

AIによるデータ駆動型研究が拓く 生命医科学と創薬

山西芳裕

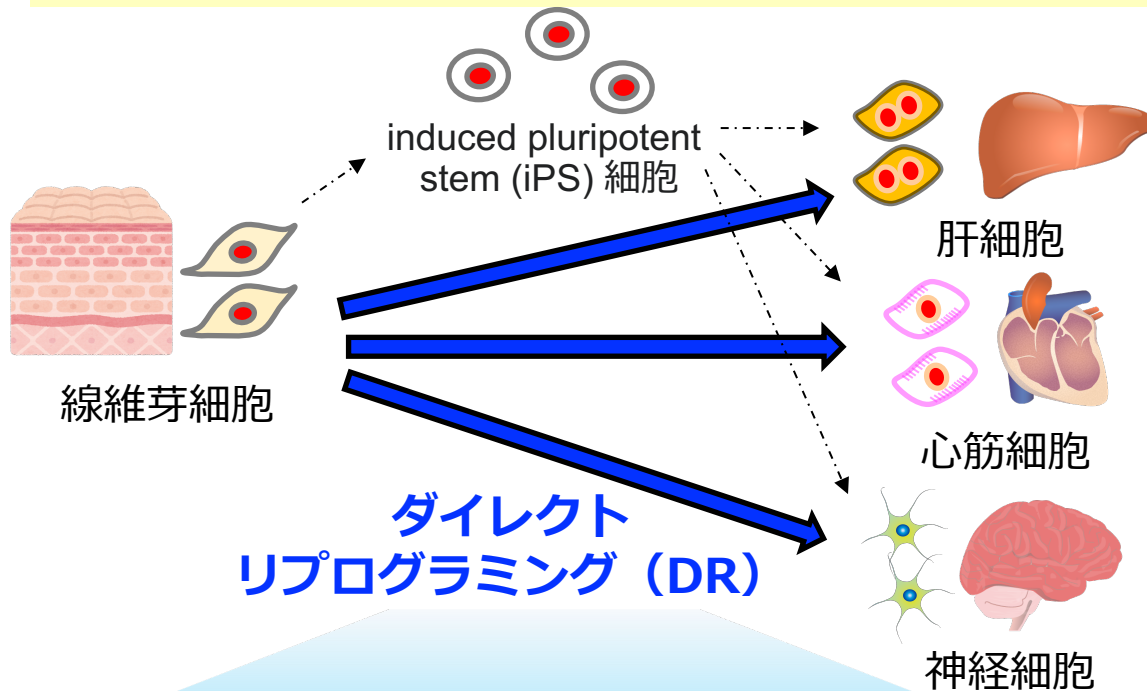
名古屋大学 大学院情報学研究科 複雑系科学専攻 生命情報論講座

内容

- 細胞直接変換
- 医薬品候補化合物予測
- シナジー創薬学

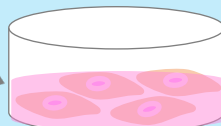
再生医療に向けたデータ駆動型ダイレクトリプログラミング

ダイレクトリプログラミング (DR) :
体細胞をiPS細胞を介さずに臓器の機能を持つ細胞へ直接変換



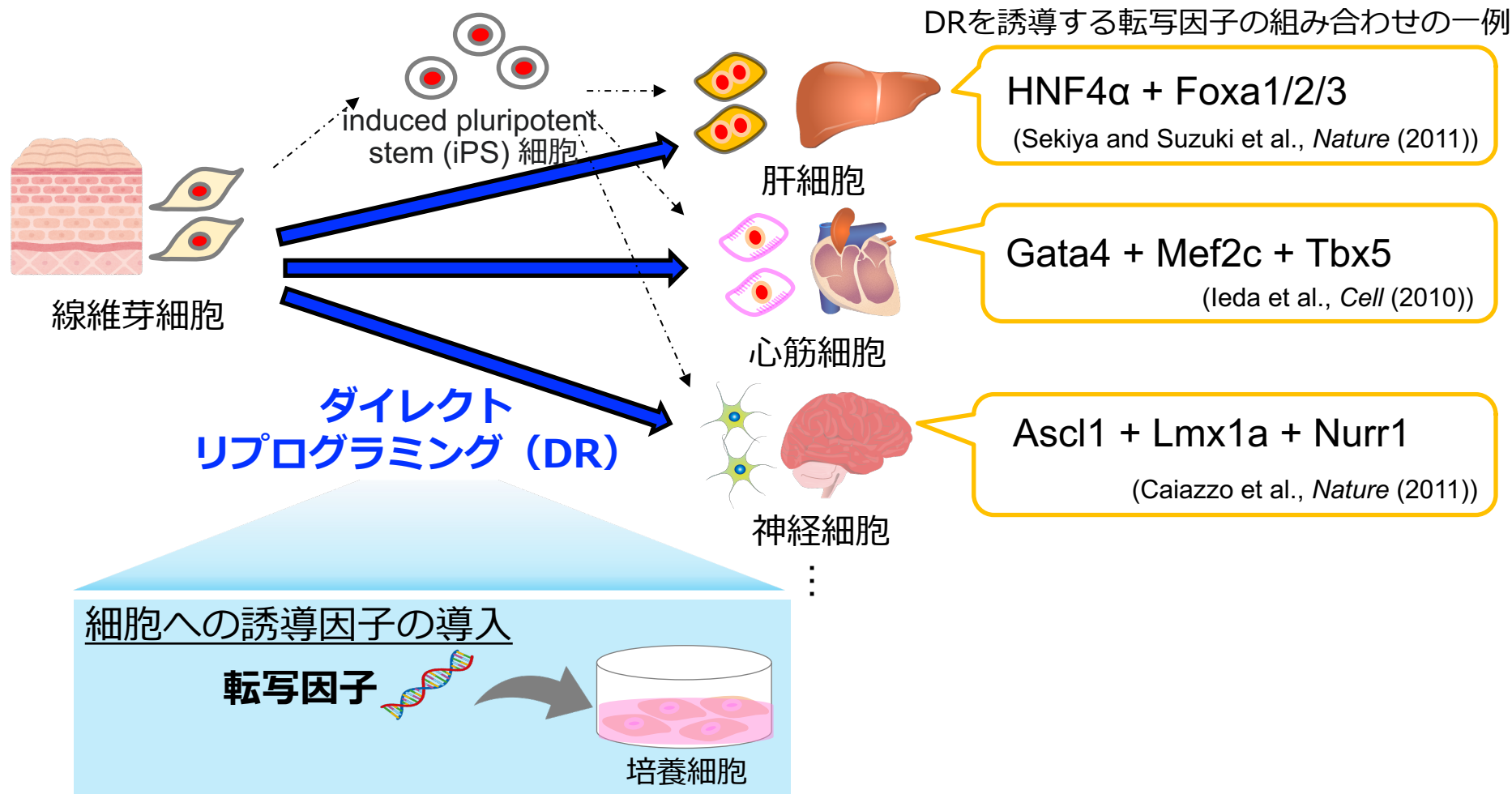
細胞への誘導因子の導入

転写因子



培養細胞

転写因子(DNAに結合するタンパク質)の組み合わせが細胞変換に重要



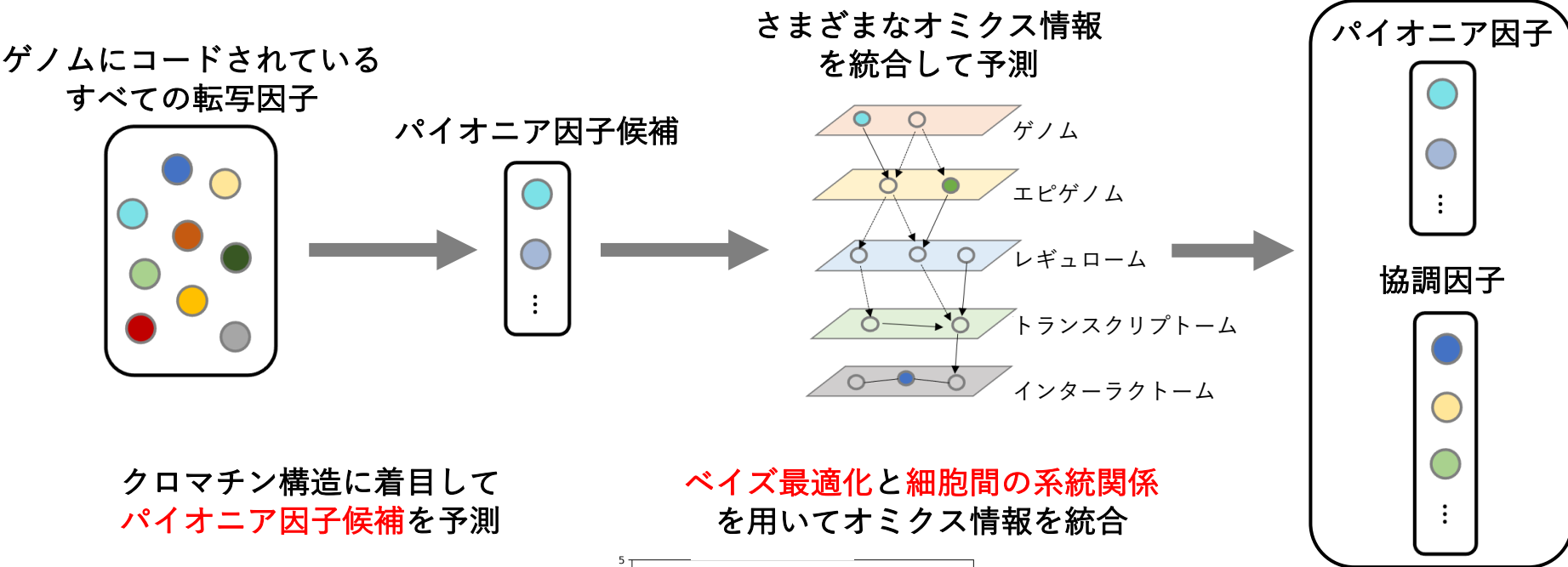
実験のみでDR誘導転写因子を同定することは困難

多階層オミクスデータから細胞変換に必要な 転写因子セットを予測する手法を提案

TRANSDIRE (TRANS-omics-based approach for Direct REprogramming)を開発

(Eguchi et al, *Bioinformatics*, 38(10), 2839-2846, 2022)

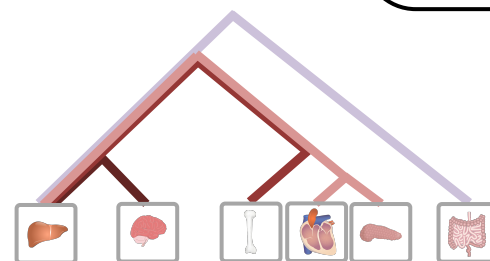
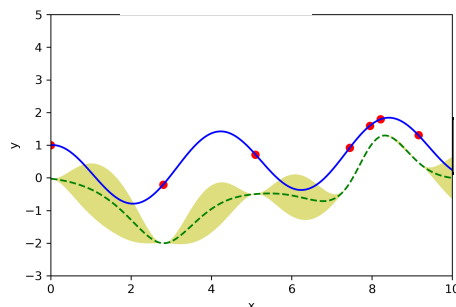
ダイレクトリプログラミングを
誘導する転写因子候補



クロマチン構造に着目して
パイオニア因子候補を予測



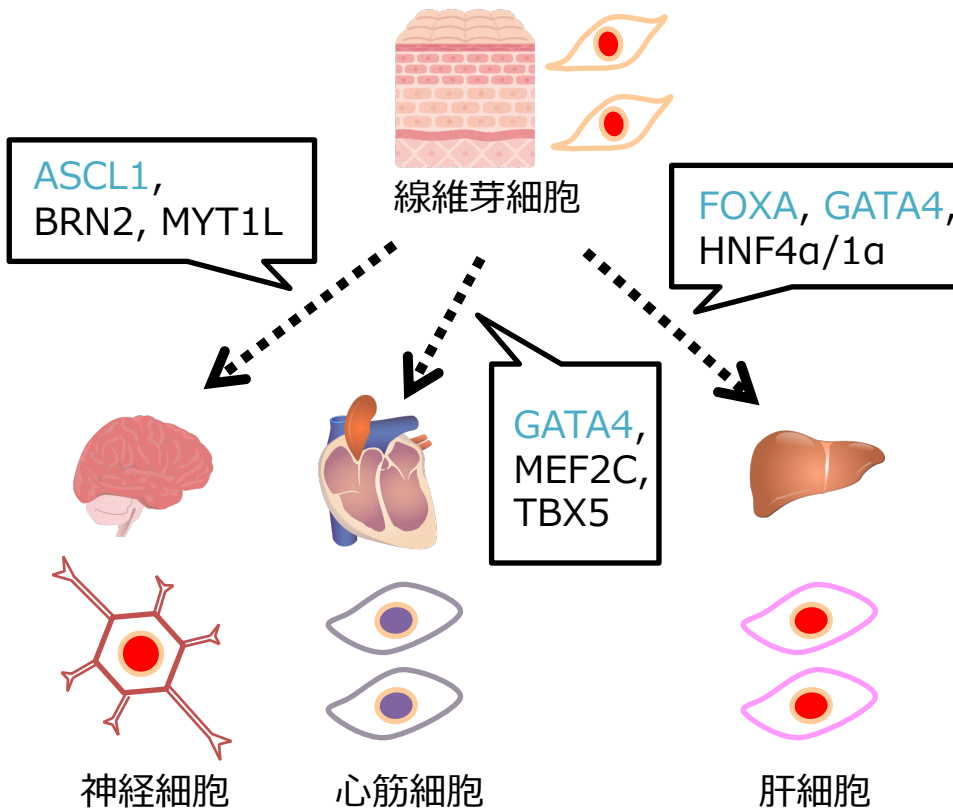
ベイズ最適化と細胞間の系統関係
を用いてオミクス情報を統合



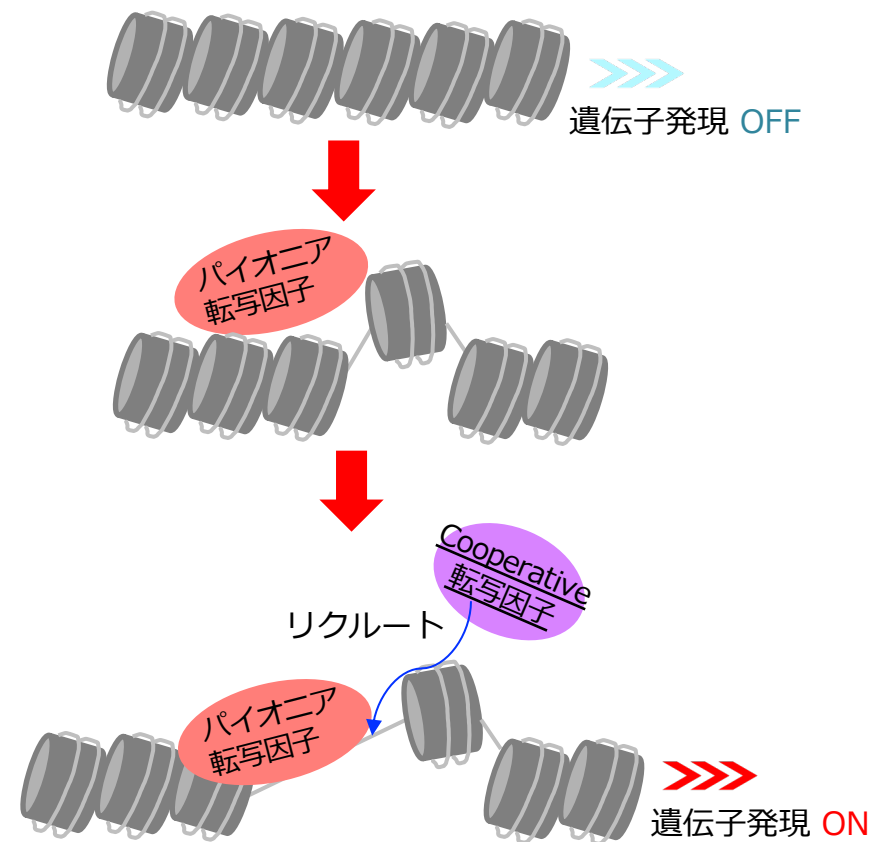
ダイレクトリプログラミングを誘導する転写因子の組み合わせにはパイオニア転写因子(PF)が含まれている

DRを誘導するのに必要な転写因子

(青字：パイオニア転写因子)



パイオニア転写因子の働き



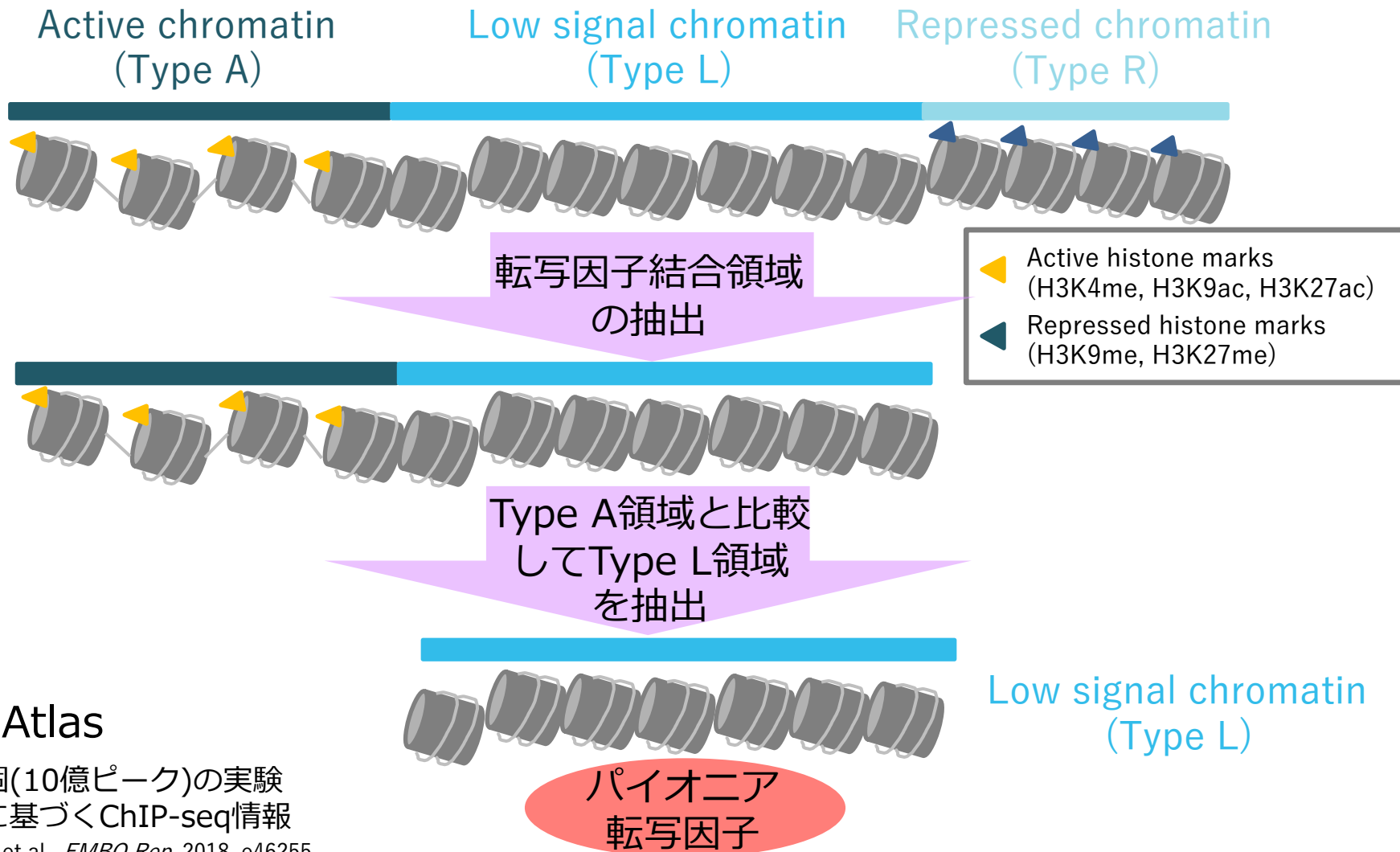
パイオニア転写因子：Pioneer factor (PF)

Horisawa et al., *Molecular Cell*, 2020, 79:660-676
Iwafuchi-Doi M and Zaret KS, *Development*, 2016, **143**:1833-1837

© 2016 DBCLS TogoTV

大規模なエピゲノムデータから パイオニア転写因子を予測する

- ① 各細胞種におけるPFが結合する領域 (Type L) を約 9 万件の大規模なChIP-seqから推定



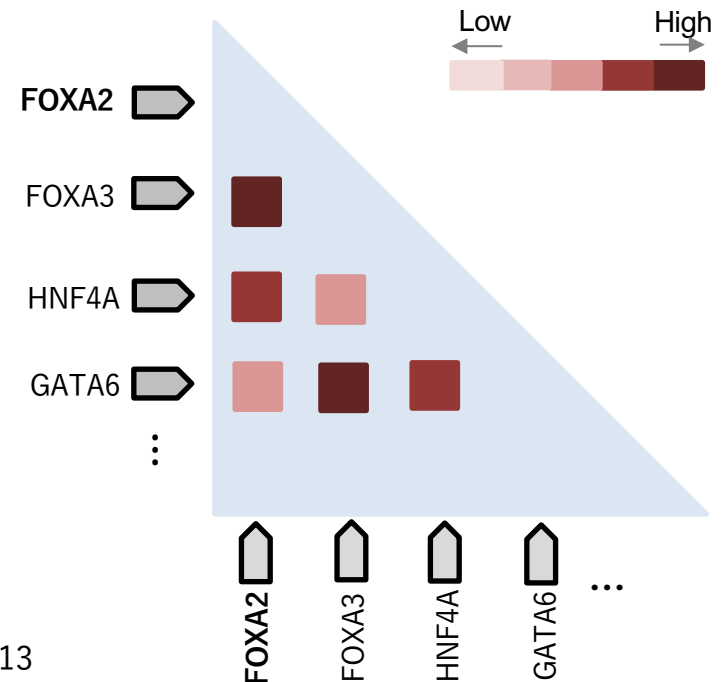
パイオニア転写因子と 共局在、共発現する協調転写因子を予測

(例) 予測したパイオニア転写因子：FOXA2

FOXA2と共局在する転写因子
のスコアを大規模なChIP-seq情報¹
を用いて算出

Cell Type	FOXA2's Colocalization partner	Score 					
			Exp. ID 1 	Exp. ID 2 	Exp. ID 3 	Exp. ID 4 	Exp. ID 5 
Hepatocyte	FOXA1						
Hepatocyte	FOXA1						
Hepatocyte	EP300						
Hepatocyte	SMARCC1						
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

FOXA2と共発現する転写因子の
スコアを大規模な分子間相互作用
情報²を用いて算出



1. ChIP-Atlas : Shinya Oki et al., *EMBO Rep.* 2018, e46255 .

2. STRING : Szklarczyk et al., *Nucleic Acids Res.*, 2019, **47**, D607-D613

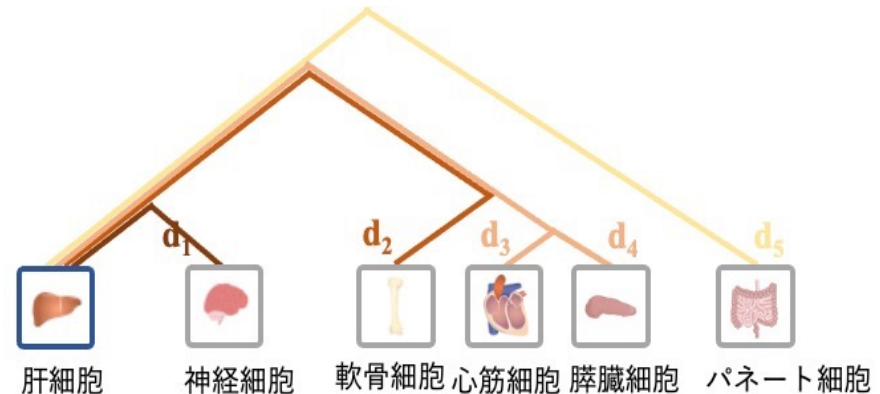
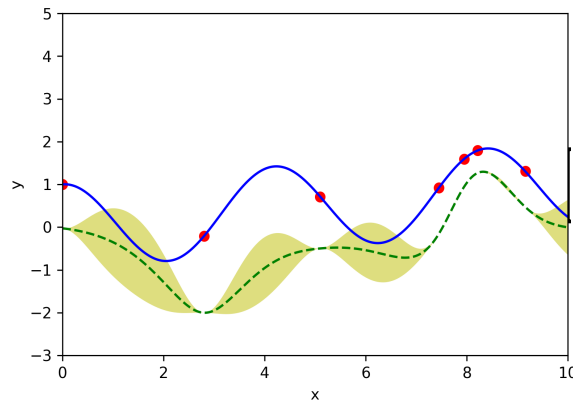
各オミクス情報から算出したスコアを融合した 予測手法を提案した

- ✓ Epigenomics method : S_{epi}
- ✓ Transcriptomics method : S_{tra}
- ✓ Regulomics method : S_{reg}
- ✓ Genomics method : S_{gen}
- ✓ Interactomics method : S_{int}

予測スコアの融合

$S_{\text{transomics}}$

ベイズ最適化と細胞間の系統関係を用いてマルチオミクス情報を統合



細胞変換に必要な転写因子のインシリコ予測

多階層オミクスデータの統合解析

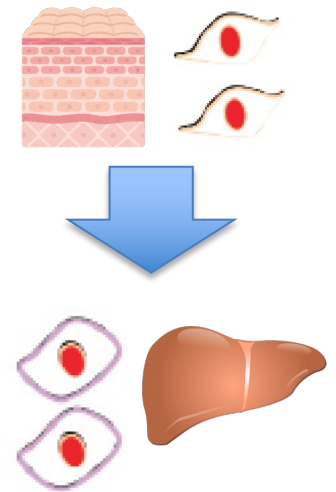
- 細胞特異的遺伝子発現データ
- 転写因子DNA結合データ
- エンハンサー領域データ
- 分子間相互作用ネットワークデータ

京都大学大学院 医学研究科
沖真弥先生との共同研究

(Eguchi et al, *Bioinformatics*, 38(10),
2839-2846, 2022)

例：皮膚細胞から肝細胞への変換を誘導する転写因子を予測

	TF	score
1	FOXA2*	41.77
2	PROX1	31.52
3	HNF4A	30.22
4	ATF5	29.95
5	SP1	29.85
6	CEBPA*	29.02
7	NR1H4	28.89
8	MLXIPL	27.96
9	HNF4G	27.83
10	CUX2	27.81



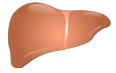
既報の転写因子
を再現

* Sekiya S, et al,
Nature(2011) 475:390-393

太字：DRを誘導することが報告されている転写因子

* パイオニア因子（PF）として予測された転写因子

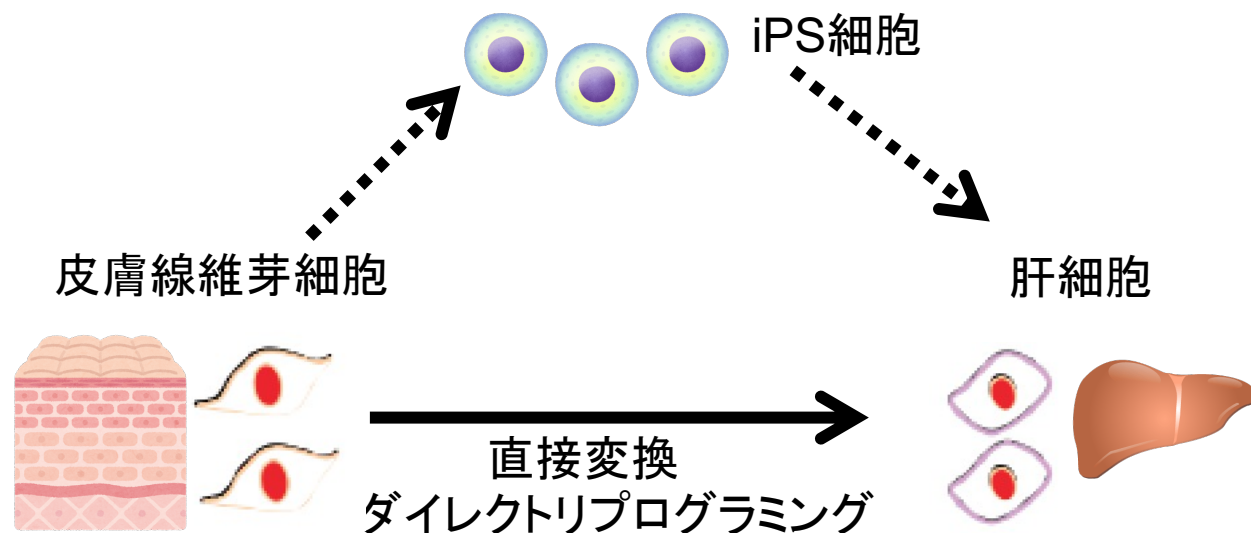
トランスオミクス解析により予測された ダイレクトリプログラミング (DR) 誘導転写因子

Source Target	Fibroblast Hepatocyte 		Fibroblast Cartilaginous cell 		Fibroblast Neuron 	
	TF	score	TF	score	TF	score
1	FOXA2*	41.77	CEBPA*	30.43	PAX6	36.09
2	PROX1	31.52	SPI1	27.05	MYT1L	35.92
3	HNF4A	30.22	FOXP1	24.49	ZIC1	34.44
4	ATF5	29.95	KMT2B	21.6	NKX2-2	33.86
5	SP1	29.85	LYL1*	20.18	NEUROD1*	30.7
6	CEBPA*	29.02	MYB	19.49	REST*	29.65
7	NR1H4	28.89	MEF2C	18.68	PEG3	28.73
8	MLXIPL	27.96	VDR	16.54	ARNT2	27.44
9	HNF4G	27.83	EPAS1	13.76	MEIS2	25.29
10	CUX2	27.81	FOXF1	13.56	SIX3	25.03
Source Target	Fibroblast Cardiomyocyte 		Fibroblast Pancreatic cell 		Fibroblast Paneth cell 	
	TF	score	TF	score	TF	score
1	ESRRA	51.27	NKX2-2	27.22	FOXA2*	22.91
2	HAND2	44.87	FOXA2*	24.94	MTF2	20.27
3	MEIS2	42.89	GATA4*	22.8	NR2F1	20.19
4	MEF2C	42.78	NR5A2	22.76	GATA2*	20.13
5	GATA4*	42.24	XBP1	20.45	PITX2	19.97
6	GATA2*	38.64	NEUROD1	20.09	MEIS2	19.32
7	ETV1	38.22	PAX6	18.15	FOXF1*	19.22
8	POU3F2	36.43	RBPJL	18.08	GATA4*	19.08
9	NKX2-5	35.62	POU3F2	17.92	SRF*	19.05
10	NR2F1	35.58	ISL1	16.65	FOXH1*	18.99

太字：DRを誘導することが報告されている転写因子

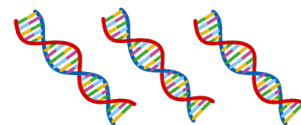
* PFとして予測された転写因子

低分子化合物によるデータ駆動型 ダイレクトリプログラミング (DR)



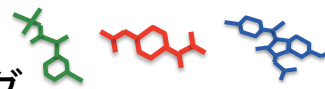
これまでの方法:

- 複数の転写因子の遺伝子導入による誘導
- 実験的な転写因子スクリーニング

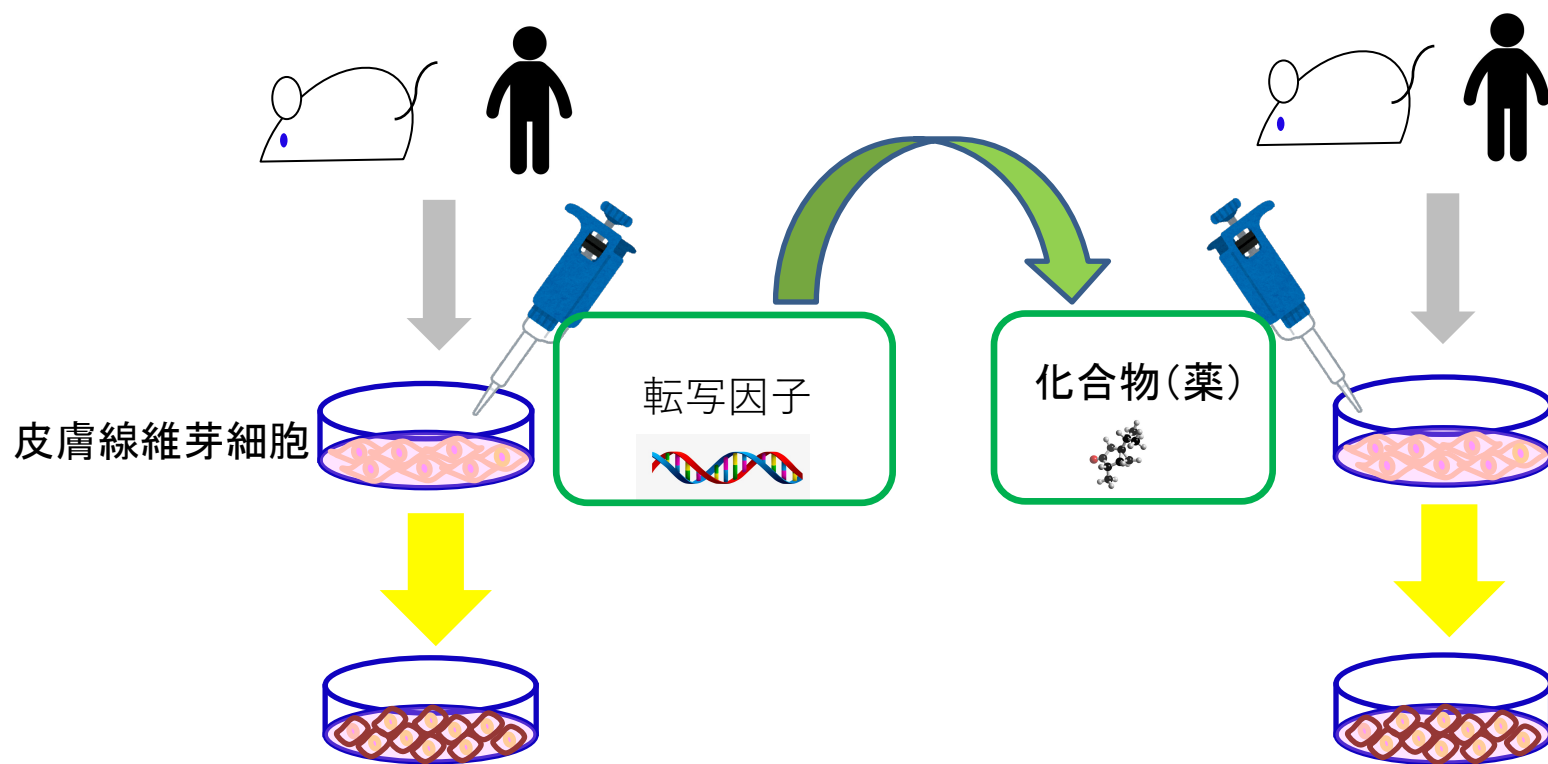


本研究の方法:

- 低分子化合物の組み合わせによる誘導
- 機械学習による低分子化合物スクリーニング



転写因子の導入を低分子化合物で置き換えたい

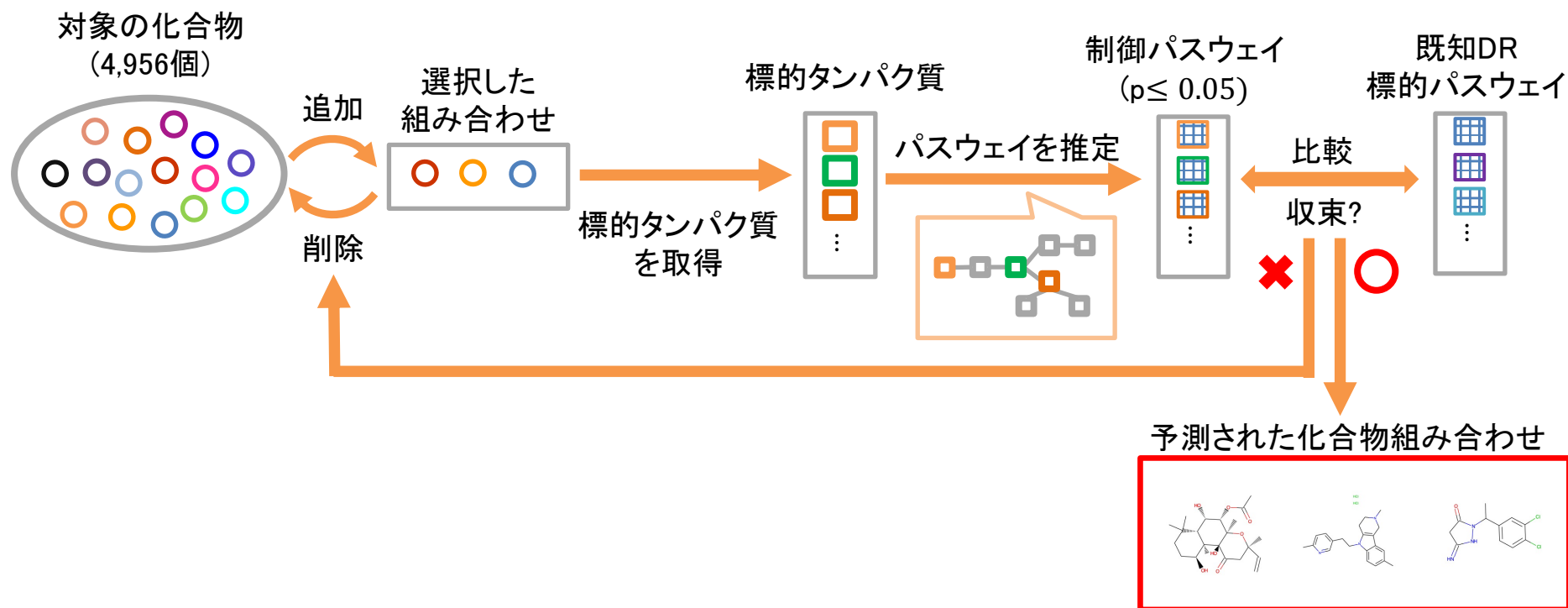


狙い:

- ウィルスに起因する発がんリスクの問題を回避する
- 生体内で目的の細胞を作成し、細胞移植を不要にする

DR関連パスウェイ情報を基に、化合物の新たな組み合わせをシミュレーテッドアニーリングで予測

(Nakamura et al, *Bioinformatics*, 2022;
Hamano et al, submitted)



- ダイレトリプログラミング (DR) 関連パスウェイを制御する化合物セットを予測した。
- 臨床応用を考慮し、対象の化合物は承認薬 (KEGG DRUG)のみとした。

心筋細胞を誘導する化合物組み合わせを予測

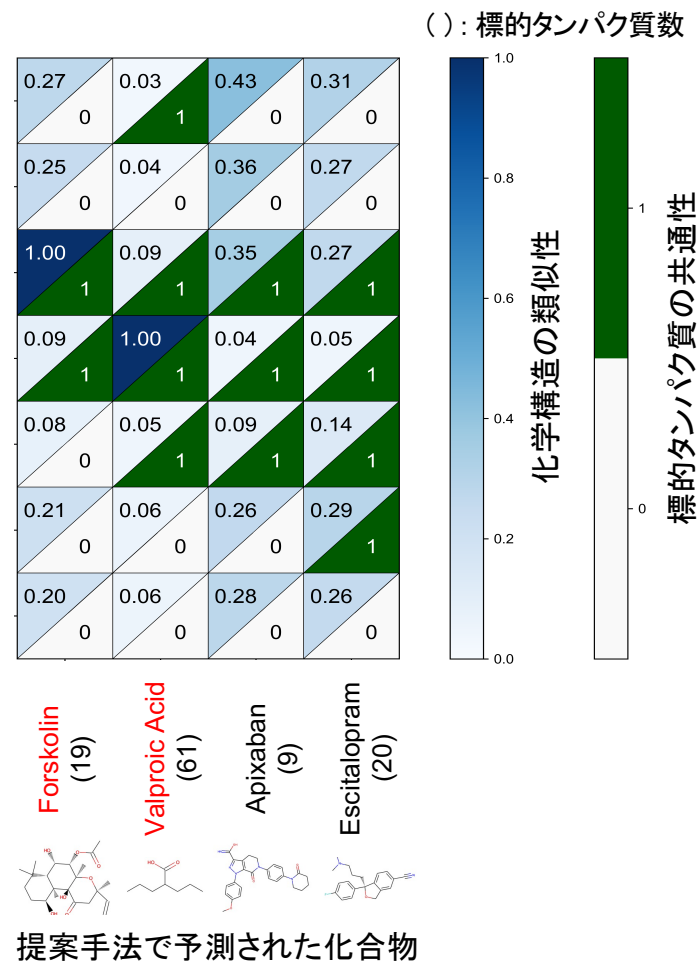
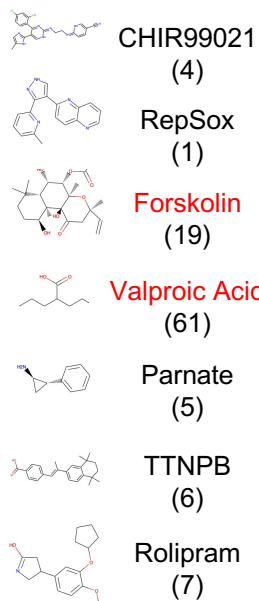


皮膚線維芽細胞 → 心筋細胞

既知の組み合わせ11番の詳細

標的タンパク質数	100
標的パスウェイ数	57
予測スコア	3.00

既知のDR誘導化合物



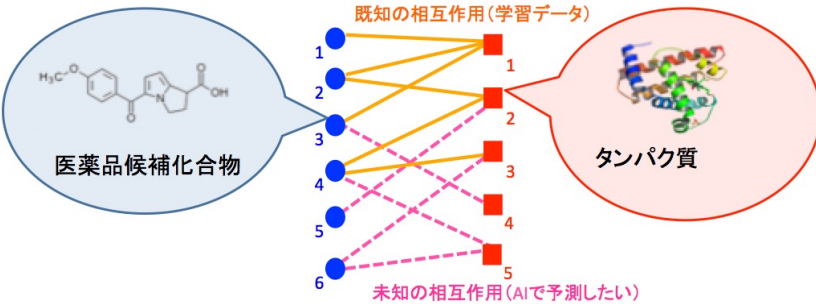
- それぞれの化合物の化学構造と標的タンパク質は関連していなかった
- 既知の5つの化合物をApixabanとEscitalopramで置き換えられる可能性がある

内容

- 細胞直接変換
- 医薬品候補化合物予測
- シナジー創薬学

ゲノムワイドに薬物（化合物）・タンパク質間相互作用を異種データから予測する手法を開発

薬物・タンパク質間相互作用予測を機械学習の視点から定式化



A pairwise model for any drug-protein pair $(\mathbf{x}', \mathbf{z}')$:

$$f(\mathbf{x}', \mathbf{z}') = \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} a_{ij} k((\mathbf{x}_i, \mathbf{z}_j), (\mathbf{x}', \mathbf{z}')) = \sum_{i=1}^{n_x} \sum_{j=1}^{n_z} a_{ij} \underbrace{k_x(\mathbf{x}_i, \mathbf{x}')}_{\text{化合物類似性}} \underbrace{k_z(\mathbf{z}_j, \mathbf{z}')}_{\text{タンパク質類似性}}$$

ペアワイズ学習

化合物・タンパク質
ペア

○ □

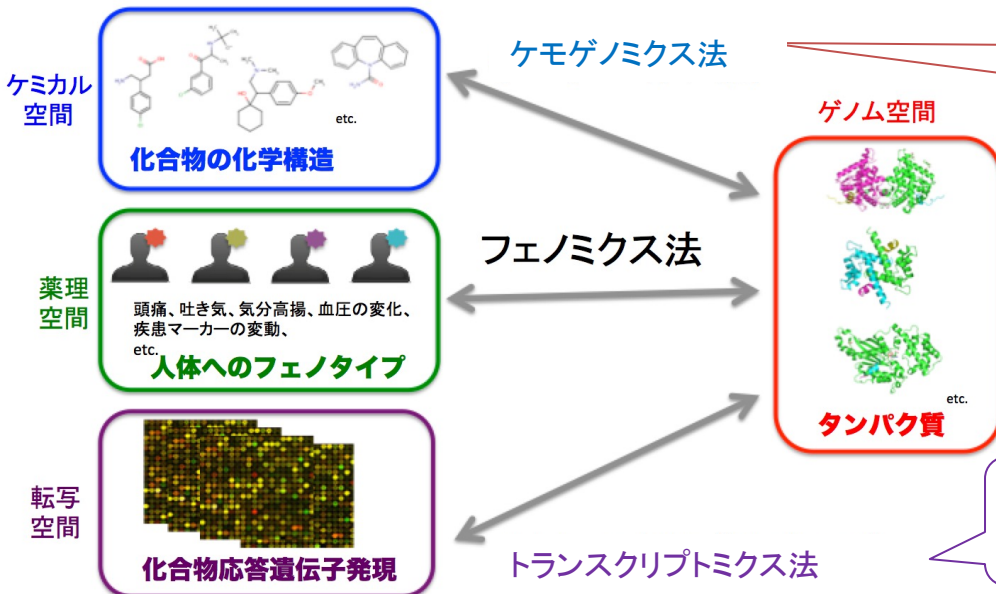
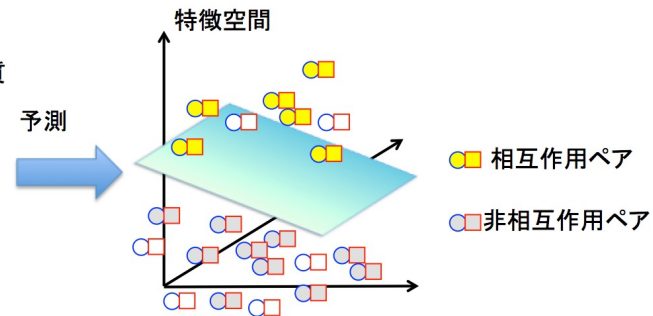
○ □

○ □

○ □

○ □

○ □



化合物の化学構造類似性から予測

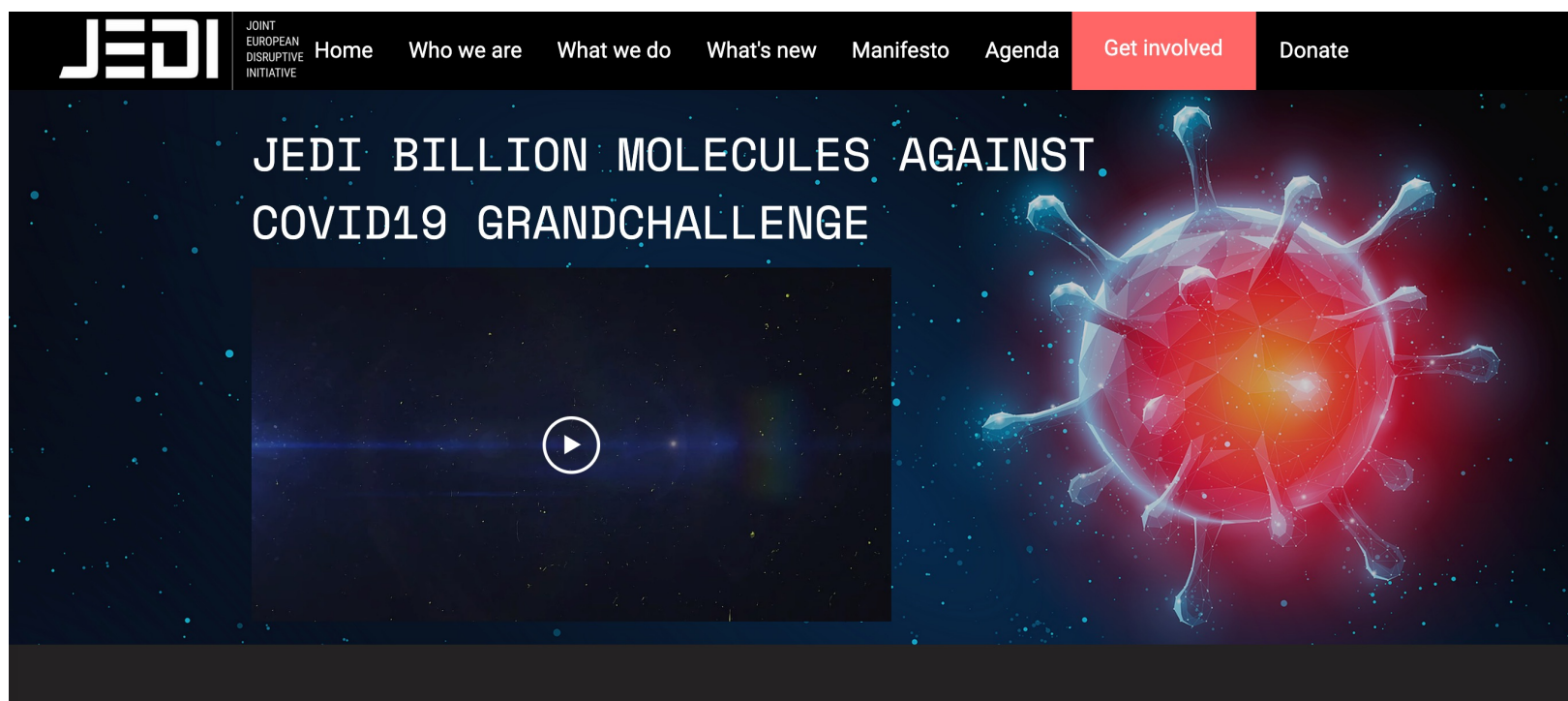
(Yamanishi et al, Bioinformatics, 2008; Yamanishi, Adv Neural Inf Process Syst, 2009; Bleakley et al., Bioinformatics, 2009; Tabei et al, BMC Systems Biology, 2013)

予測対象の化合物空間が大きい

ヒト由来細胞における化合物の遺伝子発現応用類似性から予測
(Hizukuri et al, BMC Med Genomics, 2015; Iwata et al, Sci rep, 2017; Sawada et al, Sci rep, 2018)
パスウェイの制御など作用機序の知見を得られる

新型コロナウイルス感染症への応用

COVID19に対する医薬品候補化合物の予測を行う国際プロジェクト
「JEDI Billion Molecules against Covid-19 GrandChallenge」



<https://www.jedi.foundation/billion-molecules>

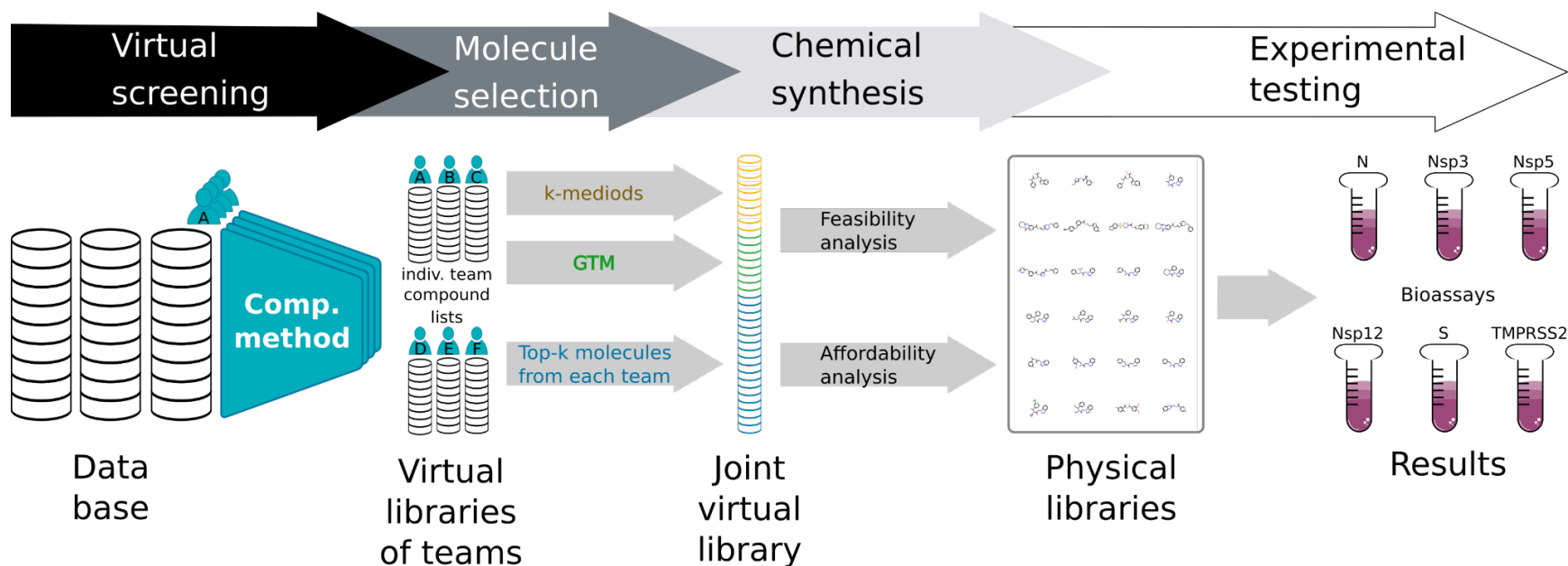
我々のチーム「KyuKen」も参画した。

多くの研究機関・大学・企業が参加

(世界中から130チーム、500人以上の科学者が参加)

Addis Ababa University, Adevinta, American University of Beirut, AnalysisMode, B.S Abdur Rahman Crescent Institute of Science and Technology, Bandung Institute of Technology, Broad Institute, Can-Fite BioPharma, Carleton University, Carnegie Mellon University, Celularity, CNRS, CompChem, Copenhagen University Hospital, Cyclica, Deeplab, DeepLife, Ecura Labs Pvt Ltd, ETH Zurich, Exponential Genomics (Xenomics), exscalate4cov, Federal University Gashua, Findimmune, Fondazione Ri.MED, Freie Universität Berlin, Georgetown University, GSI Technology, Harvard, Heidelberg University, IBM, Iktos, Inria / CNRS, institut J. Stefan, Institute of Biomedical Chemistry (IBMC), Istituto di Ricerche Farmacologiche Mario Negri, Jagiellonian University, Johannes Kepler University, Johns Hopkins, Jožef Stefan Institute, KISTI, KMD org, Korea Advanced Institute of Science and Technology (KAIST), KU Leuven, Kyushu Institute of Technology, L.J.Institute of Pharmacy, MABSilico, Medical University of Vienna, Merck KGaA, Mila, Muni University, N.D. Zelinsky Institute of Organic Chemistry, National Academy of Sciences of Belarus, National Cancer Institute (NIH), National Institute of Allergy and Infectious Diseases (NIAID), North-Caucasian State University, Northeastern University, Nyro Research India, NYU, Palacky University, Peptris, Johannes Kepler University Linz, Pirogov medical university, Prescience Insilico Private Limited, Quantiqun SRL, Rensselaer Polytechnic Institute, SAFAN Bioinformatics s.a.s, Sapienza University of Rome, Sathyabama Institute of Science and Technology, Skoltech, State University of New York, Swarnim Startup and Innovation University, Swat Youth Society, Synodis, The British University in Egypt, The University of Texas at El Paso, Thyenmed, Tokyo Institute of Technology, TU Chemnitz, TU Dresden, Tyndall National Institute, UL-FCCT, UltraNose, UNC-Chapel Hill, Universidad Industrial de Santander, Universidad Nacional Autónoma de México, Università degli Studi della Campania Luigi Vanvitelli, University Hospital Bonn, University of Abomey-Calavi, University of Birmingham, University of Bonn, University of Exeter, University of Ferrara, University of Foggia, University Of Ha'il, University of Illinois at Urbana-Champaign, University of Kentucky, University of Luxembourg, University of Michigan, University of Montana, University of Padova, University of South Florida, University of Strasbourg, University of the Punjab, University of Caen, Ural federal university, USAC, Weizmann Institute of Science, Yeungnam University.

JEDIプロジェクトの流れ



インシリコ予測した結果を、Pasteur InstituteやMIT-Broad Foundry
など様々な機関が実験検証

インシリコ予測した化合物を合成し、 実験的に有効性を確認できた化合物数

ヒット数は2位（130チーム中）

Teams contributing to hits

我々のチーム

HITS (28 total)

	covid19ddc	lambdazero	aiwinter	nuwave	way2drug	luxscreen	ai4science	JKU	belarus	cermn	deeplab	kyuken	lci	pharmai	sarswars	yoda	safan	sarstroopers	mlinch	virtualflow
predicted target																				
N	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Nsp3	0	0	0	0	1	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Nsp5	0	0	1	0	0	0	0	14	0	0	0	1	0	0	0	0	0	0	0	0
Nsp12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
S_ace2	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0
TMPRSS2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
29	2	0	1	0	1	0	1	14	0	1	1	5	0	0	0	0	0	0	2	1
hitrate vs synth (1 compound by 2 teams)	9.52	0.00	5.00	0.00	2.63	0.00	5.56	20.90	0.00	1.64	4.00	8.93	0.00	0.00	0.00	0.00	0.00	0.00	1.50	0.68
hitrate vs ranked after consensus	0.92	0.00	0.46	0.00	0.33	0.00	0.15	3.44	0.00	0.12	0.34	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.06

潜在的な抗がん作用を持つ薬物を発見

以下のような化合物を探索

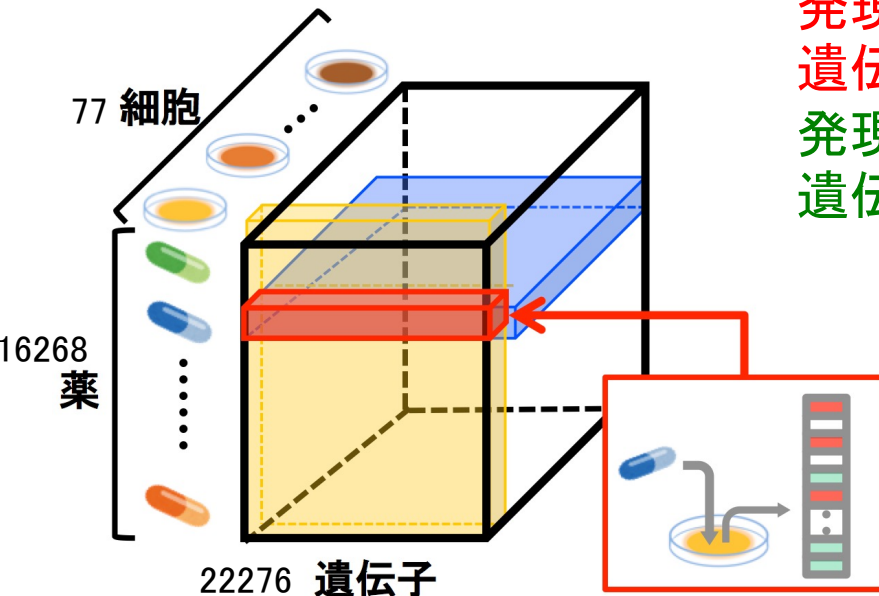
- 細胞周期パスウェイを不活性化
- p53シグナリングパスウェイを活性化
- アポトーシスパスウェイを活性化

東大・医科研

谷研究室との共同研究

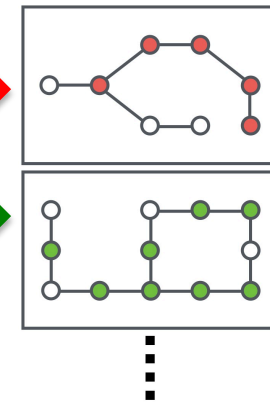
(Iwata et al, *J Med Chem*,
61, 9583–9595, 2018)

薬物応答遺伝子発現データ



発現が上がった
遺伝子

発現が下がった
遺伝子



活性化
パスウェイ

不活性化
パスウェイ

163 biological pathways

■ hypergeometric test

$$P\text{-value} = \sum_{i=z}^{\min(k, r)} \frac{\binom{k}{i} \binom{l-k}{r-i}}{\binom{l}{r}}$$

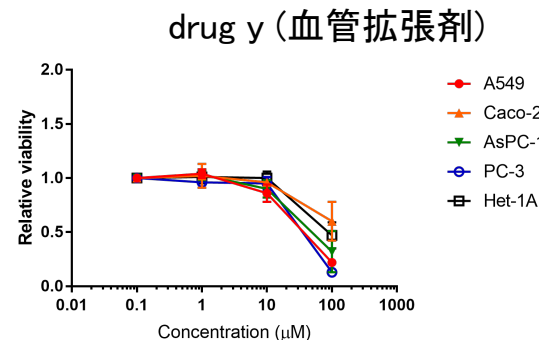
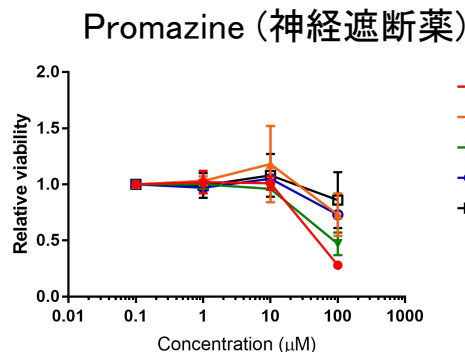
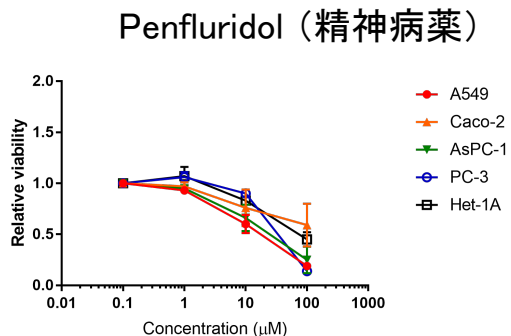
	Regulated genes	Genes
In a pathway	i	k
Not in a pathway	$r - i$	$l - k$
Total	r	l

抗がん剤以外の薬物から抗がん作用を予測し、実験検証(上位10個中、6個ヒット)

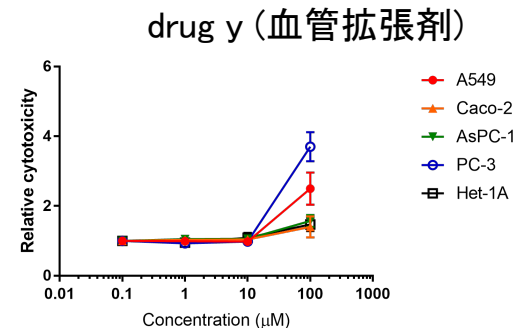
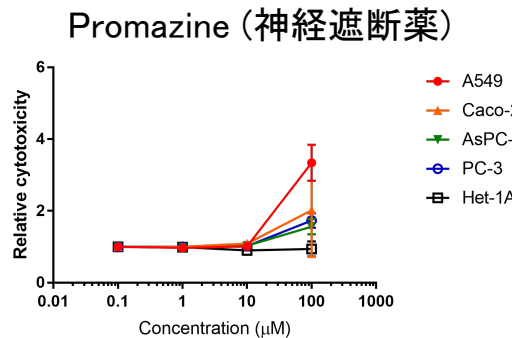
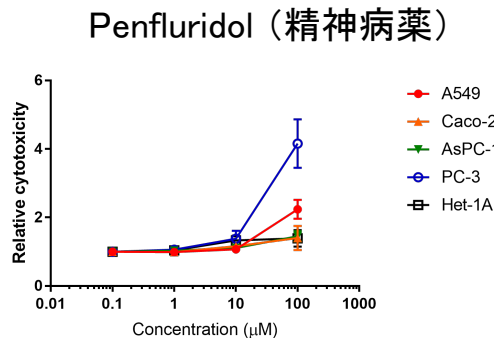
(Iwata et al, *Journal of Medicinal Chemistry*, 61, 9583–9595, 2018)

A549	肺がん
Caco-2	直腸がん
AsPC-1	膵臓がん
PC-3	前立腺がん
Het-1A	食道由来不死化細胞

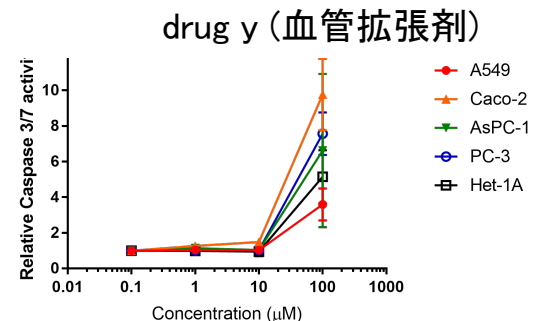
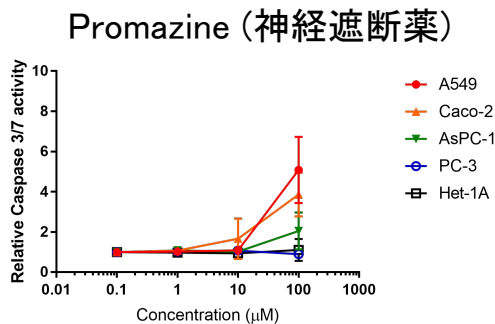
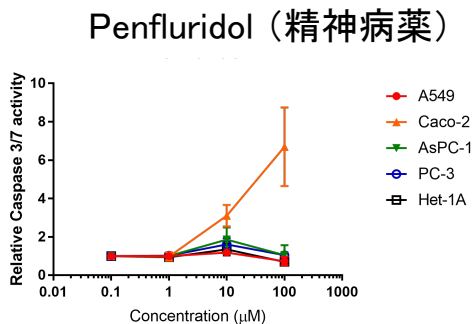
細胞生存性



細胞毒性

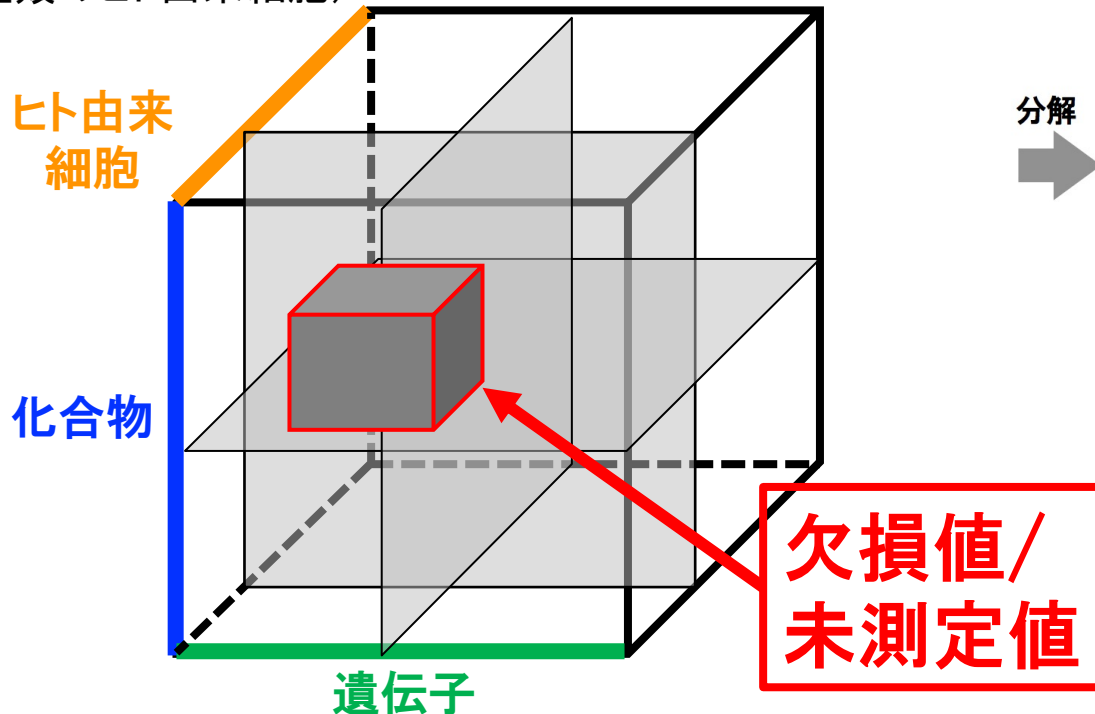


アポトーシス誘導能



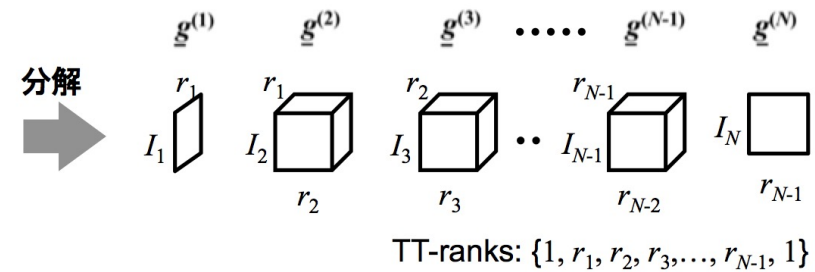
課題: オミクスデータには欠損値や未観測値が多い

化合物応答遺伝子発現データ
(16268種類の化合物、
77種類のヒト由来細胞)
(高階テンソル構造)



TensorTrainアルゴリズム

高階テンソル $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$



$$\underline{X} = \langle\langle \underline{g}^{(1)}, \underline{g}^{(2)}, \dots, \underline{g}^{(N)} \rangle\rangle$$
$$\underline{g}^{(n)} \in \mathbb{R}^{r_{n-1} \times I_n \times r_n}$$

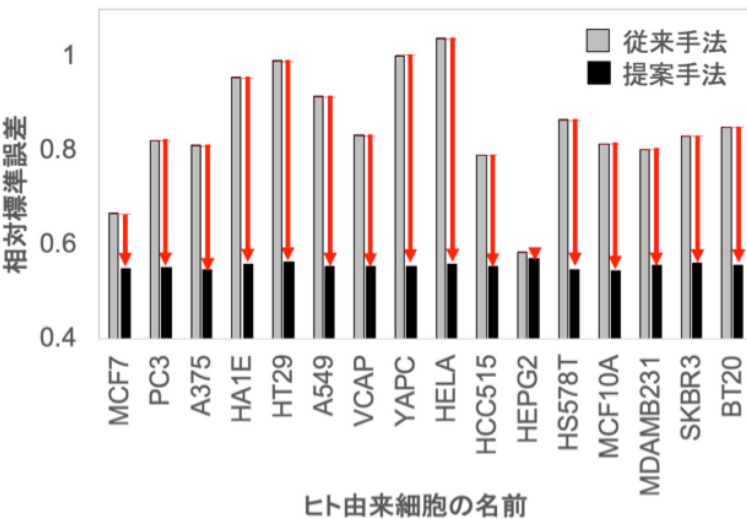
解決策: TensorTrainアルゴリズムによるテンソル分解で補完

提案手法は、データ補完精度を最大2倍改善させ、 疾患治療薬の予測性能の向上にも貢献

(Iwata et al, *Bioinformatics*, 35 i191–i199, 2019)

データ補完

さまざまな細胞を人工的に欠損させたときの
欠損値補完性能の評価結果

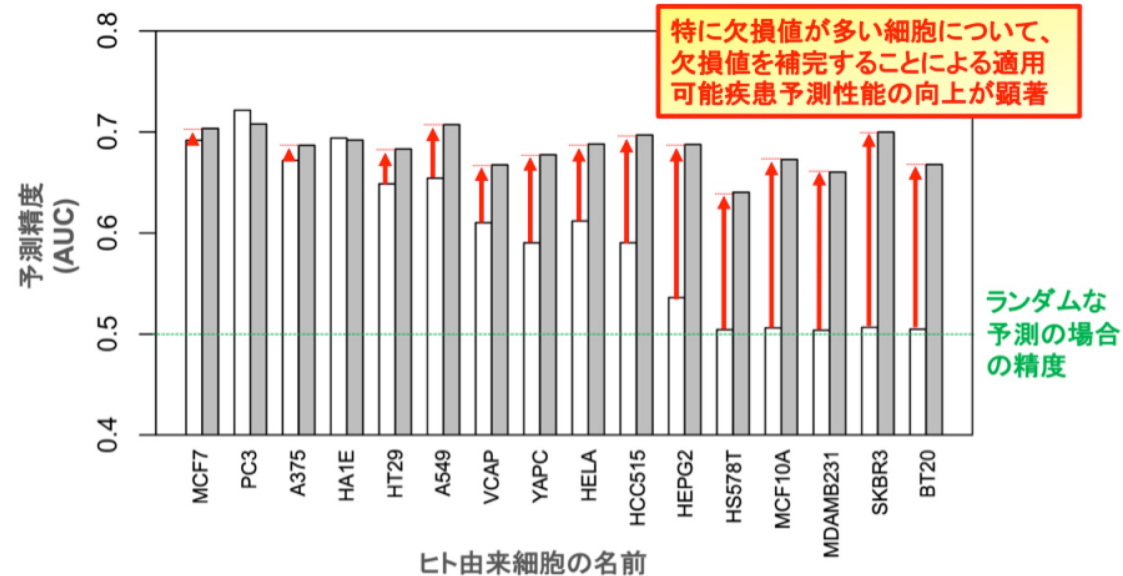


提案手法の誤差が小さい
= 提案手法の欠損値補完性能が高い

疾患治療薬の予測

薬物の適用可能疾患予測の精度評価

- 欠損値を含むデータを用いた場合
- 欠損値を予測したデータを用いた場合



特に欠損値が多い細胞について、
欠損値を補完することによる適用
可能疾患予測性能の向上が顕著

ランダムな
予測の場合の
精度

欠損値が少ない

欠損値の割合

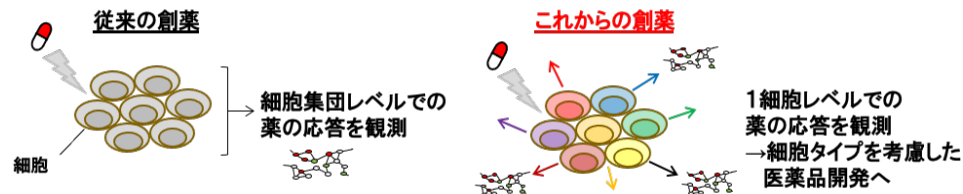
欠損値が多い

1細胞レベルの薬物応答遺伝子発現解析

(Iwata et al, *Nature Computational Science*, 2022)

1細胞レベルで薬の作用メカニズムを同定する新たな情報技術を開発 -テンソル補完アルゴリズムの創薬応用-

背景



これまでの課題

- ✓ 1細胞レベルでの遺伝子発現データは欠損値(未観測値)が多い

解決方法

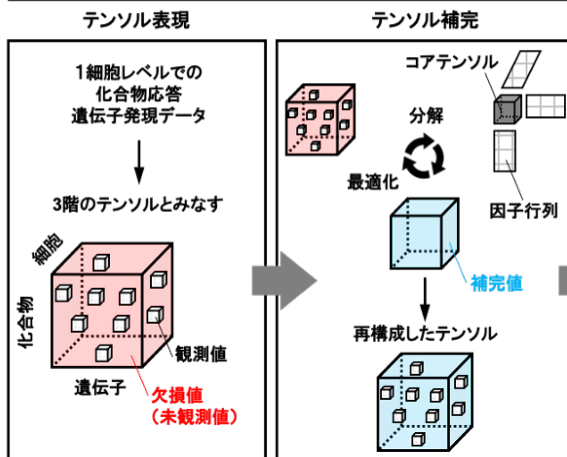
- ✓ テンソル補完技術を用いて欠損値(未観測値)を補完

※補完: 欠損値の値を補って、欠損値のない完全なデータを再構成すること

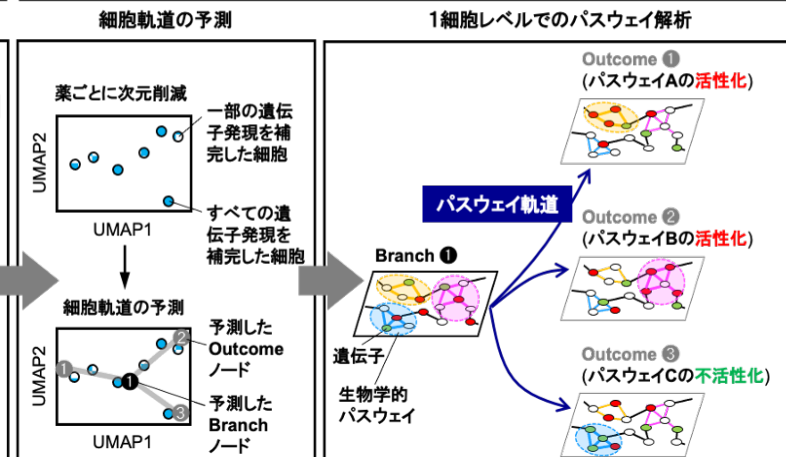
成果

- ✓ 1細胞レベルで薬の作用メカニズムを同定可能に

1細胞レベルでの化合物応答遺伝子発現データの予測



パスウェイ軌道解析

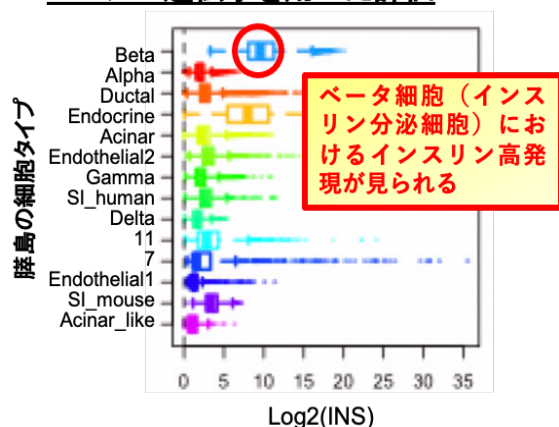


シングルセルレベルで薬物の作用機序を同定

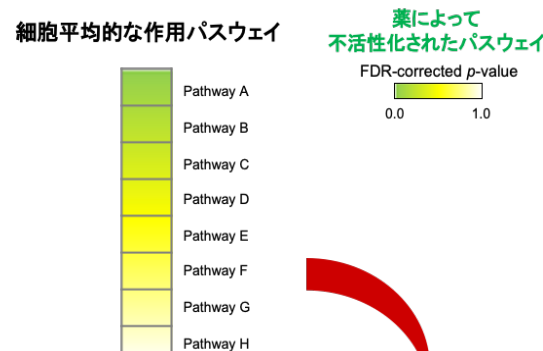
(Iwata et al, *Nature Computational Science*, 2022)

細胞のマーカー遺伝子に対して妥当な発現値、細胞タイプ特異的な遺伝子発現パターン、薬（例：アルテメテル）の潜在的な抗がん作用やその作用機序を予測できた

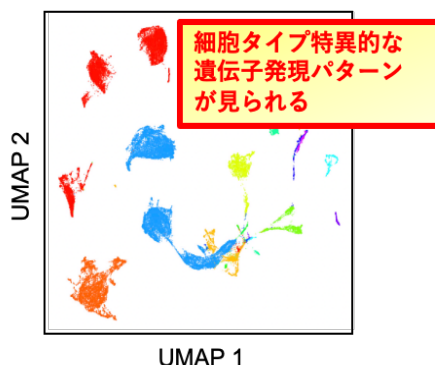
マーカー遺伝子を用いた評価



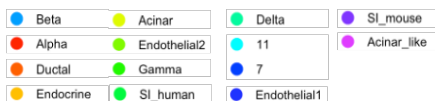
薬の作用パスウェイを用いた評価



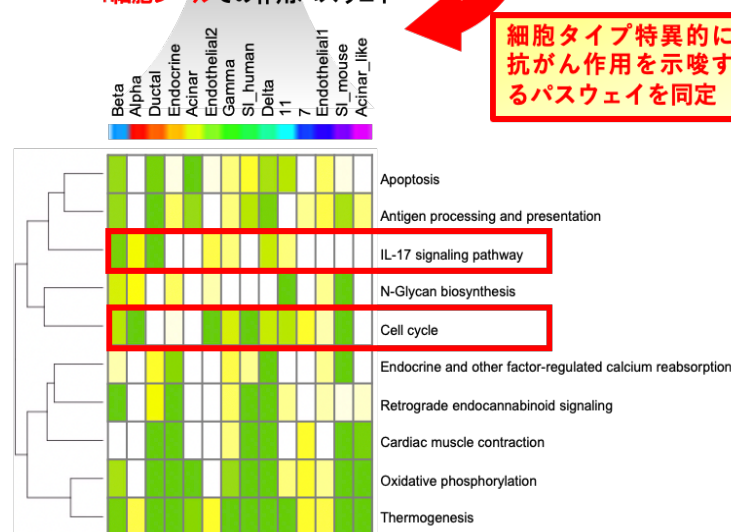
次元削減による評価



膵島の細胞タイプ



1細胞レベルでの作用パスウェイ



臨床試験

- 脊髄小脳失調症

- 運動失調を症状とする神経疾患。
- 小脳の神経細胞が徐々に破壊、消失していく難病。

久留米大学病院・神経内科の
三浦史郎先生との共同研究
(Miura et al, *Clin Case Rep*, 2023)

内容

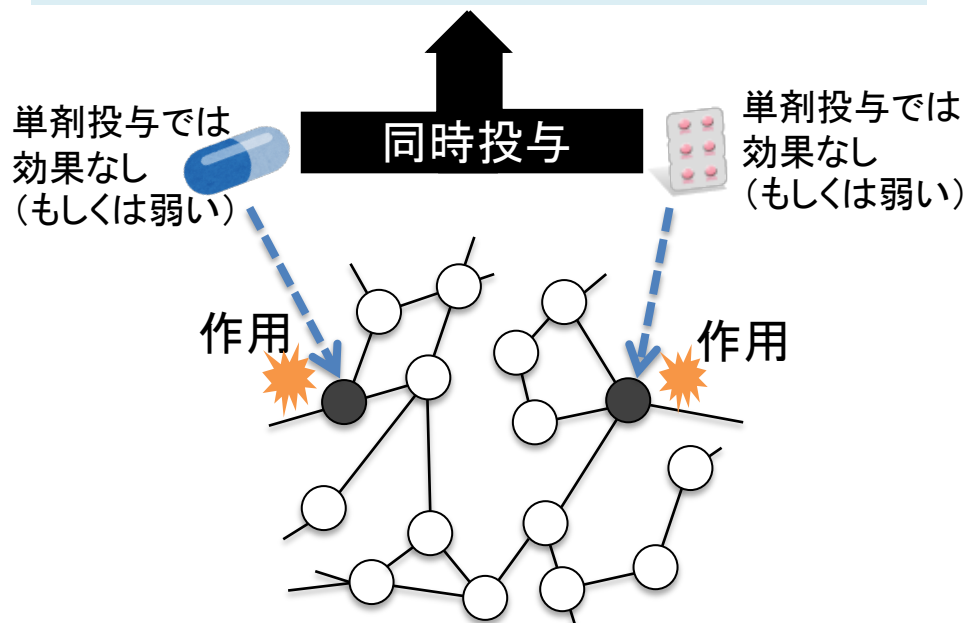
- 創薬標的探索
- 医薬品候補化合物予測
- シナジー創薬学

学術変革領域研究(B)「シナジー創薬学」

複数の薬剤・化合物の組み合わせによる相乗効果
(薬剤シナジー) の同定は生命医科学で重要課題

薬剤シナジー:

複数の薬剤の組み合わせによる相乗効果



○ 生体分子 — ネットワーク ● 治療標的

従来の研究・問題点

- 薬剤シナジーは偶発的に発見
- シナジー効果のメカニズムは不明
- 化合物の全ての組み合わせは実験検証不可能

本研究の提案

- 情報科学: AIによる高速・網羅的な予測
- 物質科学: 自由自在な化合物の設計と創製
- 生命科学: 広範な作用機序・制御機構の解明

情報科学、物質科学、生命科学の協奏
による「シナジー創薬学」の確立

問題: 最適な薬剤の組み合わせの同定は困難

まとめ

- 機械学習によるビッグデータ解析で、様々な細胞変換を誘導する転写因子や化合物の予測が可能。
- 様々な疾患に対する医薬品候補の探索や作用機序の同定が可能。
- 所望の性質を持つ医薬品候補の新しい化学構造を生成可能。

公共データベースをフル活用

ChIP-Atlas (沖真弥先生)

- 細胞直接変換
- 医薬品候補化合物予測
- シナジー創薬学

KEGG (金久實先生)

Disease/Drug/Pathway/Brite/Genes

公共データベースの課題と期待

- AIの学習セットになるので、データ駆動型研究に不可欠。
- 継続して維持できる枠組みが必要（消えてしまうデータベースが非常に多い）。データベースも過度な選択と集中は危険では？
- 個々のデータベースの場所はバラバラでもいいが、異なるデータベース間で同じ分子や表現型の情報は紐付けされているとユーザは使いやすい。