

令和 3 年 6 月 23 日現在

機関番号：82626

研究種目：基盤研究(C) (特設分野研究)

研究期間：2018～2020

課題番号：18KT0070

研究課題名(和文) オミクスデータを統合する数理的手法による免疫系モデル構築

研究課題名(英文) Modeling immune system by a mathematical method integrating omics data

研究代表者

熊谷 雄太郎 (Kumagai, Yutaro)

国立研究開発法人産業技術総合研究所・生命工学領域・主任研究員

研究者番号：00528408

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究においては我々が考案したモデリング手法、Stochastic binary modeling (SBM)を適用し、免疫システムの計算、予測可能なモデルの構築を目指した。bulkでのRNA-seqデータと1細胞オミクスデータとを統合できる解析手法を研究し、自然免疫応答を記述しうるモデルの構築が可能となった。さらに、本手法を細胞分化のスナップショットデータに適用し、細胞分化に必須と考えられる因子の候補を同定した。また、細胞間相互作用を考慮することで分岐を示すようなシステムの非線形な挙動も記述しうることを見出した。

研究成果の学術的意義や社会的意義

本研究で開発された手法により免疫システムが簡便にモデル化できるようになった。また、細胞間相互作用を考慮することで分岐を示すような非線形なシステムの挙動もモデル化することが可能性が示唆された。これらの結果は、免疫システムの分子-細胞-個体の階層性に配慮・統合しつつの迅速なモデル構築が可能であることを示唆しており、モデルに基づいた合理的な実験仮説の設定や創薬、医療への道が開けることが期待される。また、全身レベルの免疫応答の計算予測が可能になることも暗示され、疾病の発症や病態の進行、エビデンスに基づく予測要因の検証等の研究へとつながる。

研究成果の概要(英文)：The current study aimed at developing computable and predicatable model of immune system by stochastic binary modeling (SBM), which we have devised. We evolved a method to integrate bulk RNA-seq and single-cell omics data and enabled development of model which describes innate immune response. By applying the method to single-cell dendritic cell development data, we have found several candidate genes though to be important for dendritic cell development. Furthermore, by considering cell-cell interaction, we could describe a non-linear system dynamics with bifurcations.

研究分野：システム生物学

キーワード：免疫システム 数理モデリング

## 1 研究開始当初の背景

免疫システムは、感染症、自己免疫疾患などのみならず、神経系との相互作用、代謝システムとの相互作用、がん免疫等を通じ、種々の疾患に関与し、健康状態全般をコントロールする鍵であることがわかりつつあり、その正確な制御、それを旨とした創薬、医療などは喫緊の課題である。そしてその正確な制御のためには個別化したプロファイルと、それを統一的、合理的な見地から扱うことができるような数理モデルが必要となると考えられるが、残念ながら免疫のような多数の要素が相互作用している高次システムのモデル化はパラメータ爆発等の問題に直面し困難である。

## 2 研究の目的

本研究においては我々が考案したモデリング手法、Stochastic binary modeling (SBM) を適用することで、迅速なモデル化を可能にしてパラメータ爆発の問題を回避し、免疫システムの多種類のオミクスデータからデータドリブンに計算、予測可能なモデルの構築を目指したものである。

SBM においては遺伝子、分子、モジュール等をノードとし、それらの間の相互作用を重み、向き、種類（活性化か不活性化かなど）をもったエッジで表したグラフによってネットワークを記述し、細胞の応答は同定が容易な線形システムとして記述できる。ノードは状態 0/1 をとるものとし、ネットワークによって規定されたダイナミクスにより状態間を確率的に遷移すると考える。これは各状態をとる確率を  $2^N$  次元のベクトル  $p$  で、遷移を  $2^N \times 2^N$  次元の行列  $Q$  として、 $\dot{p} = Qp$  と表現できる。また、観測は  $x = Hp$  とする。系のダイナミクスは  $Q$ ,  $H$  の二つによって完全に説明され、数値計算によって挙動を知ることができるのみならず、これらを同定すればシステムを完全に記述できる。さらに、離散化した時系列のシステム方程式はシステムノイズ、観測ノイズをそれぞれ  $v, w$  として、 $p_{t+1} = f(p_t, Q, v_t), x_t = Hp_t + w_t$  と状態空間モデルとして表現できる。

本研究では状態空間モデルとしての記述を利用し、網羅的 RNA 発現量測定データなどを使用し、データドリブンに種々の免疫細胞のモデルを同定することを目指した。

## 3 研究の方法

本研究においては SBM を用い、システムを状態空間モデルで記述する。これは、免疫細胞内のシグナル伝達分子や遺伝子をノードとし、その相互作用ダイナミクスを確率過程として記述し、その内部状態を何かしらの形で観測することに相当する。前の研究 (Teraguchi et al, Phys Rev E, 2011) において TNF-NF- $\kappa$ B シグナル伝達経路をモデル化した際には、観測はそれぞれのノードが ON になっている周辺確率を計算することに相当していた。しかしながら、この描像においては観測されるシグナル伝達分子、遺伝子ごとにノードが必要となり、システムの次元、そしてパラメータ数が爆発してしまう。そこで、次のようにシステムをモデル化することとした

- (1) 観測は、各ノードそのものとは限らず、各ノードが何かしらの状態をとっている場合（たとえばノード 1, 2, 3 が ON, OFF, ON の場合、ON/OFF, ON, ON の場合など）を観測しているものとする。これにより、ノード数  $N$  に対して観測される状態は  $3^N - 1$  個となる。
- (2) 観測されるダイナミクスは遺伝子や分子それぞれではなく、それらのダイナミクスのパターンとする。例えば遺伝子発現の網羅的時系列定量データを例にとると、数万の遺伝子それぞれを 1 つのノードと捉える

のではなく、そのダイナミクスから  $k$ -means クラスタリング等によって抽出された  $k$  個のパターンが観測されているものとする。これにより、ノードは最大でも  $k \leq 3^N - 1$  を満たす個数  $N$  だけあれば良いこととなる。

この描像を採用することにより、生物学的解釈可能性を残しつつも十分に多様なダイナミクスを記述できる。以下ではこれに基づいて解析を進めた。

## 4 研究成果

### 4.1 システムのノード数推定

SBM においてシステムを決定する重要なパラメータにノード数  $N$  がある。ノード数が多ければ多様なダイナミクスを持つが、一方でパラメータ数が爆発しシステムの同定は困難になる。そこで、まずはシステムのノード数  $N$  を統計的に決定することにより、以降のシステム同定を簡便化することを試みた。

前述の描像においては、システムの内部状態の初期値を  $p_0$  として、時刻  $t$  において  $p(t) = \exp(tQ)p_0$  となり、その時刻で観測される量  $x_k$  は  $x_k(t) = H \cdot \exp(tQ)p_0$  となる。これから、システム行列  $Q$  の固有値  $\lambda_i (i = 1, \dots, D; D \leq 2^N)$  とすると、 $x_k(t) = \sum_i w_{ki} \exp(t\lambda_i)$  とできる。すなわち、観測された量  $x_k$  を用い、 $w_{ki}$  と  $\lambda_i$  を非線形フィッティングにより決定可能であり、 $D$  を変えてフィッティングを行い、AIC を用いてモデル選択を行うことで、ノード数を  $D \leq 2^N$  を満たす数  $N$  と決定できる。

まず、 $Q$  や  $H$  を予め与えシミュレートしたデータを用いて本方法を用い、モデル選択が可能なことを示した。試行した全ケースにおいて、ノード数は本来の数と等しいかまたは 1 個少なく推定された。少なく推定されるのは、観測行列をランダムに与えた際、観測される量の中にシステム行列の固有値のうちダイナミクスに含まれないものが出てくるためと推測された。

次に、樹状細胞を TLR4 リガンドである LPS で刺激したものの遺伝子発現時系列データ (Patil et al, PLoS CB, 2011) に本手法を適用した。まず遺伝子発現データに対し、クラスター数  $k$  を変えながら  $k$ -means クラスタリングを行った。得られたクラスターを用い、上記の手法を用いたところ、 $k \geq 8$  程度にしても最大で  $N = 4$  となった。この結果から、本手法の有効性が示されると同時に、この遺伝子発現ダイナミクスは概ね 4 つのノードがあれば記述しうる可能性が示唆された。

### 4.2 システム行列の同定

前述の方法でノード数  $N$  を同定した後、システム行列  $Q$  の同定を目指した。この際、 $Q$  だけでなく、観測行列  $H$ 、初期状態  $p_0$  も同定する必要がある。しかしながら、前述の描像を採用することにより、 $H$  は特定の内部状態の周辺確率を計算する 0/1 のみを含む行列、初期状態は  $p_0 = {}^t(0, 0, \dots, 0, 1, 0, \dots)$  のように  $2^N$  の内部状態のどれかを取る、と考えることができ、探索は簡便になる。これを用い、 $N = 2, 3$  で  $Q, H$  をランダムに与えシミュレートしたデータを用い、非線形フィッティングによりパラメータ推定を行ったところ、 $N = 2$  ではほぼ正確に、 $N = 3$  においても概ね元のデータを再現しうることを見出した。一点注意すべきこととして、行列として異なってもその固有値  $\lambda_i$  が近い場合、与えた  $Q$  とは違う行列が発見されることである。すなわち、同じダイナミクスを実現しうるシステムが複数存在することを示しており、生物学的解釈を考えた場合、同じダイナミクスを実現しうるシグナル伝達のシステムが複数あることを意味しており興味深い。

しかしながら、ここにおいてもパラメータ数爆発の問題が起こりうるということがわかった。すなわち、

- (1) 上記のように  $H, p_0$  を単純化したとしても、 $H$  は  ${}_{3^{N-1}}C_k$  通り、 $p_0$  は  $2^N$  通りの場合が存在し、 $N = 3, K = 4$  程度でも  $10^5$  オーダーになる。ここで与えた方法では、 $H, p_0$  をランダムにピックアップした後に  $Q$  を非線形フィッティングで計算しており、 $10^5$  オーダーの回数フィッティングを行うことになり現実的でない。
- (2) ノード数決定の項においても述べたように、観測量の組み合わせによっては  $Q$  のダイナミクスが欠落し、推定が難しくなる。
- (3)  $H, p_0$  は所与のものとしても、 $Q$  の推定は  $N \geq 4$  になると探索範囲が広すぎるために初期条件に依存するようになり非常に難しい。

ということが判明した。

これを解決する方法として、マルコフ連鎖モンテカルロ法 (MCMC) の一つ、hybrid Monte Carlo 法 (HMC) を  $Q$  の推定に、 $H, p_0$  の推定には Metropolis 法を利用した。この組み合わせ手法により  $N = 3$  においては  $H, p_0$  を未知としても概ね正確な推定が可能になった。しかしながら、 $N = 3$  の場合においても頻回、 $N \geq 4$  においてはほとんど全てのケースでマルコフ連鎖が収束せず、有効な推定が困難であることがわかった。これをさらに克服する方法として、マルチカノニカル法の利用を検討し研究を継続している。

### 4.3 1 細胞解析との統合

ここまでの結果から、当初目標としていた bulk の網羅的解析データからのモデル同定においては、内部状態を観測に「縮約」する際にデータポイント数の減少、 $Q$  の固有値の脱落等といった問題が生じ、システムの同定が難しくなりうることが示唆された。

近年、1 細胞 RNA-seq (scRNA-seq) の発展により臓器内の種々の細胞集団の同定だけでなく細胞の分化、活性化等の過程の詳細な解析が可能になっている。本来 SBM は 1 細胞レベルの応答の多様性・不均一性・確率性を積極的に活用する枠組みであり、1 細胞データは一つ一つの細胞の状態を確認できるため、SBM の描像ではそれらが各ノードの状態に直接対応している、すなわち内部状態を直接表していると考えられる。これは上記の観測にまつわる問題を回避することが可能なことを意味しており、1 細胞データを活用することによりシステム同定が容易になることが推測できる。

Shalek et al, Nature, 2014 では LPS 刺激後 0, 1, 2, 4, 6h の樹状細胞の 1 細胞 RNA-seq データが取得されている。これを用い、次の手順で解析を行った。

- (1) 全タイムポイントの全細胞のデータを統合し、PCA を行うことで、全タイムポイントで共通の主成分  $\xi$  を計算する。もとの細胞  $c$  のデータを  $x_c$  とすれば、PCA によって決まる回転行列  $R$  を用いて  $\xi_c = Rx_c$  となる。
- (2) 適当な数 ( $M$  とする) の主成分軸において 2 値化を行い、時刻  $t$  において第 1 成分が 1, 第 2 は 0, 第 3 は 0...etc となる細胞の割合を  $q_{010...}^t$  として計算する。ここで、 $M$  はノード数、各  $q$  は SBM の内部状態にあたる。
- (3) 観測行列  $H$  を、各主成分軸への射影 (i.e. 周辺確率を計算している) を  $H_{sc}$  として、 $H = {}^tRH_{sc}$  とした。ただし、第  $M + 1$  以降の主成分に対応する要素は全て 0 とする。

以上で求めた内部状態  $q$  を用い、前述の方法で時間発展を記述する行列  $Q$  の推定を行った。現在解析を継続している。

上記では時系列データを用い内部状態の時系列を定義しシステム同定を試みたが、同様の方法は細胞分化のスナップショットデータに用いることもできる。その場合、各細胞へのラベルを  $l$  として、 $q_{010\dots}^l$  等を同様に計算し、 $Q$  とともに  $l$  の順序を推定することとなる。この解析により、システムの時間発展ダイナミクス  $Q$  とともに各細胞の分化の順序が推定できる。本手法を Balan et al, Cell Rep, 2018 や Brown et al, Cell, 2019 のデータ、また in house のデータに適用することで、樹状細胞分化に必須と推定される遺伝子の候補をいくつか同定し、研究を継続している。

#### 4.4 多細胞の結合による非線形システム化

研究協力者の寺口俊介さんは、SBM に細胞間相互作用を導入することによりシステムが非線形となり複雑なダイナミクスを記述できることを見出した。このモデルの研究を共同で進め、あるネットワークを持つシグナル伝達と液性因子の産生システムのダイナミクスはトランスクリプショナル分岐を持ち、液性因子の産生パターンが大きく変化するを見出した。また、液性因子とその受容体の相互作用に非線形性を取り入れることで、ダイナミクスが Hopf 分岐を持ち、細胞の状態がリミットサイクルになることがわかった。この結果は、個々の細胞は確率的に状態を変化させるにもかかわらず、細胞間相互作用によって系全体が”安定”に振動することを意味しており、生物学的にも興味深いため、実際にそのようなシステムが実現されているかを探索している。

## 5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件/うち国際共著 2件/うちオープンアクセス 5件）

1. 著者名 Miyazaki Yoshihiro, Oda Tatsuya, Inagaki Yuki, Kushige Hiroko, Saito Yutaka, Mori Nobuhito, Takayama Yuzo, Kumagai Yutaro, Mitsuyama Toutai, Kida Yasuyuki S.	4. 巻 11
2. 論文標題 Adipose-derived mesenchymal stem cells differentiate into heterogeneous cancer-associated fibroblasts in a stroma-rich xenograft model	5. 発行年 2021年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 4690
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-021-84058-3	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Maruyama Kenta, Kidoya Hiroyasu, Takemura Naoki, Sugisawa Erika, Takeuchi Osamu, Kondo Takeshi, Eid Mohammed Mansour Abbas, Tanaka Hiroki, Martino Mikael M., Takakura Nobuyuki, Takayama Yasunori, Akira Shizuo, Vandenbon Alexis, Kumagai Yutaro	4. 巻 32
2. 論文標題 Zinc Finger Protein St18 Protects against Septic Death by Inhibiting VEGF-A from Macrophages	5. 発行年 2020年
3. 雑誌名 Cell Reports	6. 最初と最後の頁 107906 ~ 107906
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.celrep.2020.107906	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 Sugisawa Erika, Takayama Yasunori, Takemura Naoki, Kondo Takeshi, Hatakeyama Shigetsugu, Kumagai Yutaro, Sunagawa Masataka, Tominaga Makoto, Maruyama Kenta	4. 巻 182
2. 論文標題 RNA Sensing by Gut Piezo1 Is Essential for Systemic Serotonin Synthesis	5. 発行年 2020年
3. 雑誌名 Cell	6. 最初と最後の頁 609 ~ 624.e21
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.cell.2020.06.022	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Maruyama Kenta, Kumagai Yutaro, Tominaga Makoto他	4. 巻 6
2. 論文標題 The ATP Transporter VNUT Mediates Induction of Dectin-1-Triggered Candida Nociception	5. 発行年 2018年
3. 雑誌名 iScience	6. 最初と最後の頁 306 ~ 318
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.isci.2018.08.007	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Takayama Yuzo, Kushige Hiroko, Akagi Yuka, Suzuki Yutaka, Kumagai Yutaro, Kida Yasuyuki S.	4. 巻 10
2. 論文標題 Selective Induction of Human Autonomic Neurons Enables Precise Control of Cardiomyocyte Beating	5. 発行年 2020年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 9464
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-020-66303-3	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Vandenbon Alexis, Kumagai Yutaro, Lin Mengjie, Suzuki Yutaka, Nakai Kenta	4. 巻 19
2. 論文標題 Waves of chromatin modifications in mouse dendritic cells in response to LPS stimulation	5. 発行年 2018年
3. 雑誌名 Genome Biology	6. 最初と最後の頁 138
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s13059-018-1524-z	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計4件 (うち招待講演 4件 / うち国際学会 0件)

1. 発表者名 熊谷雄太郎
2. 発表標題 術後遷延性疼痛の分子機構解析
3. 学会等名 第126回日本解剖学会総会・全国学術集会 / 第98回日本生理学会大会合同大会 (招待講演)
4. 発表年 2021年

1. 発表者名 熊谷雄太郎
2. 発表標題 免疫システムの情報処理機構 - 細胞の情報伝達と全体最適化
3. 学会等名 第5回理論免疫学ワークショップ (招待講演)
4. 発表年 2021年

1. 発表者名 熊谷雄太郎
2. 発表標題 免疫系の情報処理機構
3. 学会等名 第4回理論免疫学ワークショップ（招待講演）
4. 発表年 2020年

1. 発表者名 熊谷雄太郎
2. 発表標題 免疫システムの情報処理機構
3. 学会等名 第3回理論免疫学ワークショップ（招待講演）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------