



大阪大学  
OSAKA UNIVERSITY

INSTITUTE for  
  
PROTEIN RESEARCH  
OSAKA UNIVERSITY

# 言語モデルを用いた 細胞シミュレーションとデータ統合

トーゴーの日シンポジウム2023

日本科学未来館

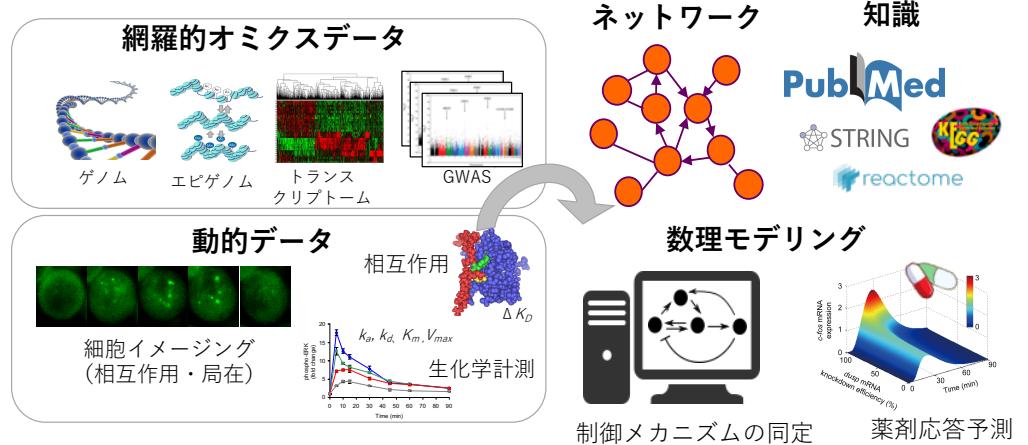
2023年10月5日

岡田眞里子  
大阪大学蛋白質研究所

# 数理モデルは生命のダイナミクスを表現し、制御機構の理解と操作の手助けをする

## 細胞の数理モデリング (常微分方程式(ODE)モデル)

- 多様な実験データを統合、解析できる  
メカニズムの同定が可能
- 通常のAIと異なり、解釈しやすい

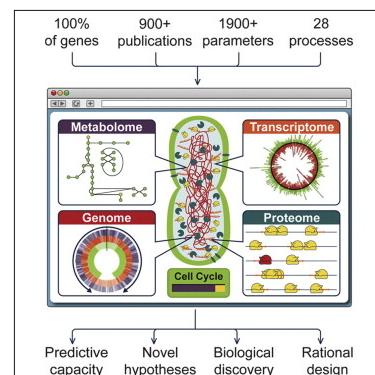


ここ10年で様々な生命データベースが整備  
詳細な細胞モデリングが可能に

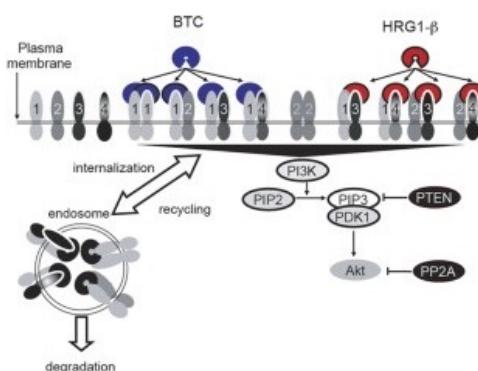
<https://www.simulations-plus.com/>

薬の毒性評価も  
イン・シリコ(米国FDA)

既存データを使いリアル実験前に  
シミュレーションで最適な候補を絞る



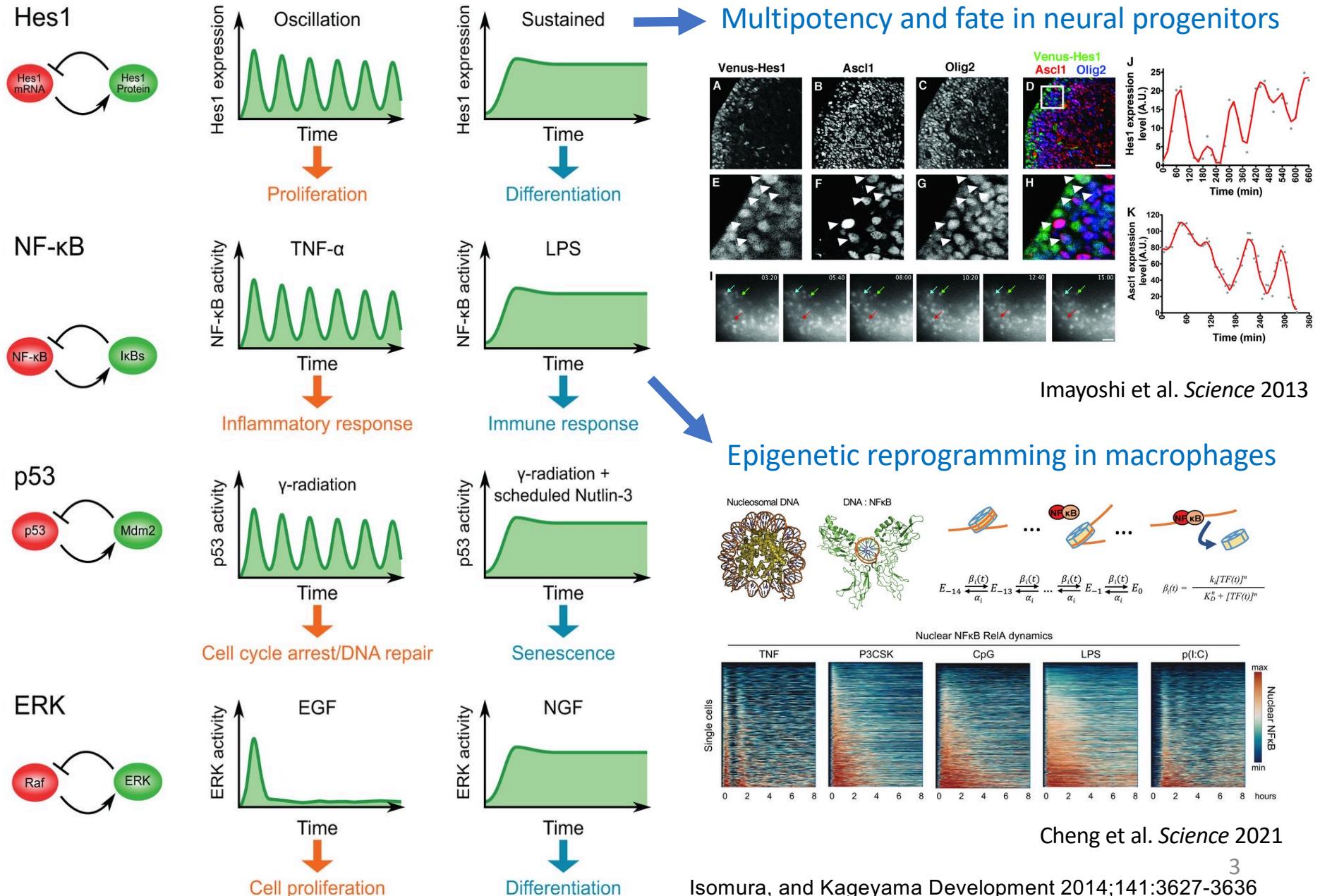
微生物の全細胞  
モデルの構築  
(Covert, 2012,  
2020)



抗ErbB3抗体  
Seribantumab  
(MM-121)の開発

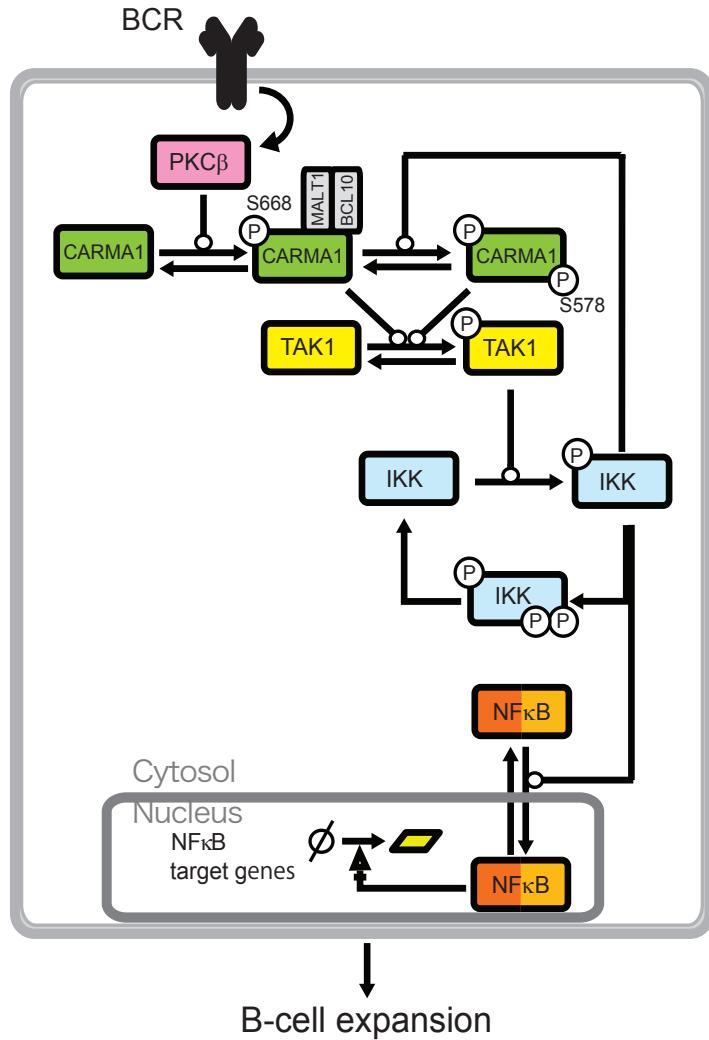
臨床試験(Phase II)  
への決断は  
イン・シリコ計算で  
(Schoeberlら 2017)

# 分子活性の有無ではなく、ダイナミクスが細胞運命を決定

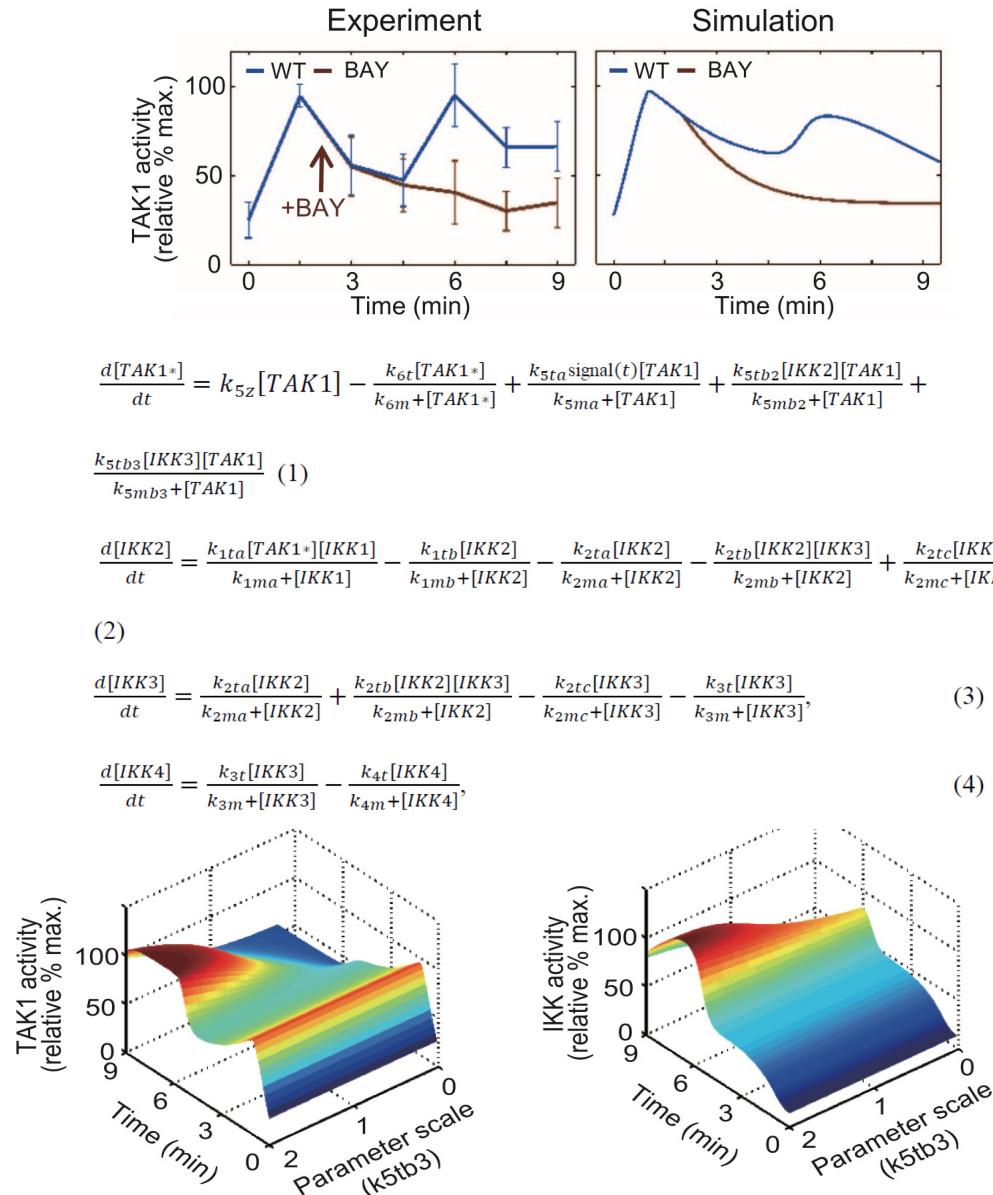


# NF-κB 転写因子の数理モデル解析

## Elucidating the Mechanisms Behind Dynamics

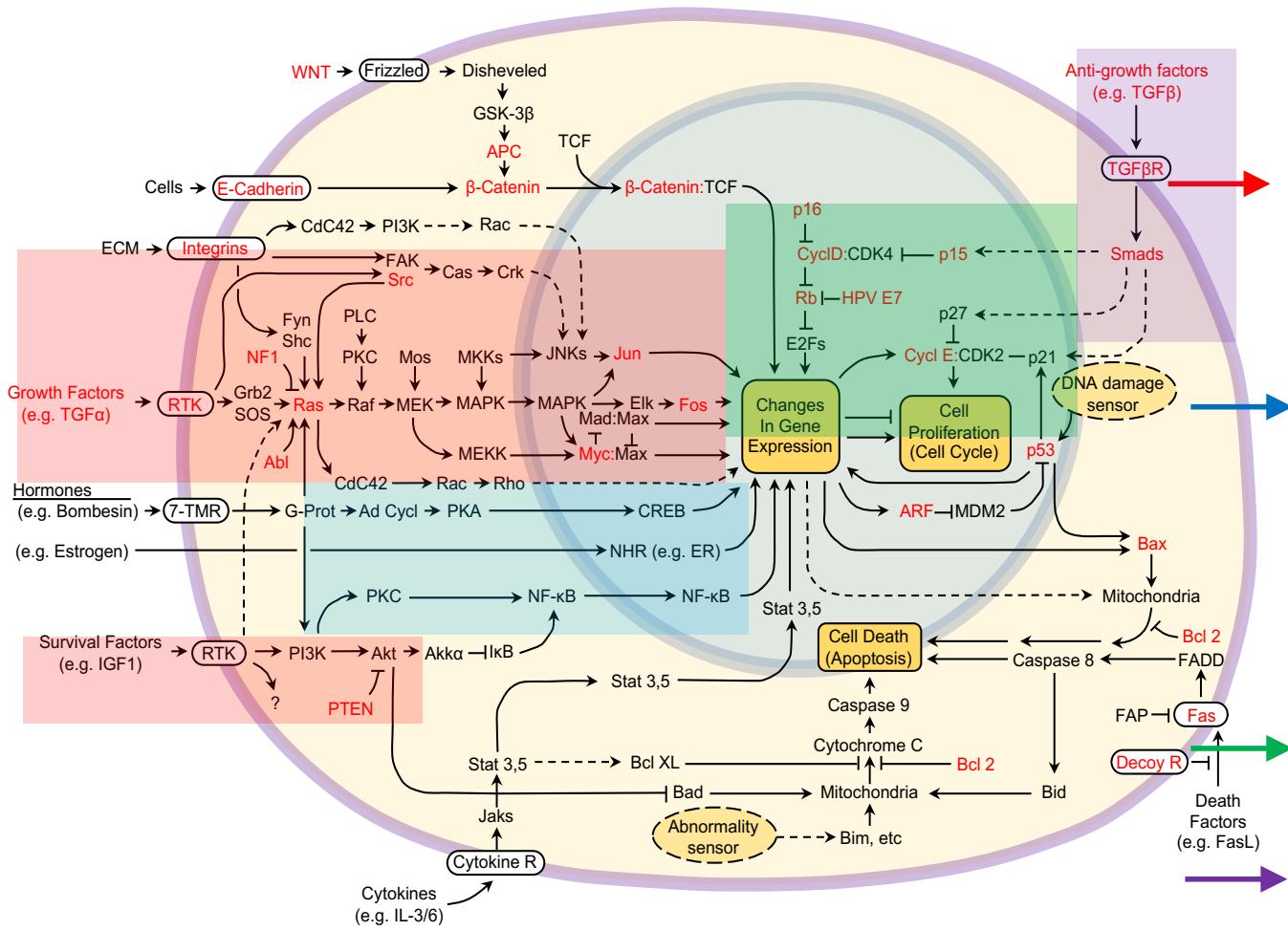


Shinohara et al. *Science* 2014



# がんシグナルネットワークの数理モデル

## 多様な時系列変化を示すサブネットワークの集合体



### ErbB receptor pathway

Hatakeyama, et al. *Biochem. J.* 2003  
 Birtwistle, et al. *Mol. Syst. Biol.* 2007  
 Nakakuki, et al. *Cell* 2010  
 Imoto, et al. *Cancers* 2020  
 Imoto, et al. *iScience* 2022

### NF- $\kappa$ B pathway

Shinohara, et al. *Science* 2014  
 Inoue, et al. *npj Syst. Biol. Appl.* 2016  
 Michida, et al. *Cell Rep.* 2020  
 Ando, et al. *npj Syst. Biol. Appl.* 2021  
 Wibisana, et al. *PLoS genetics*, 2022  
 Tabata, et al. *bioRxiv* 2023.

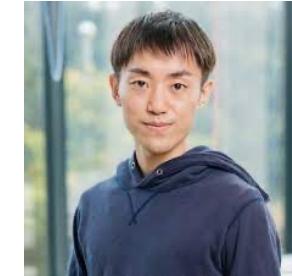
### Cell cycle regulation

Imoto, et al. *Curr. Syst. Biol.* 2018  
 Ichikawa, et al. *in prep.*

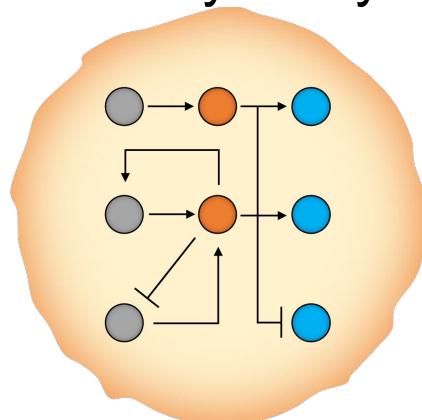
### TGF $\beta$ pathway

Haga, et al. *iScience*, *in revision*.

# 数理モデリングの計算ツールの開発



1. Model construction
2. Parameter Estimation
3. Sensitivity Analysis



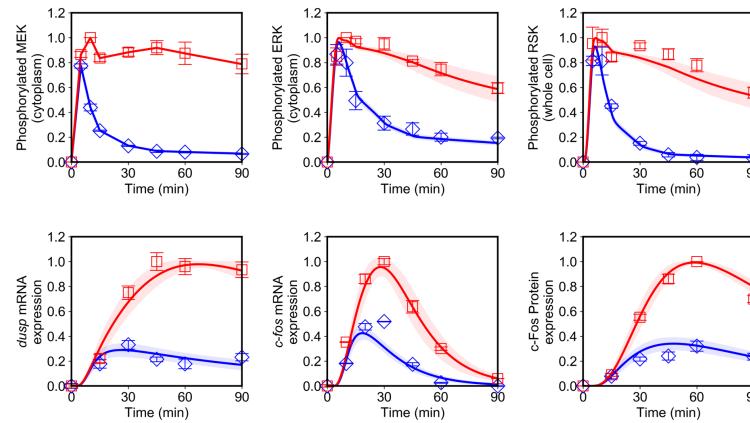
## BioMASS

Modeling and Analysis of Signaling Systems

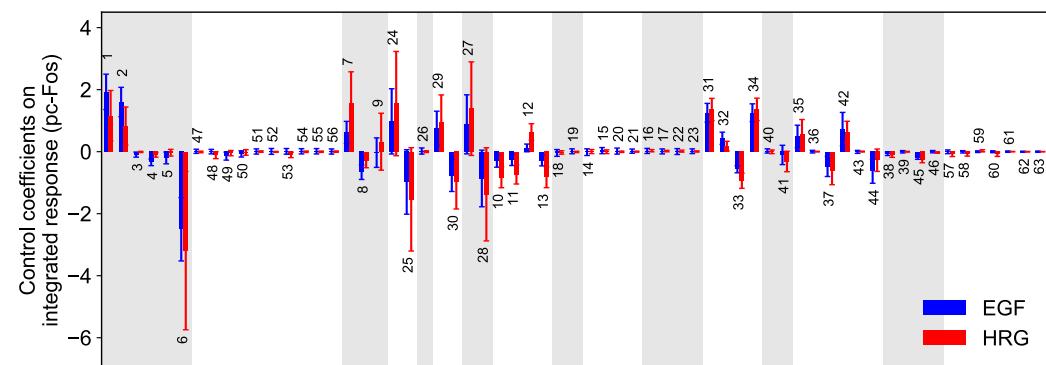
<https://github.com/okadalabipr/biomass>

Imoto et al. Cancers 2020  
Arakane et al. Submitted

## パラメータフィッティング

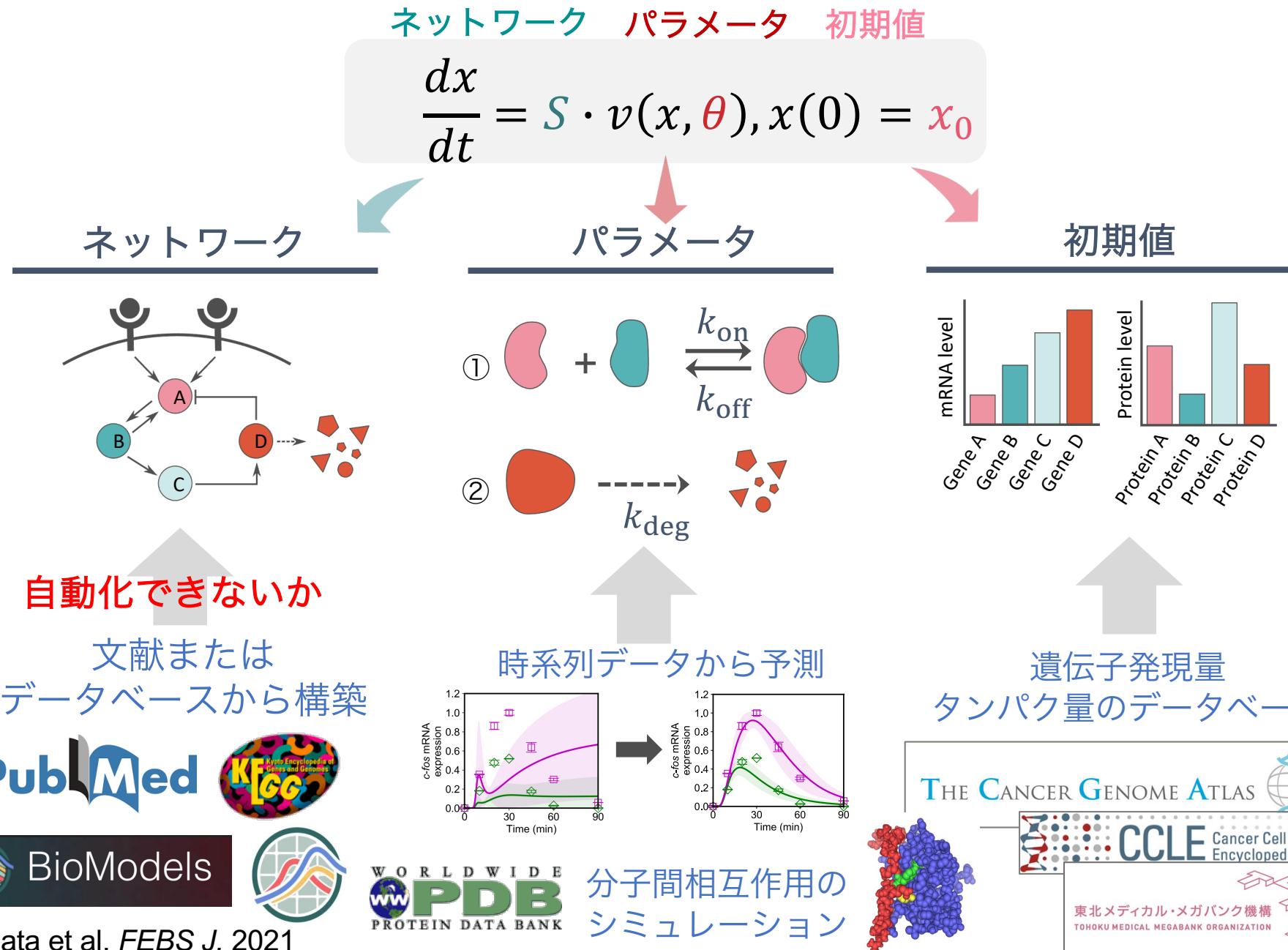


## 感度解析



61,290 downloads after release

# 細胞・疾患モデリングに必要な情報

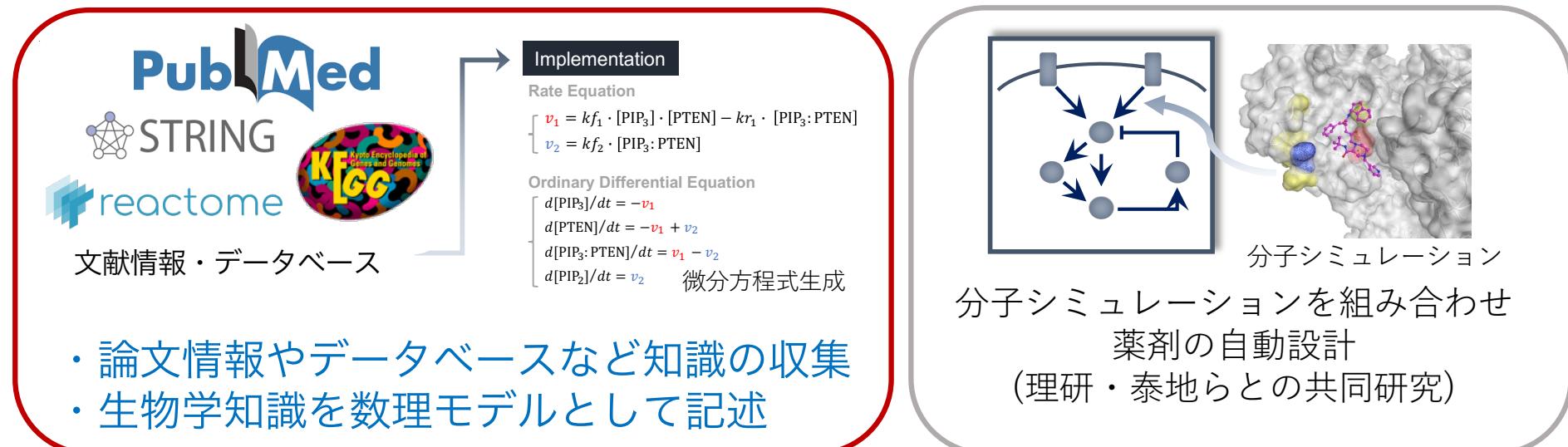


# 数理モデリング研究におけるボトルネック

## シミュレーション研究の流れ

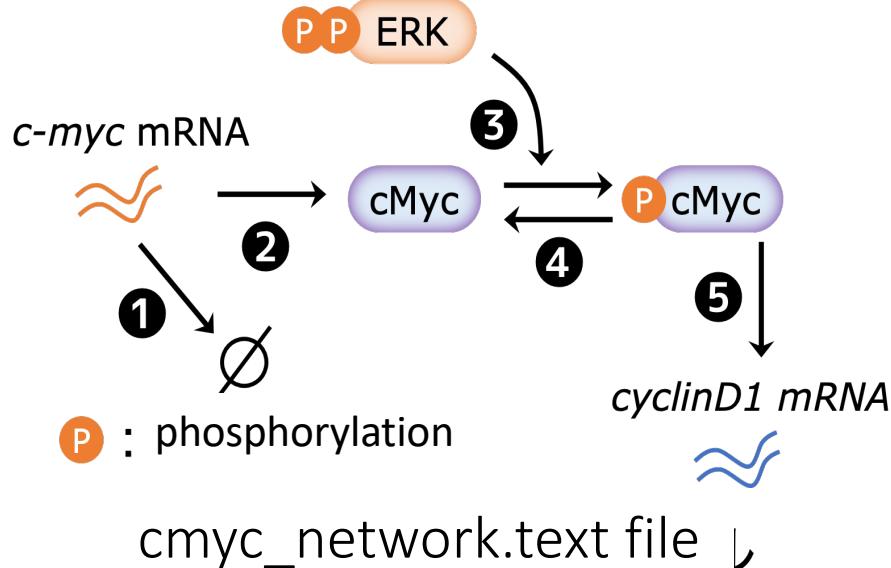


解決すべきボトルネック



# Text2model: Modeling from biochemical text

数式を入力しないで数理モデルを構築



- 1** *cmyc*RNA is degraded
- 2** *cmyc*RNA is translated into cMyc
- 3** ppERK phosphorylates cMyc --> pcMyc
- 4** pcMyc is dephosphorylated --> cMyc
- 5** pcMyc transcribes *cyclinD1*mRNA

```
>>> from biomass import Text2Model  
>>> d = Text2Model("cmyc_network.txt")  
>>> d.convert()
```

実行可能な数理モデルの自動生成

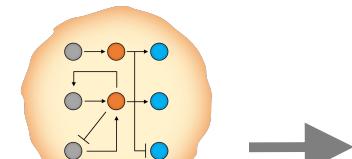
```
# reaction_network.py ファイル  
class ReactionNetwork(object):  
  
    def __init__(self) -> None:  
        super(ReactionNetwork, self).__init__()  
  
    @staticmethod  
    def flux(t, y, x) -> dict:  
        """  
        フラックススペクトル  
        """  
        v = {}  
        v[1] = x[C.kf1] * y[V.cmymRNA]  
        v[2] = x[C.kf2] * y[V.cmymRNA]  
        v[3] = x[C.V3] * y[V.ppERK] * y[V.cMyc] \  
            / (x[C.K3] + y[V.cMyc])  
        v[4] = x[C.V4] * y[V.pcMyc] \  
            / (x[C.K4] + y[V.pcMyc])  
        v[5] = x[C.V5] * y[V.pcMyc] ** x[C.n5] \  
            / (x[C.K5] ** x[C.n5] + y[V.pcMyc] ** x[C.n5])  
  
        return v
```

# Text2model: Modeling from biochemical text

数式を入力しないで数理モデルを構築

Pattern matching between text and equations

Mathematical modeling platform



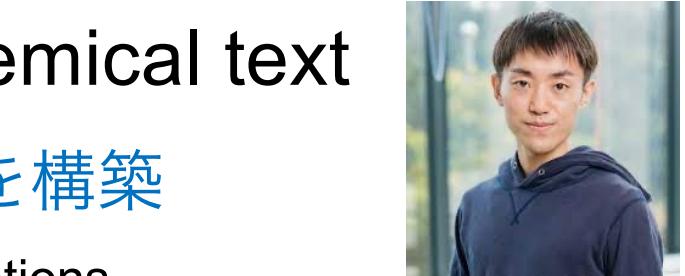
**BioMASS**

A Python framework for  
Modeling and Analysis of  
Signaling Systems

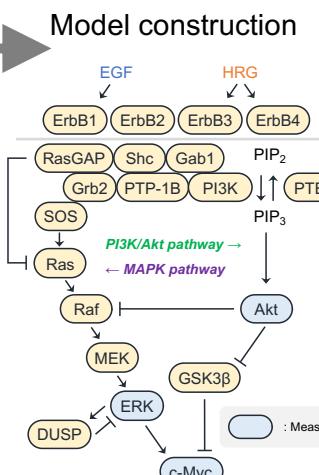
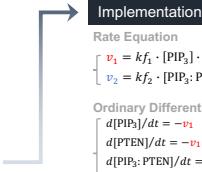
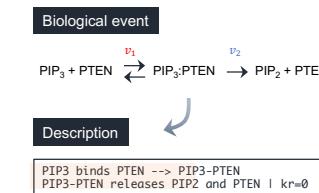
<https://github.com/okadalabipr/biomass>

Imoto et al., Cancers 2020

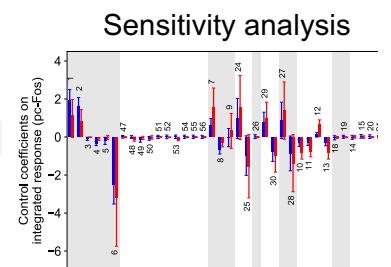
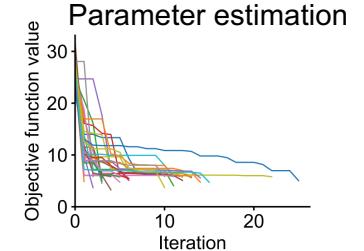
Event & Word	Rate equation	Mass balance
	<b>Mass action</b>	$d[A]/dt = -\nu$ $d[B]/dt = -\nu$ $d[AB]/dt = +\nu$
<b>A binds B --&gt; AB</b>	$\nu = kf[A][B] - kr[AB]$	
	<b>Michaelis-Menten equation</b>	$d[A]/dt = -\nu$ $d[Ap]/dt = +\nu$
<b>B phosphorylates A --&gt; Ap</b>	$\nu = V[B][A] / (K + [B])$	
	<b>Michaelis-Menten equation</b>	$d[A]/dt = +\nu$ $d[Ap]/dt = -\nu$
<b>Ap is dephosphorylated --&gt; A</b>	$\nu = V[Ap] / (K + [Ap])$	
	<b>Hill equation</b>	$d[a]/dt = +\nu$
<b>B transcribes a</b>	$\nu = V[B]^n / (K^n + [B]^n)$	
	<b>First-order kinetics</b>	$d[A]/dt = +\nu$
<b>a is translated into A</b>	$\nu = kf[a]$	
	<b>Zero-order kinetics</b>	$d[A]/dt = +\nu$
<b>A is synthesized</b>	$\nu = kf$	



Hiroaki Imoto



319 rate equations  
228 species  
648 parameters



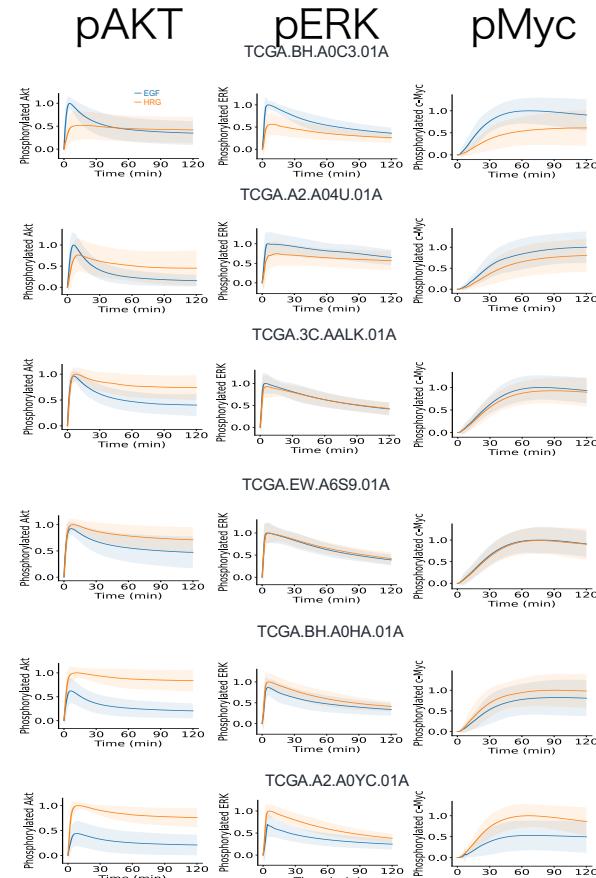
大規模なODEモデルの構築に威力を発する

Imoto et al. *iScience* 2022

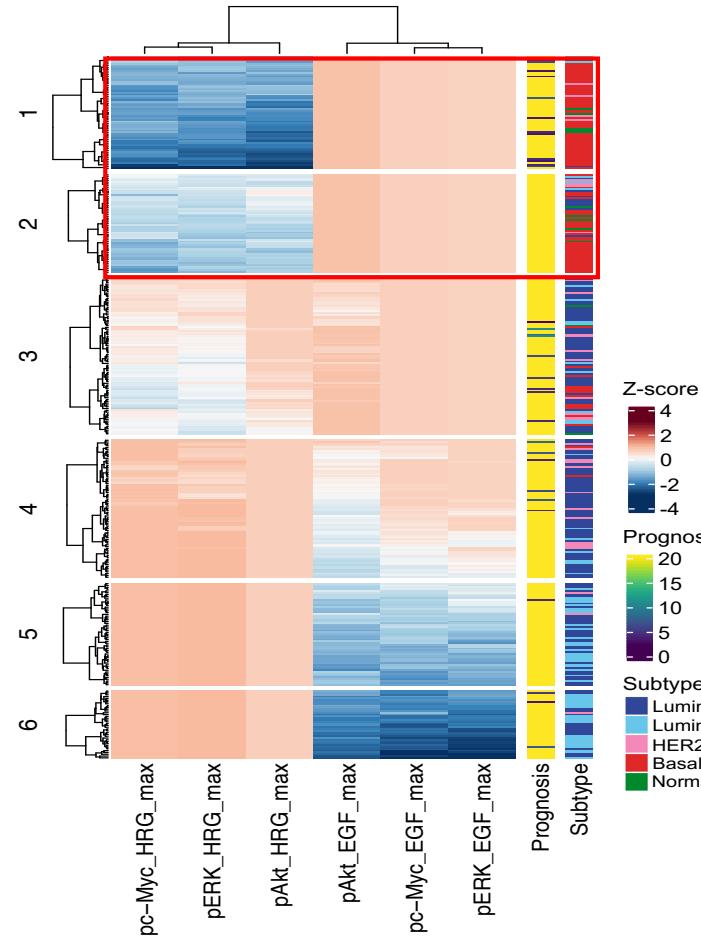
# 患者ひとりひとりの遺伝子発現データをモデルに与え In silicoの分子活性ダイナミクスから予後予測

患者ごとの分子活性のシミュレーション結果

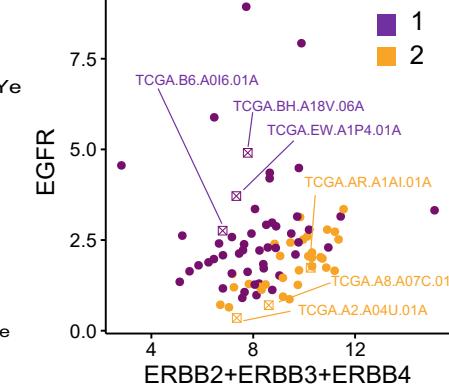
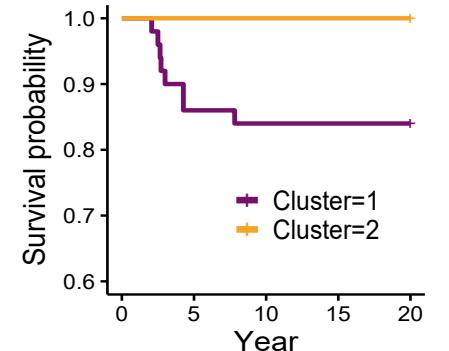
TCGA乳がん遺伝子発現データ  
(n=377)



Triple negative (TNBC) 予後分類



膜受容体比が予後に関連



EGFRキナーゼ阻害剤への  
応答予測・検証

ダイナミクスそのものが疾患のバイオマーカーとなり得る

# 自然言語処理(NLP)を用いた数理モデリング

## Data Source

(1) Databases



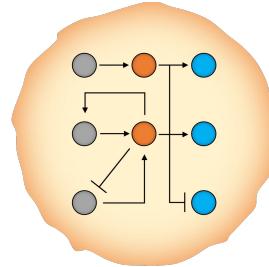
(2) Literatures



- 2.1. Large Language Model (LLM) eg: GPT3
- 2.2. Distant Supervision (遠距離教師あり学習)

## Text2Model

### Intermediate language



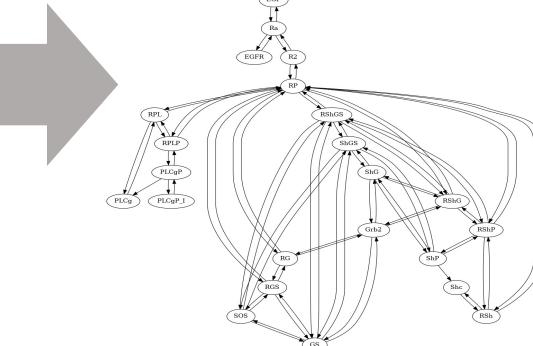
## BioMASS

A Python framework for  
Modeling and Analysis of  
Signaling Systems  
Imoto *et al.*, Cancers (2020)

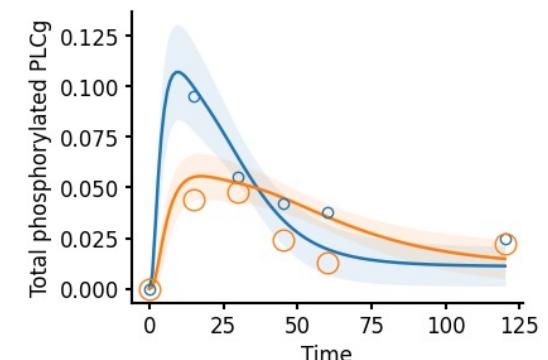
- Detailed balance
- Parameter optimization
- Sensitivity analysis

## Executable model

```
1 EGF binds EGFR <--> Ra | kf=0.003, kr=0.06 | EGFR=10
2 Ra dimerizes <--> R2 | kf=0.01, kr=0.01
3 R2 is phosphorylated <--> RP | kf=1, kr=0.01
4 RP is dephosphorylated --> R2 | V=450, K=50
5 RP binds PLCg <--> RPL | kf=0.06, kr=0.2 | PLCg=105
6 RPL is phosphorylated <--> RPLP | Kf=1, kr=0.05
7 RPLP is dissociated into RP and PLCgP | kf=0.3, kr=0.05
8 PLCgP is dephosphorylated --> PLCg | V=1, K=100
9 RP binds Grb2 <--> RG | kf=0.003, kr=0.05 | Grb2=85
10 RG binds SOS <--> RGS | kf=0.01, kr=0.06 | SOS=34
11 RGS is dissociated into RP and GS | kf=0.03, kr=4.5e-05
12 GS is dissociated into Grb2 and SOS | kf=1.5e-3, kr=0.003
13 RP binds Shc <--> RSh | kf=0.09, kr=0.6 | Shc=150
14 RSh is phosphorylated <--> RShP | kf=6, kr=0.06
15 RShP is dissociated into Shc and RP | kf=0.3, kr=0.05
```



## Simulation (validation)

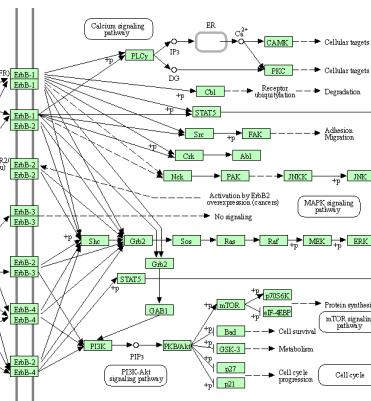


# (1) KEGG2Model



Pathway Structure

KEGG ErbB pathway  
in KGML format



Convert to Text2Model format

```

TGFA activates EGFR_EGFR --> a_EGFR_EGFR
a_EGFR_EGFR is deactivated --> EGFR_EGFR
NRG3 activates ERBB4_ERBB4 --> a_ERBB4_ERBB4
a_ERBB4_ERBB4 is deactivated --> ERBB4_ERBB4
NRG3 activates ERBB4_ERBB2 --> a_ERBB4_ERBB2
a_ERBB4_ERBB2 is deactivated --> ERBB4_ERBB2
NRG2 activates ERBB3_ERBB3 --> a_ERBB3_ERBB3
a_ERBB3_ERBB3 is deactivated --> ERBB3_ERBB3
NRG2 activates ERBB4_ERBB4 --> a_ERBB4_ERBB4
NRG2 activates ERBB3_ERBB2 --> a_ERBB3_ERBB2
a_ERBB3_ERBB2 is deactivated --> ERBB3_ERBB2
NRG2 activates ERBB4_ERBB2 --> a_ERBB4_ERBB2
HBEGF activates EGFR_EGFP
  
```

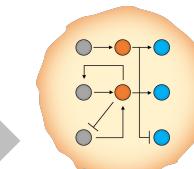
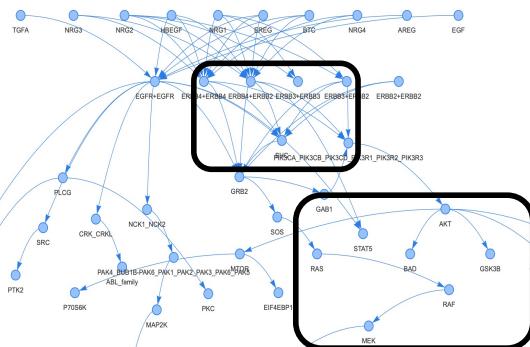


PubMed全文  
アブスト

Text-Mined  
Knowledge

Highlight & Extract  
Context-Dependent  
Subpathways

Transfer to Python format  
(General pathway)

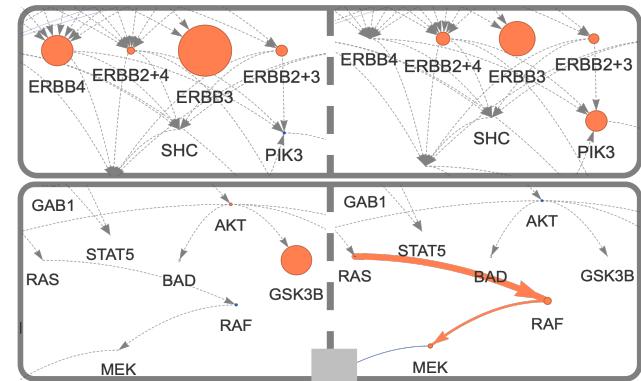


BioMASS

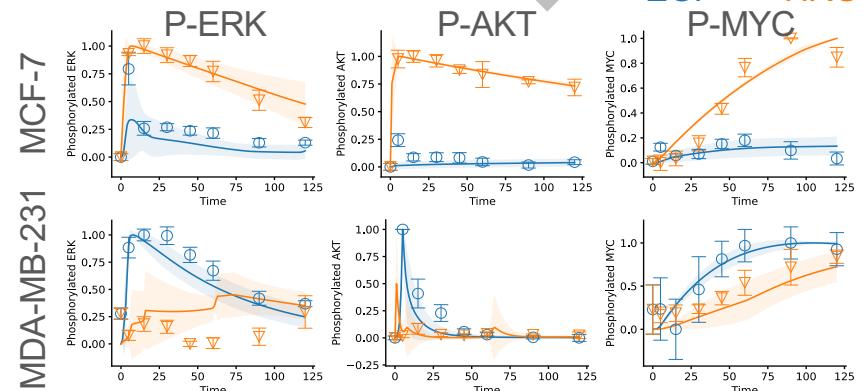


Kiwamu Arakane

Context-dependent pathway  
MDA-MB-231 MCF-7

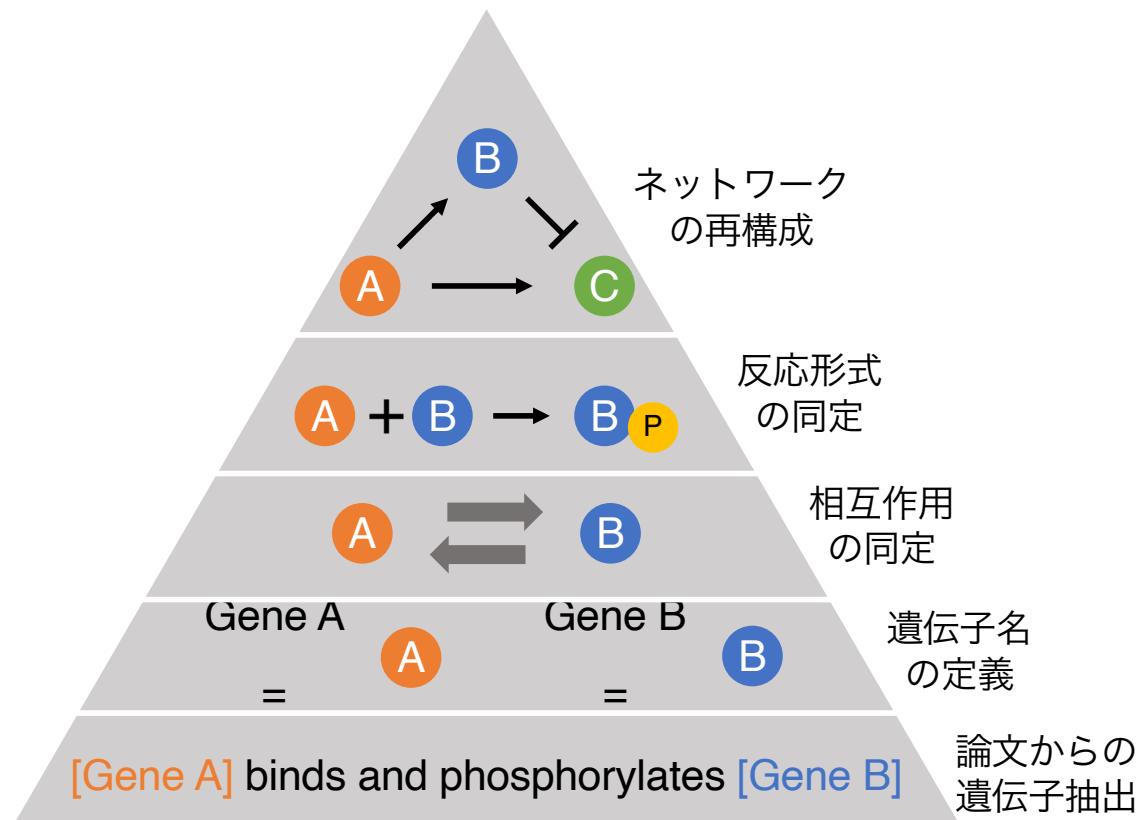


Parameter fitting



## (2) 文献からのネットワーク構築: 主な2つの手法

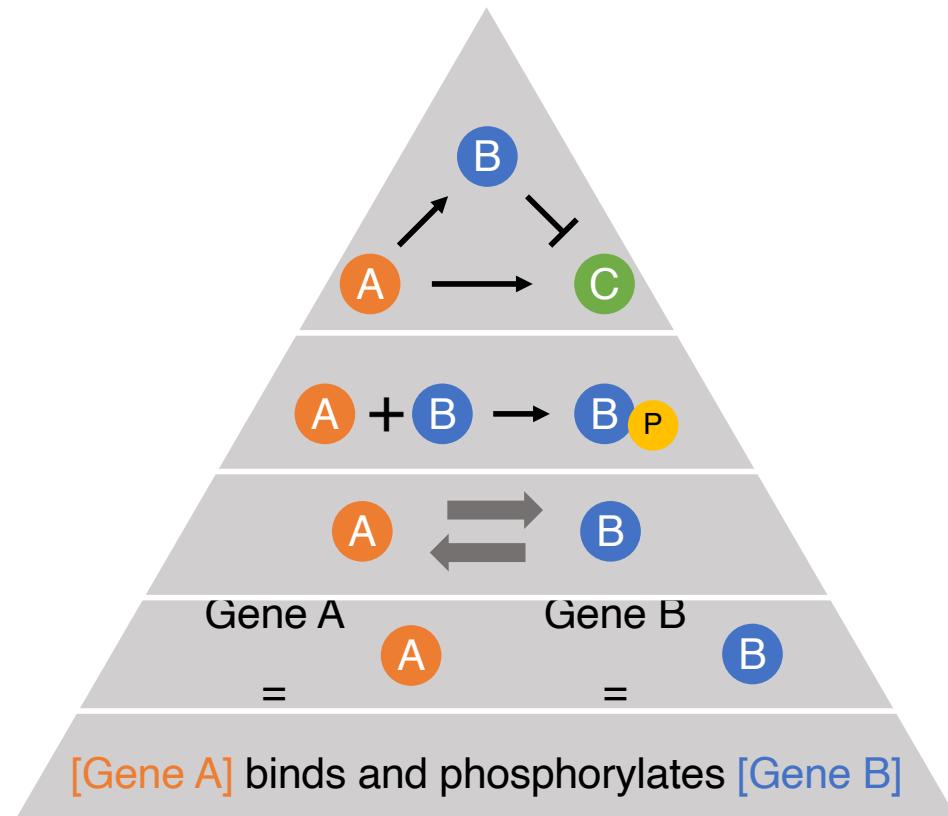
Follow the Pyramid Step-by-Step



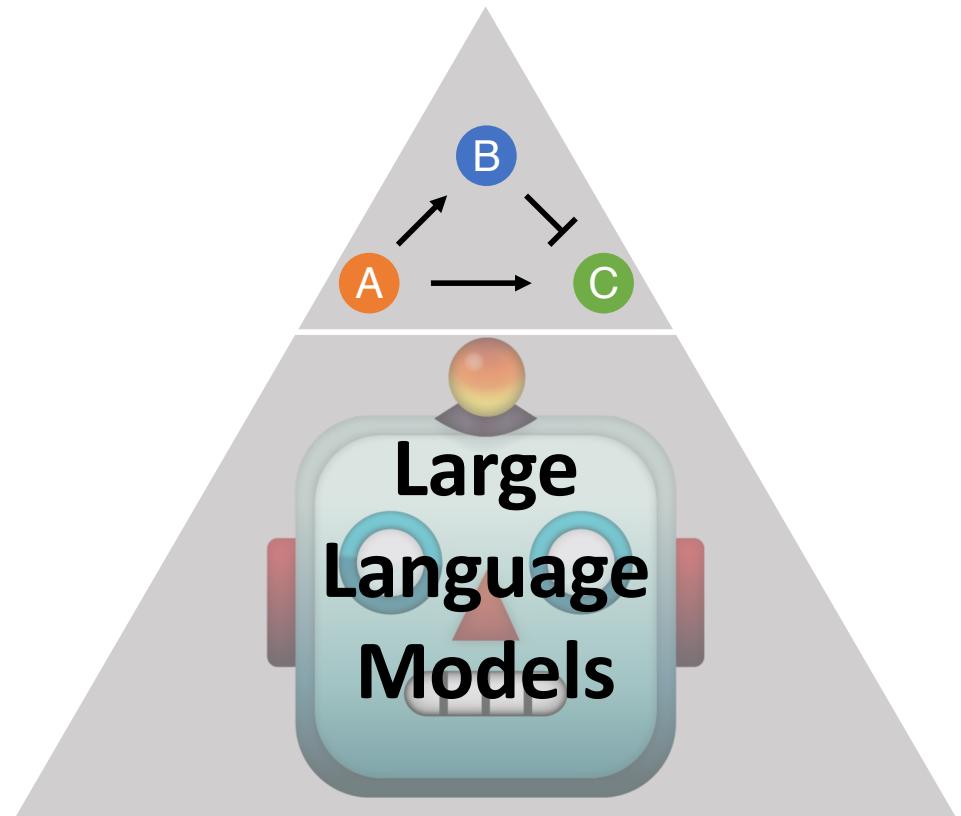
Adapted from: Li, C., et al. *Briefings in Bioinformatics*, 2013.

## (2) 文献からのネットワーク構築: 主な 2 つの手法

Follow the Pyramid Step-by-Step



Harness the Power of LLMs



Adapted from: Li, C., et al. *Briefings in Bioinformatics*, 2013.

# (2.1) LLM using GPT3 (text-davinci-003)

## Task

Given the current state of a list and a prompt, extrapolate as many reactions of molecules as possible from the prompt and update the list. Every line in the list contains gene names which act as the reactants and type of reaction of the left and products on the right, connected with different types of arrows depending on the reversibility of the reaction.

## Task of example

current state:

```
[  
A binds B <--> AB  
AB phosphorylates C --> pC  
pC is dephosphorylated --> C  
]
```

prompt: The phosphorylated protein C exhibits kinase activity and phosphorylates its substrate D...

new state:

```
[  
A binds B <--> AB ...  
]
```

← Input-Output format of  
Text2Model

← prompt  
← Text2Model-format output

## Problem

current state:

```
[  
EGF binds EGFR <--> Ra  
Ra dimerizes <--> R2  
R2 is phosphorylated --> RP  
RP is dephosphorylated --> R2  
]
```

prompt: In step 1, EGF binds to the extracellular domain of the monomeric EGFR (designated as R in the kinetic scheme) and forms the EGF-EGFR complex (designated as Ra). EGF binding drives...

← Input a text

THE JOURNAL OF BIOLOGICAL CHEMISTRY  
© 1999 by The American Society for Biochemistry and Molecular Biology, Inc.

Vol. 274, No. 42, Issue of October 15, pp. 30169–30181, 1999  
Printed in U.S.A.

## Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor\*

### EGFR-Rasカスケードのモデル論文

(Received for publication, July 29, 1998, and in revised form, August 4, 1999)

Boris N. Kholodenko<sup>‡\$</sup>, Oleg V. Demin<sup>‡†</sup>, Gisela Moehren<sup>‡</sup>, and Jan B. Hoek<sup>‡</sup>

From the <sup>‡</sup>Department of Pathology, Anatomy and Cell Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107 and the <sup>§</sup>A. N. Belyovskiy Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899, Russia

#### Kinetic Analysis

Schematic Representation of Protein-Protein Interactions Induced by EGF Binding

For a quantitative analysis of the EGFR signaling network, an adequate description is required of the reactions that contribute to the experimentally detected protein-protein interactions and tyrosine phosphorylation events. The kinetic scheme presented in Fig. 1 forms the basis for the integration of the experimental study and the computational analysis.

In step 1, EGF binds to the extracellular domain of the monomeric EGFR (designated as R in the kinetic scheme) and forms the EGFR-EGFR complex (designated as Ra). EGF binding drives the association of two receptor monomers into an activated receptor dimer (step 2). Recent studies (26, 27) have shown that a 2:2 (EGF:EGFR) complex is the predominant form of the receptor dimer (designated as R<sub>2</sub>). The phosphorylation of tyrosine residues by receptor tyrosine kinase is combined into a single step 3, yielding a form designated as RP. Although multiple tyrosine residues in the cytoplasmic tail of the receptor are targets for autop phosphorylation, we did not attempt to distinguish experimentally between different phosphorylation forms of the receptor, and, as we will discuss below, the initial computational analysis also does not require a functional distinction to be made. Step 4 is the dephosphorylation of RP, catalyzed by one or more phosphotyrosine phosphatase(s) (28, 29).

Tyrosine phosphorylation triggers the binding of cytosolic proteins to the receptor. We consider here three proteins that directly interact with phosphotyrosine residues on the receptor, namely Grb2, Shc, and PLC $\gamma$ (4). Although several other proteins bind to the activated EGFR, it is helpful to consider a limited number of target proteins as an initial core model. It is not entirely clear whether these multiple proteins can bind simultaneously to their target phosphotyrosine residues on the same receptor molecule or whether the order of, for example, Grb2 to the receptor occurs before the binding of PLC $\gamma$  (competitive binding). The model depicted in Fig. 1 considers the binding of cytosolic proteins to occur by a competitive mechanism. The advantage of a model with competitive binding is that it allows us to consider receptor phosphorylation as a single step rather than monitoring different phosphorylated forms of R<sub>2</sub> as distinct entities. We also assume that, when Grb2, Shc, or PLC $\gamma$  are bound to EGFR, the corresponding phosphotyrosine residues are not available to receptor phosphotyrosine phosphatases. The implications of these assumptions for the dynamic patterns of EGFR signaling will be considered below. Which mechanism of interactions of EGFR and adapter proteins occurs *in vivo* remains to be identified.

The entire network of reactions of the receptor with its cytosolic target proteins can now be divided into three coupled cycles of interactions with Grb2, PLC $\gamma$ , and Shc, respectively. One receptor cycle includes the binding of PLC $\gamma$  (step 5 in Fig. 1, resulting in the formation of the complex designated as R-PL) and phosphorylation of PLC $\gamma$  at two tyrosine residues by receptor tyrosine kinase (step 6, yielding the complex R-PLP). The partial cycle of the receptor is completed by the dissociation of R-PLP into phosphorylated phospholipase C $\gamma$  (PLC $\gamma$ P) and RP in step 7. Tyrosine phosphorylation of PLC $\gamma$  is thought to be necessary for its activation and the subsequent formation of inositol 1,4,5-trisphosphate (IP<sub>3</sub>) and diacylglycerol (DAG), which elicit a Ca<sup>2+</sup> response (30, 31). PLC $\gamma$ P can translocate to cytoskeletal or membrane structures (step 25), which yields bound PLC $\gamma$ -I (32, 33).

Another partial receptor cycle starts with the binding of Grb2 to a receptor phosphotyrosine (step 9, forming the complex R-G). The complex of the EGFR receptor with the adapter protein Grb2 is a branch point that leads to several signaling pathways through binding to different potential targets. Here we consider the link of EGFR to the Ras signaling pathway. The SH3 domains of Grb2 bind to proline-rich regions of the Ras-specific GDP-GTP exchange factor SOS. In step 10, SOS binds to the receptor-bound Grb2, resulting in the formation of the ternary complex R-G-S. The binding of SOS to the EGFR-Grb2 complex localizes SOS to the vicinity of the membrane and is amenable to the cell membrane. The ternary complex R-G-S dissociates (step 11), yielding the phosphorylated receptor (RP) and the complex G-S, which further dissociates into Grb2 and SOS (step 12).

The final EGFR cycle considered here includes the formation of the complex of Shc with EGFR (R-SH) (step 13 in Fig. 1) and its subsequent phosphorylation at Tyr<sup>317</sup> by receptor tyrosine kinase (step 14, yielding R-Shc<sup>p</sup>). This allows Grb2 to also bind to EGFR indirectly through phosphorylated Shc, forming a ternary complex (R-Shc<sup>p</sup>) (step 17). There are three embedded EGFR cycles that involve Shc protein. The shortest of these cycles is completed in step 15, where the complex

TABLE I Kinetic equations comprising the computational model

$d[EGF]/dt = -v_1$
$d[R]/dt = -v_1$
$d[R_2]/dt = -2v_1 + 2v_2$
$d[R]/dt = v_2 + v_3 - v_3$
$d[RP]/dt = v_3 + v_4 + v_{11} + v_{15} + v_{18} + v_{20} - v_4 - v_5 - v_9 - v_{11}$
$d[R-PL]/dt = v_5 - v_6$
$d[R-G]/dt = v_6 - v_7$
$d[R-SH]/dt = v_{10} - v_{11}$
$d[R-SH-G]/dt = v_{13} - v_{14} - v_{24} - v_{16} - v_{17}$
$d[R-Sh-G]/dt = v_{17} - v_{14} - v_{19}$
$d[R-Sh-G-S]/dt = v_{19} - v_{20} - v_{24}$
$d[G-S]/dt = v_{11} + v_{23} - v_{12} - v_{24}$
$d[Shp]/dt = v_{13} + v_{23} - v_{21} - v_{16}$
$d[Sh-G]/dt = v_{18} + v_{21} - v_{22}$
$d[PLC\gamma]/dt = v_8 - v_6$
$d[PLC\gamma-P]/dt = v_6 - v_8 - v_{25}$
$d[PLC\gamma-P-II]/dt = v_{25}$
$d[Grb]/dt = v_{14} - v_9 - v_{17} - v_{21}$
$d[Shc]/dt = v_{16} - v_{13}$
$d[SOS]/dt = v_{12} - v_{10} - v_{19} - v_{22}$

R-SH dissociates, yielding the phosphorylated receptor (RP) and phosphorylated Shc (Shp). The second cycle is completed in step 18, where the ternary complex R-Sh-G dissociates into RP and the complex Sh-G. The longest of the three embedded cycles includes SOS binding to R-Sh-G, leading to the formation of a four-protein complex, R-Sh-G-S (step 19). The complex R-Sh-G-S can also be formed by association of R-Sh-P and G-S complexes in step 24. The third cycle is completed in step 20, where the complex R-Sh-G-S dissociates, releasing the phosphorylated receptor (RP) and the complex Sh-G.

It is unknown whether the binding of the phosphorylated target protein to EGFR protects them from specific phosphatases. The kinetic scheme of Fig. 1 assumes that PLC $\gamma$ -P and Shp<sup>p</sup> are dephosphorylated only after they dissociate from the receptor (steps 16 and 8). However, this assumption is not critical, provided the dephosphorylation of bound target proteins proceeds no faster than that of their unbound phosphorylated forms.

After phosphorylated Shc dissociates from the receptor (Shp<sup>p</sup>), it retains its ability to bind various SH2 domain-containing targets. The remaining steps in Fig. 1 constitute the cycle of Shp. The scheme shows that Grb2 binds to Shp, forming the complex Sh-G (step 21). The GDP-GTP exchange factor SOS is able to bind to Grb2 complexed with phosphorylated Shc, forming the ternary complex Sh-G-S (step 22). The dissociation of the complex Sh-G-S yields G-S and Shp (step 23).

#### Derivation of a Kinetic Model

**Kinetic Equations**—In order to integrate the experimental observations in a description of the dynamic behavior of the EGFR signaling network, we converted the reaction schemes of Fig. 1 into a set of mathematical equations known as chemical kinetics equations (34). For changes with time of the concentration of any component, e.g., the receptor form RP, one can write the following.

$$\text{Rate of change of RP concentration} = \text{total rate of RP production} - \text{total rate of RP consumption} \quad (\text{Eq. 1})$$

Here the total rate is the sum of the rates that produce or consume RP according to the kinetic diagram. For instance, the total rate of RP production equals the sum of the (net) rates of six steps (steps 3, 7, 11, 15, 18, and 20; see Fig. 1). A complete set of chemical kinetic equations describing the reactions of Fig. 1 is provided in Table I.

Kinetic equations are usually written in terms of concentrations (rather than mole numbers), since the equations are functions of concentrations. If the same compound participates in reactions taking place in different compartments with different volumes, the effective concentration of that compound will be different depending on the volume of the corresponding compartment. Step 1 (EGF binding to EGFR) could be considered as taking place in the extracellular compartment with a given initial concentration of EGF. The concentration of EGFR in the extracellular compartment would then be calculated as the number of the receptors on the cell surface divided

Kholodenko et al., 1999

## (2.1) Result: LLM-based model

```

1 EGF binds EGFR <--> Ra
2 Ra dimerizes <--> R2
3 R2 is phosphorylated <--> RP
4 RP is dephosphorylated --> R2
5 RP binds PLCg <--> RPL
6 RP phosphorylates PLCg --> RPLP
7 RPLP is dissociated into RP and PLCgP
8 PLCgP is dephosphorylated --> PLCg
9 RP binds Grb2 <--> RG
10 RG binds SOS <--> RGS
11 RGS is dissociated into RP and GS
12 GS is dissociated into Grb2 and SOS
13 RP binds Shc <--> RSH
14 RP phosphorylates Shc --> RShP
15 RShP is dissociated into ShP and RP
16 ShP is dephosphorylated --> Shc
17 RShP binds Grb2 <--> RShG
18 RShG is dissociated into RP and ShG
19 RShG binds SOS <--> RShGS
20 RShGS is dissociated into ShGS and RP
21 ShP binds Grb2 <--> ShG
22 ShG binds SOS <--> ShGS
23 ShGS is dissociated into ShP and GS
24 RShP binds GS <--> RShGS
25 PLCgP is translocated to cytoskeletal or membrane structures <--> PLCgP_I
    
```

# RPL is phosphorylated <--> RPLP

# missing

# RSh

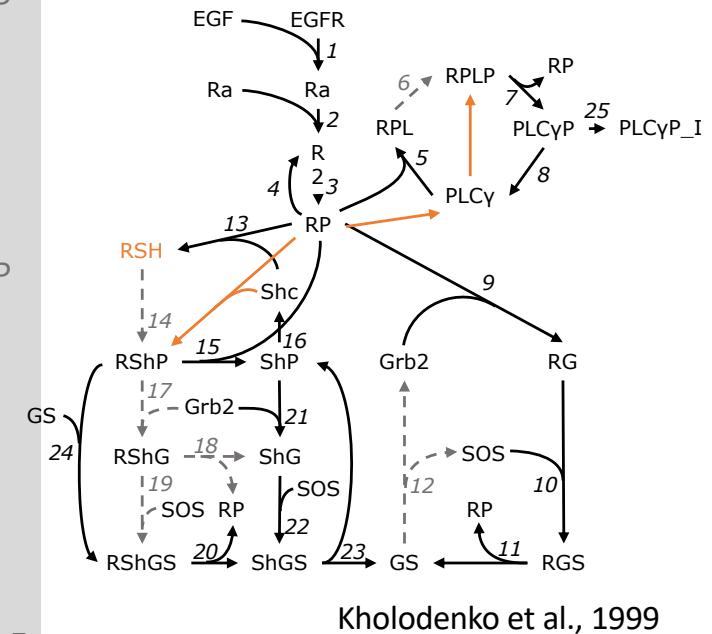
# RSh is phosphorylated <--> RShP

# missing

# missing

# missing

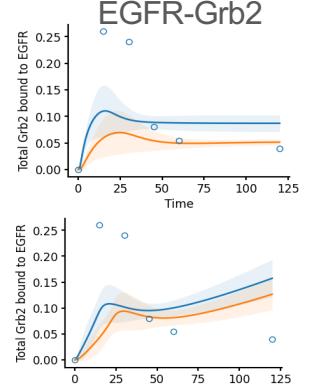
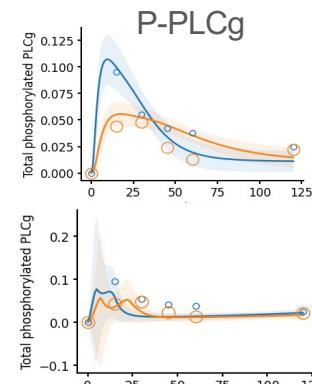
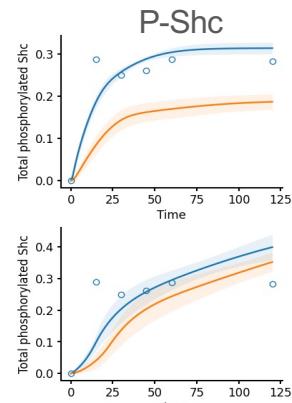
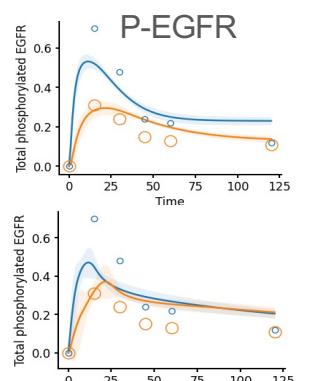
## Network given from the text



Kholodenko et al., 1999

## Simulation result

Original

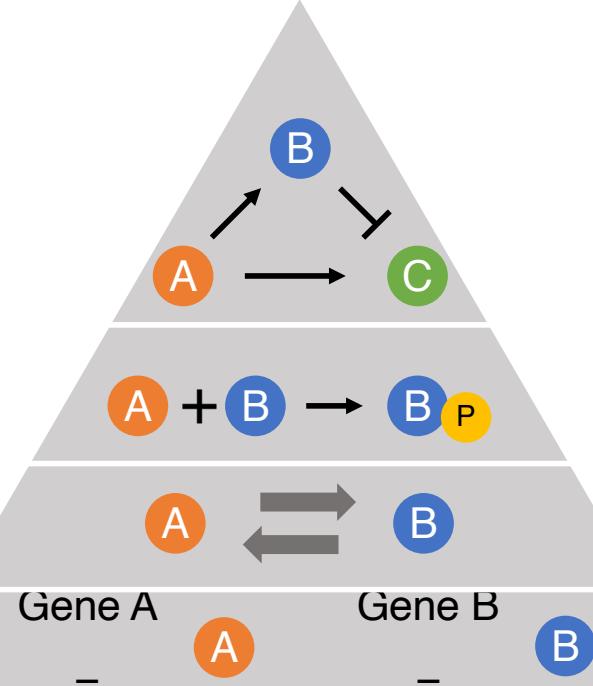


blue: 20nM  
EGF  
orange:  
2nM EGF

LLM

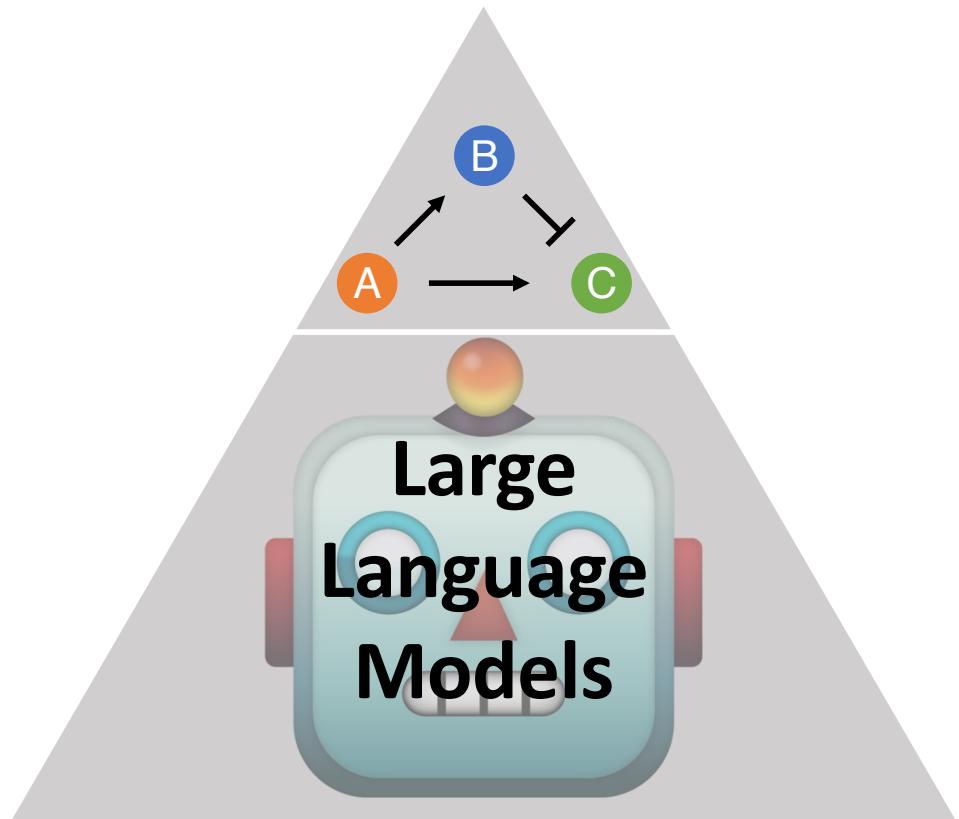
## (2) 文献からのネットワーク構築: 主な 2 つの手法

### Follow the Pyramid Step-by-Step



Adapted from: Li, C., et al. *Briefings in Bioinformatics*, 2013.

### Harness the Power of LLMs



## (2.2) 遠距離教師あり学習 (Distant Supervision)

生物医学系論文DB

PubTator

PubMed PMC

... A is known to phosphorylate B...  
... A, B, and C are the main targets of...  
... examining the expression of A and B in...

遺伝子名、疾患名  
などの固有表現の  
自動アノテーション

入力 : AとBを含む文

一文中の固有表現の共起を抽出

PMID11156524

2000

Overexpression of ErbB2 in cancer and ErbB2-targeting strategies.

Yu D, Hung MC • Oncogene

BioCXML

This past decade has witnessed the remarkable advances in the understanding of the role of the erbB2 gene in cancers and the stunning progress in developing targeted therapies for erbB2-overexpressing cancers. Activation of the ErbB2 receptor signaling pathways can enhance various metastasis-associated properties that lead to an increase of cancer metastasis. Additionally, ErbB2 overexpression confers therapeutic

BIOCONCEPTS

GENE  DISEASE

CHEMICAL  MUTATION

SPECIES  CELLLINE

<https://www.ncbi.nlm.nih.gov/research/pubtator/>

非公開

## (2.2) 遠距離教師あり学習：手法

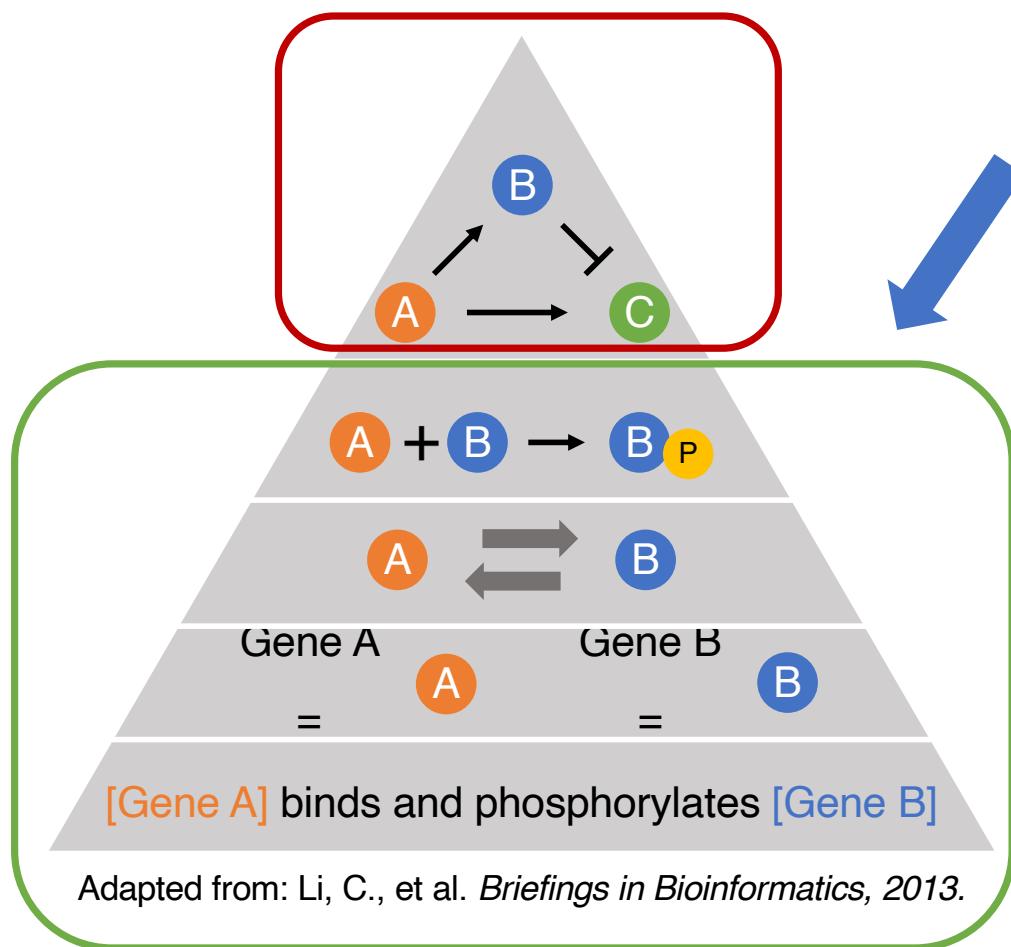
非公開

## (2.2) 遠距離教師あり学習：結果

非公開

## (2) 文献からのネットワーク構築: 主な2つの手法

カスケード構造の構築をどうするか



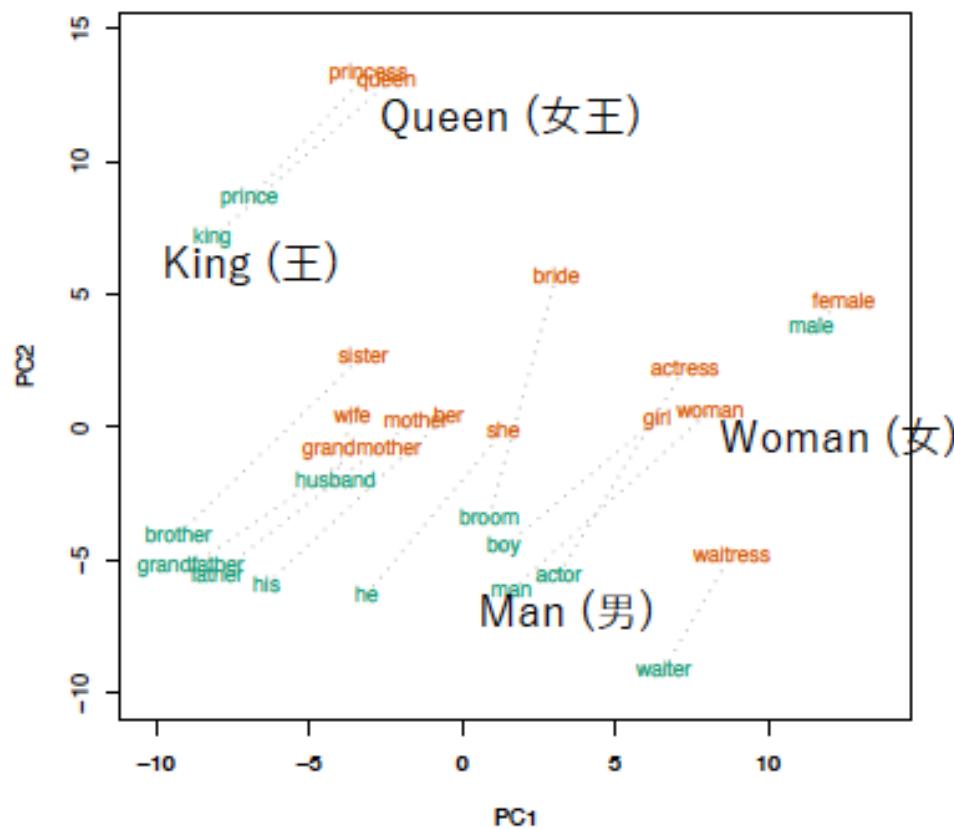
非公開

# がんシグナルのベクトルデータベースの構築

非公開

# 言語をベクトル化する

加法構成性

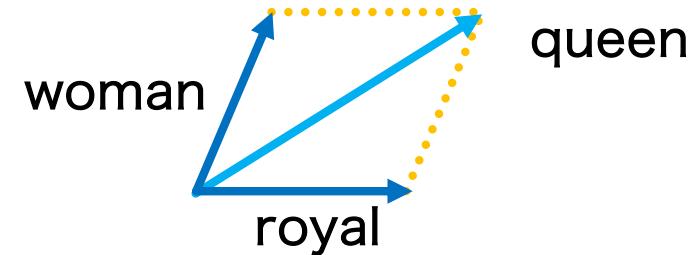


$$\text{king} + (\text{woman} - \text{man}) = \text{queen}$$

ベクトルの和で意味を構成できる  
単語や句の関係性を計算できる

Wikipedia などから skipgram [Mikolov et al., 2013] で学習した分散表現では、  
以下の関係が時として成立する

Mikolov et al. Proc. of NAACL-HLT 2013



$$\text{woman} + \text{royal} = \text{queen}$$

山際、下平（京大）

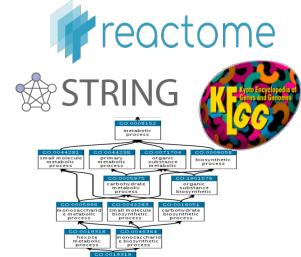
単語ベクトルにより単語間の関係性を抽出・演算する

非公開

# まとめ 言語は多種多様なデータを統合できる可能性を秘める

## 自然言語処理を拡張したグラフ埋め込み

データベース  
(パスウェイ  
オントロジー)



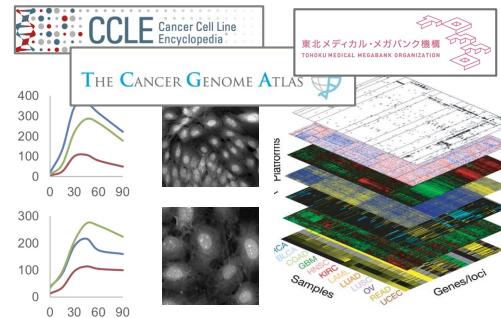
通常の単語埋込み  
(自然言語処理)

文献テキスト



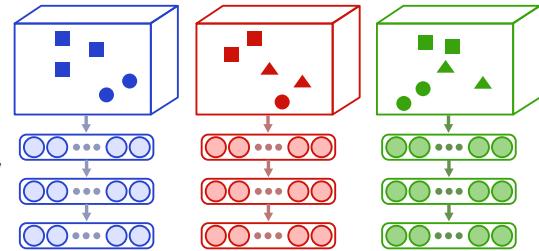
Pcclust: an R package for assessing the uncertainty in hierarchical clustering  
Ryota Sasaki, Makoto Shimodaira  
Affiliations: 1 Institute for Molecular Medicine  
Published online: DOI: 10.1186/bioinformatics/2012/10/102  
Abstract  
Pcclust is an R package for a statistical software for assessing the uncertainty in hierarchical clustering. This tool can be used easily for general statistical problems, such as DNA microarray analysis, to perform the bootstrap analysis of clustering, which has been popular in bioinformatics. In addition, it can be used for other hierarchical clustering methods using bootstrapping resampling techniques. Two types of p-values are available: approximately unbiased (AU)

イメージング・オミクス  
(実験データ)

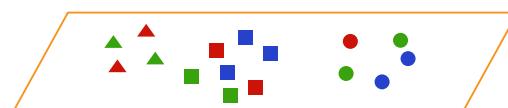


Nat. Genet. 45, 2013

ニューラル  
ネットワーク



共通空間



遺伝子、薬剤、相互作用等を表現  
特徴ベクトル (関連性の近さがわかる)

(京大・下平研との共同研究)

## 特徴ベクトルのアナロジー計算

自動翻訳(word2vec, transformer) の発展

### 単語と画像のアナロジー計算

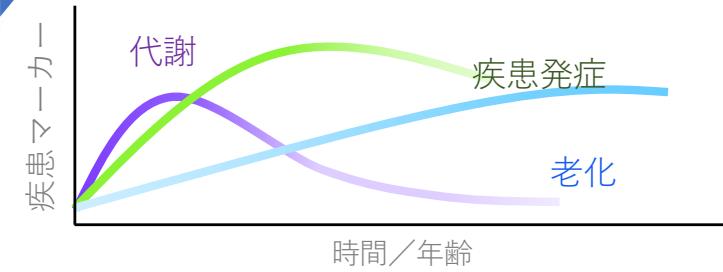
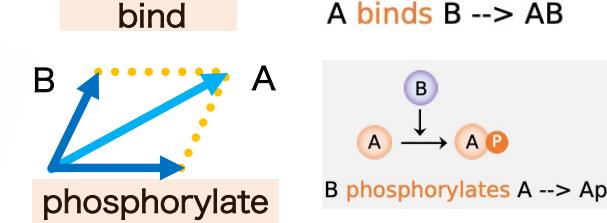
king – man + woman = queen



– day + night =



– brown + white =



生命のダイナミクスの再現  
メカニズム同定・時間発展予測

# Lab of Cell Systems, IPR, Osaka Univ

NLP modeling

**Kiwamu Arakane (M2)**

Modeling platform

**Hiroaki Imoto (海外学振, UCD)**

Keita Iida (Associate prof)

Ayaka Ichikawa (Assistant prof)

Sho Soeda (SA assistant prof)

**Ken Murakami (M.D., D3)**

Masatoshi Haga (D2)

Ririn Rahmala Febri (D1)

**Masato Tsutsui (D1)**

Keita Matsuda (M2)

Hayate Takagishi (M2)

Katsunari Saito (M2)

Muhammad Neanaa (B4)

Former members

Sawa Yamashiro

Hiroki Michida

Johannes Nicolaus Wibisana

Sho Tabata, Kenshiro Nagai



**Collaborators**

**CREST BioDX**

Hidetoshi Shimodaira (Kyoto U)

Makoto Taiji (RIKEN)

**Live cell imaging**

Yasushi Sako (RIKEN)

**Single cell sequence**

Yutaka Suzuki (Tokyo Univ)

**Metabolome**

Takeshi Bamba (Kyusyu Univ)

**Modeling**

Boris Kholodenko (UCD)

Alexander Hoffmann (UCLA)

