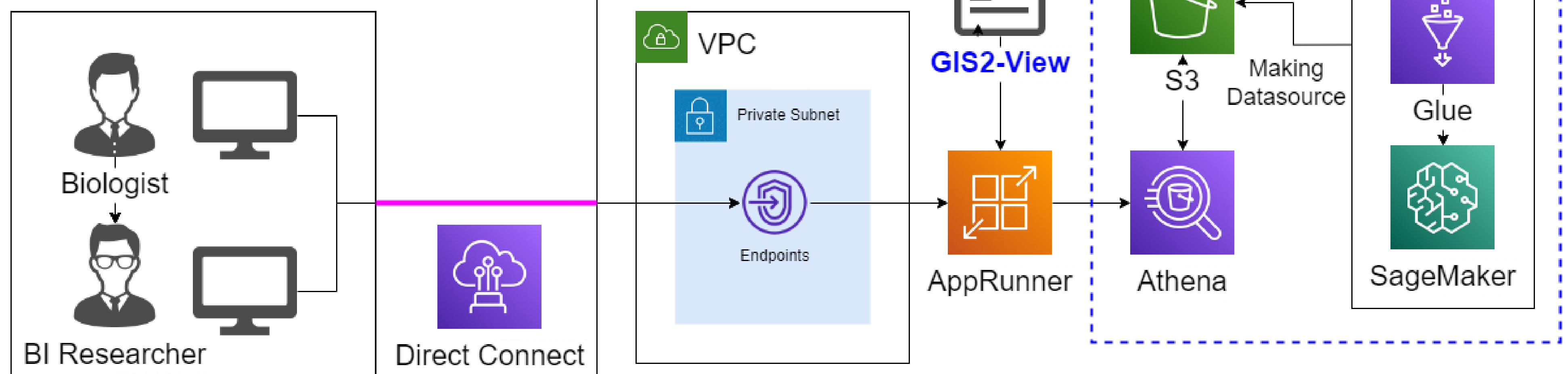
*「有害事象」のみ、商用DB OFF-X (Clarivate社)を活用。

主な参照元データベース

ページ	データベース
遺伝子	HGNC
発現	GTEX, Human Protein Atlas (HPA)
局在	Open Targets
分子機能	Open Targets, Gene Ontology
ドラッグビリティ	TargetDB, Cavity Space
薬剤	BindingDB
他種相同性	Ensembl Compara
疾患 (文献)	Open Targets, PubMed
治験	Open Targets

- BindingDB: <https://www.bindingdb.org/rwd/bind/index.jsp>
- CavitySpace: <http://www.pkumdl.cn:8000/cavityspace/>
- Ensembl: <https://useast.ensembl.org/index.html>
- GO: <https://geneontology.org/>
- GTEX: <https://gtexportal.org/home/>
- HGNC: <https://www.genenames.org/>
- HPA: <https://www.proteinatlas.org/>
- OpenTargets: <https://www.opentargets.org/>
- PubMed: <https://pubmed.ncbi.nlm.nih.gov/>
- TargetDB: <https://github.com/sdecresco/targetDB/>

ITアーキテクチャ



GIS2-Viewのページ構成と内容

遺伝子 (Gene) ID → トランスクリプトID / タンパク質ID → 遺伝子もしくはタンパク質と紐づけ

- 発現**: Exp. Tissue, Exp. SingleCell, Exp. ImmuneCell
- 局在**: Localization
- 分子機能**: GO
- 他種相同性**: Homology
- 化合物**: EffectiveCompound
- ドラッグビリティ**: Druggability (TDB), Druggability (CS)
- 疾患 (文献)**: Disease (literature)
- 治験**: ClinicalTrial
- 有害事象**: AdverseEvent

遺伝子 Gene

遺伝子リストBOX

#	GIS2ID	HGNC Symbol	Entrez GeneID	ENSGID	GeneName	Input	GeneType
1	2110	AVPR2	2110	ENSG0000013190	arginine vasopressin receptor 2	AVPR2	hssvm
2	4175	CDK4	4175	ENSG00000006728	cyclin dependent kinase 4	CDK4	hssvm
3	4175	CDK6	4175	ENSG00000040274	cyclin dependent kinase 6	CDK6	hssvm
4	7091	DRD2	7091	ENSG00000032259	dopamine receptor D2	DRD2	hssvm
5	7092	DRD3	7092	ENSG00000030849	dopamine receptor D3	DRD3	hssvm
6	9200	FGFR2	9200	ENSG00000054252	fibroblast growth factor receptor 2	FGFR2	hssvm
7	12418	HTR1A	12418	ENSG00000191001	5-hydroxytryptamine receptor 1A	HTR1A	hssvm
8	25100	HTR2A	25100	ENSG00000191002	5-hydroxytryptamine receptor 2A	HTR2A	hssvm
9	25100	PDE3A	25100	ENSG00000191003	phosphodiesterase 3A	PDE3A	hssvm
10	25102	PDE3B	25102	ENSG00000191004	phosphodiesterase 3B	PDE3B	hssvm

複数遺伝子を一括検索可能

他種相同性 Homology

Amino Acid Sequence Homology by Ensembl Compara

HGNC Symbol	Animal Symbol	Target Gene ID	Homology Type
AVPR2	Avpr2	ENSG00000003190	ortholog_oneZone
CDK4	Cdk4	ENSG00000006728	ortholog_oneZone
CDK6	Cdk6	ENSG00000040274	ortholog_oneZone
DRD2	Drd2	ENSG00000032259	ortholog_oneZone
DRD3	Drd3	ENSG00000030849	ortholog_oneZone
FGFR2	Fgfr2	ENSG00000054252	ortholog_oneZone
HTR1A	Htr1a	ENSG00000191001	ortholog_oneZone
HTR2A	Htr2a	ENSG00000191002	ortholog_oneZone
PDE3A	Pde3a	ENSG00000191003	ortholog_oneZone
PDE3B	Pde3b	ENSG00000191004	ortholog_oneZone

マウス、ラット、イヌ、マカクオソログとの相同性の程度をヒートマップ化

疾患 (文献) Disease (literature)

疾患リストBOX

疾患名を選択

エビデンスとなる文献について、発表年とエビデンスレベルを表示

発現 Exp. Tissue (組織発現)

Expression Specificities

発現量 (TPM)

化合物 EffectiveCompound

Compounds and Effective Doses

HGNC Symbol	UniProtID	Min(IC50)	Min(EC50)	Min(Ki)	Min(Kd)	Nch
AVPR2	P11802	0.007	0.05	0.36	1.38	915
CDK4	P00534	0.1	0.16	0.8020	63	
CDK6	P14416	0.0	0.02	0.0149	10257	
DRD2	P35462	0.0038	0.012	0.0258	5220	
DRD3	P21802	0.1	0.13	1.3	2.5	406
FGFR2	P22607	0.1	30	2.2	1.8	406
HTR1A	P08908	0.1	0.00021	0.00158	0.08	5183
HTR2A	P28223	0.025	0.27	0.01	0.097	4390
PDE3A	Q14432	0.11	10000	24	40	322
PDE3B	Q13370	0.049	6000	56	320	103

Distribution of Effective Doses: IC50, EC50, Ki, Kd

治験 ClinicalTrial

ClinicalStages1 (recruiting, completed)

HGNC Symbol	DrugType	Phase I	Phase II	Phase III	Phase IV
AVPR2	Protein	0	0	0	4
AVPR2	Small molecule	0	1	3	2
CDK4	Small molecule	4	4	0	4
CDK6	Small molecule	2	3	0	4
DRD2	Small molecule	2	10	13	53
DRD2	Unknown	0	1	0	0
DRD3	Small molecule	0	5	6	22
FGFR2	Antibody	1	1	0	0
FGFR2	Protein	0	0	2	0
FGFR2	Small molecule	6	8	5	2
FGFR2	Unknown	0	1	0	0
FGFR3	Small molecule	4	8	8	2
FGFR3	Unknown	0	1	0	0

ClinicalStages2 (terminated, withdrawn, etc.)

Exp. SingleCell (細胞発現)

Relative Expression Levels between Single Cells

遺伝子ごとの細胞の発現量 (TPM)

ドラッグビリティ Druggability (TDB)

Druggabilities & Pocket Score (by TargetDB)

HGNC Symbol	UniProtID	Druggability(TDB)	PocketScore	Druggability(CS)
AVPR2	P30518	0.919	0.288	
CDK4	P11802	0.985	0.934	
CDK6	Q00534	0.97	0.8	
DRD2	P14416	0.98	0.886	
DRD3	P35462	0.965	0.77	
FGFR2	P21802	0.95	0.848	
FGFR3	P22607	0.9	0.292	0.98
HTR1A	P08908			
HTR2A	P28223	0.998	0.679	
PDE3A	Q14432			
PDE3B	Q13370	0.992	0.825	

Relationships between Druggabilities and Pocket Scores

ドラッグビリティとポケットスコアに閾値設定し、超えた場合に「ドラッグダブル」と判定

謝辞 Lifematics

ライフマティックス社の方々には、DBに関する各種調査、GIS2-DBのPoC等にご協力いただきました。この場を借りて、感謝申し上げます。