

令和元年6月12日現在

機関番号：82401

研究種目：基盤研究(C) (特設分野研究)

研究期間：2016～2018

課題番号：16KT0196

研究課題名(和文) 生命システムの破綻にひそむ秩序を解明する

研究課題名(英文) Elucidate latent rules in disruption of biological system

研究代表者

川上 英良 (Kawakami, Eiryo)

国立研究開発法人理化学研究所・科技ハブ産連本部・ユニットリーダー

研究者番号：30725338

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：世界中のヒト・マウスのChIP-seqのデータを再解析することで、ヒトとマウス両方に対応した遺伝子制御ネットワークを構築した。医学・生物学で見られる、散発的で不均一な時系列データに対して、多項目測定データに基づいて状態を地形として表現する手法を用いて患者の状態の層別化と状態変化予測アルゴリズムの構築を行った。機械学習を導入することで卵巣腫瘍に関して術前の血液検査データに基づく予後予測アルゴリズムの構築も行った。今後は、健康診断データなど、疾病発症前のデータを対象とすることで、未病の状態、疾病予備軍の状態を発見し、疾病の発症を予見・予防する研究を展開して行きたいと考えている。

研究成果の学術的意義や社会的意義

膨大な実験データに基づく遺伝子制御ネットワークは、既に様々な生命現象の遷移過程を明らかにしており(関連論文6報)、オミクスデータからの情報抽出、仮説形成に今後も活用されていくと考えられる。機械学習を用いた精度の高い診断・予後予測は、治療方針を決めるのに役立つ情報を得るのに活用されると考えられる。また、機械学習や数理科学により、既存の分類や数値などを予測するだけでなく、今まで臨床医や実験研究者も気づかなかった複雑なパターンを発見することを示せたのも大きな成果である。

研究成果の概要(英文)：We reconstructed human and mouse gene regulatory network based on the enormous public ChIP-seq data. For sporadic and unequal time series data often obtained in medical/biological research, we introduced an algorithm that represent states of data as "landscape" based on multidimensional variables. Utilizing the method, we could stratify patient states and predict state transition of disease progression. We also applied machine learning for the accurate diagnostic and prognostic prediction for epithelial ovarian cancer based on preoperative blood test data. Hereafter, we would like to expand our research aiming at forecasting and preventing disease development by identifying presymptomatic disease state based on health check data and real-world measurement data obtained from various IoT devices.

研究分野：システム医学

キーワード：多階層ネットワーク 制御ネットワーク 状態遷移モデル 時系列解析 機械学習

## 様式 C-19、F-19-1、Z-19、CK-19（共通）

### 1. 研究開始当初の背景

生命は生体、組織、細胞などの多階層にまたがる動的恒常性を維持しており、近年多くの疾病がこの動的恒常性の乱れや破綻として説明できることが明らかになってきている。しかしながら、疾病状態においても生命システムは一定の動的恒常性を維持しており、正常な状態とは異なるものの、疾病は生命システムの一つの安定状態と考えることができる。したがって、疾病の発症もシステムの無秩序な崩壊ではなく、一定の規則・秩序に従った「正常な安定状態」から「異常な安定状態」への遷移過程として捉えられる。正常な安定状態から異常な安定状態への遷移過程にひそむ秩序・規則を明らかにすることは、疾病の発症メカニズムの理解に重要であるだけでなく、疾病発症及び病態進行の予測を行なう上で理論的土台になると考えられる。

近年、反応の詳細なパラメータを全て決定するのが難しい大規模ネットワークにおいて、分岐が生じるパラメータの臨界点付近で構成要素の挙動の傾向が類似してくることを利用して分岐を検出する手法が提案されている (Chen et al. 2012, Moon et al. 2015)。しかし、これらの大規模ネットワーク状態遷移理論は生命ネットワークに見られるような質の異なる階層構造や反応タイムスケールの違いは考慮していない。一方、近年の次世代シーケンシング技術と質量分析法の進歩により、トランスクリプトーム、プロテオーム、メタボロームを始めとした各種網羅的オミクスデータが膨大に蓄積されている。これらの網羅的オミクスデータを細胞内で起こっているプロセスに立脚して取り込むことは、先行研究では考慮されていない、タイムスケールの違い、多階層性、階層間フィードバックを反映した大規模ネットワーク状態遷移理論の基盤として極めて重要であると考えられる。

### 2. 研究の目的

近年膨大に蓄積されつつある多種オミクスデータを取り入れることで制御の強度や構造を考慮した多階層ネットワークの構築を行う。また、階層間フィードバック・時間スケールを考慮した多階層ネットワーク解析手法を確立し、ネットワーク状態遷移理論を用いて疾病発症に至る状態遷移ルートを同定することで、分子レベルのダイナミクスと大域的な状態遷移をつなぎ、疾病の発症を予測する汎用的な理論研究手法を確立する。

### 3. 研究の方法

#### **[遺伝子制御ネットワーク] ChIP-seq および CLIP-seq データを制御の強度として取り入れた転写因子・マイクロRNA 遺伝子制御ネットワークの構築**

転写因子-遺伝子制御ネットワーク構築時に確立した大量の次世代シーケンシングデータ解析ワークフローを利用することで、近年普及しつつあるマイクロRNAとそのターゲットである mRNA の結合を網羅的に検出する CLIP-seq のデータを取り入れ、統合的な遺伝子制御ネットワークを構築する。

#### **[タンパク質相互作用ネットワーク] 局在情報およびプロテオミクスデータを取り入れ相互作用の確からしさと構造を考慮したタンパク質相互作用ネットワークの構築**

複数の相互作用スクリーニング結果を統合して取り入れることで「相互作用の確からしさ」を算出、タンパク質の細胞内局在情報を UniProt データベースから取り入れる、タンパク質修飾を網羅的に測定したプロテオミクスデータを取り入れることで相互作用の向きや制御の強度を反映し、実際の細胞内プロセスに立脚した PPI を構築する。

#### **[多階層ネットワーク] 遺伝子制御ネットワーク・タンパク質相互作用ネットワークおよび代謝ネットワークを統合した多階層ネットワークの構築**

遺伝子制御ネットワーク、タンパク質相互作用ネットワークに加え、既に申請者らのグループで構築されている代謝ネットワークを統合し、多階層ネットワークを構築する。

#### **[ネットワーク解析] 階層間フィードバック・タイムスケールを考慮した多階層ネットワーク解析手法の確立**

階層間のタイムスケールが異なるフィードバックを考慮した多階層ネットワーク解析手法を確立する。

#### **[状態遷移モデル] 疾病発症に対応するネットワーク状態遷移モデルの確立**

モジュール化した多階層ネットワークを用いて、疾病に至る状態遷移過程を予測するモデルを確立する。

### 4. 研究成果

疾病発症の予測モデルとしてのネットワークに基づく状態遷移理論構築の基盤として、1. 遺伝子制御ネットワークの拡張、2. 実際の細胞内プロセスに立脚したタンパク質相互作用ネットワークの構築、3. 不均一で散発的な時系列データ解析の枠組みの整備、を行った。

#### **1. 遺伝子制御ネットワークの拡張**

研究代表者のグループは既に 3500 以上の世界中のマウスの ChIP-seq のデータを再解析することで、従来考慮されて来なかった制御の強度を反映した遺伝子制御ネットワークを構築した。マウスのネットワークに加え、3500 以上のヒトのデータを加え、ヒトとマウス両方に対応した遺伝子制御ネットワークを構築した。また、研究代表者らは、これらの遺伝子制御ネットワークに基づき、上流の制御転写因子を推定する手法 (wPGSA 法)

を開発した (Kawakami et al. 2016 *Nucleic Acids Res.*)。この手法を実際の実験データに適用することで、様々な生命現象の遷移過程を明らかにした。転写制御およびタンパク質相互作用ネットワークに基づくトランスクリプトーム解析によって、1. 制御性 T 細胞における BATF の役割 (Hayatsu et al. 2017 *Immunity*)、2. Langerhans 細胞における TGFβ receptor signaling の役割 (Tenno et al. 2017 *J. Exp. Med.*)、3. B 細胞分化過程における転写制御の変化 (Miyai et al. 2018 *Genes & Dev.*)、4. Germinal Center における B 細胞の転写制御変化 (Ise et al. 2018 *Immunity*)、5. 大腸がんにおける制御性 T 細胞の応答 (Fujimoto et al. 2018 *J. Immunology*) を明らかにした。また、GWAS データと転写制御ネットワークを組み合わせることで、Amyopathic Dermatomyositis においてリスクとなる転写制御変化を明らかにした (Kochi et al. *Ann. Rheum. Dis.* 2018)。

## 2. 実際の細胞内プロセスに立脚したタンパク質相互作用ネットワークの構築

網羅的スクリーニングに基づいたタンパク質相互作用ネットワーク (PPI ネットワーク) を用いて、ヒトとマウスのシグナル伝達モデルの網羅的再構成を行った。イーストツーハイブリッド法や免疫共沈降-質量分析に基づく PPI ネットワークには多くの非特異的な相互作用が含まれている。そこで、タンパク質の細胞内局在や発現量の情報を取り入れることで、非特異的な相互作用を排除し、従来知られているシグナル伝達パスウェイの多くを含んだ網羅的シグナル伝達ネットワークモデルを構築することに成功した。この成果は、第 89 回日本生化学会大会シンポジウムにおいて発表を行った。構築したタンパク質相互作用ネットワークを用いて、上流制御因子探索アルゴリズムを開発中である。

## 3. 不均一で散発的な時系列データ解析の枠組みの整備

機械学習に基づく次元削減・圧縮手法として、ランダムフォレスト法を用いた距離計算 (Random Forest distance: RF dist, Shi and Horvath 2006) を応用した、高次元データの次元圧縮と教師なしクラスタリング手法を整備した。医学・生物学データは分布が正規分布しないものも多く、症状の有無といった二値変数や、癌の組織型のようなカテゴリ変数も存在するため、PCA のような正規分布を前提にした距離計算やクラスタリング手法では、データの特徴を適切に捉えることは難しい。RF dist はランダムフォレスト法で構成される多数の決定木において、サンプルが同じ葉に分類される頻度からサンプルの類似度を計算する。RF dist を用いることで、変数の分布によらず外れ値に強いクラスタリングが可能になった。実際に、RF dist を卵巣がんの診断時年齢と術前血液検査 32 項目に適用した結果、従来早期がんと一律に診断されていた患者が、「良性腫瘍に近いタイプ」と「末期がんに近いタイプ」に分かれることが明らかになった (図 1 A)。末期がんに近いタイプの早期がんは、良性腫瘍に近いタイプに比べて再発のリスクが有意に高いことが示され (図 1 B)、RF dist を用いた教師なしクラスタリングが疾患の特徴をよく捉えられることが示唆された (Kawakami et al. 2019 *Clin. Cancer Res.*)。

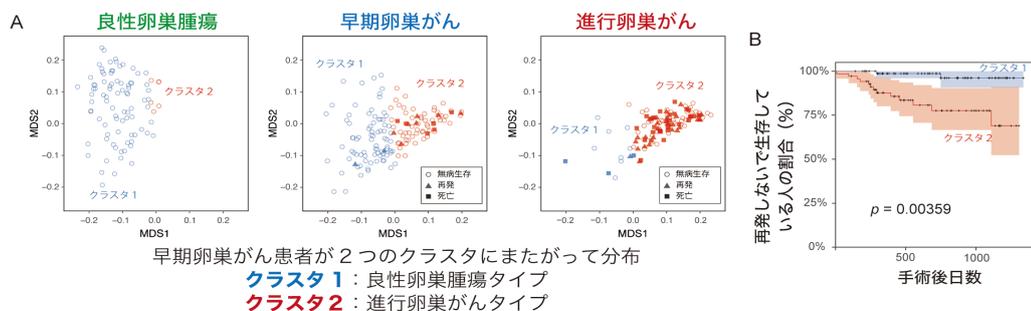


図 1. 術前血液検査データに基づく卵巣がんの教師なしクラスタリング

また、高次元で短い時系列データに対しては、物理学のイジングモデルに基づいて状態を記述するエネルギーランドスケープ分析 (Ezaki et al. 2016) を導入した。本手法は、従来エピソードランダムスケープを始めとして、生物学において概念的に使われてきたエネルギー (安定性) をデータから計算し、状態遷移のパターンや傾向を明らかにすることができる。実際に、本手法を更年期障害患者の経時的質問票データ (初診時、1ヶ月後、4ヶ月後の3タイムポイント) に適用することで、更年期障害の状態と患者ごとの状態遷移を可視化することができた (図 2)

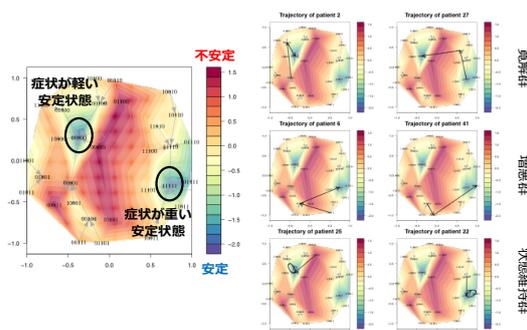


図 2. エネルギーランドスケープモデルによる更年期障害患者の状態遷移の可視化

## 5. 主な発表論文等

〔雑誌論文〕（計 13 件 すべて査読有り）

- (1). **Kawakami E**, Tabata J, Yanaihara N, Ishikawa T, Koseki K, Iida Y, Saito M, Komazaki H, Shapiro JS, Goto C, Akiyama Y, Saito R, Saito M, Takano H, Yamada K, Okamoto A., Application of Artificial Intelligence for Preoperative Diagnostic and Prognostic Prediction in Epithelial Ovarian Cancer Based on Blood Biomarkers., *Clin. Cancer Res.*, DOI: 10.1158/1078-0432.CCR-18-3378, 2019
- (2). Ono R, Watanabe T, **Kawakami E**, Iwasaki M, Tomizawa-Murasawa M, Matsuda M, Najima Y, Takagi S, Fujiki S, Sato R, Mochizuki Y, Yoshida H, Sato K, Yabe H, Kato S, Saito Y, Taniguchi S, Shultz LD, Ohara O, Amagai M, Koseki H, Ishikawa F., *EBioMedicine.*, 41, 584-596, DOI: 10.1016/j.ebiom.2019.02.001, 2019
- (3). Ackerman EE, **Kawakami E**, Katoh M, Watanabe T, Watanabe S, Tomita Y, Lopes TJ, Matsuoka Y, Kitano H, Shoemaker JE, Kawaoka Y., *American Society for Microbiology*, 9, DOI: 10.1128/mBio.02002-18, 2018
- (4). Sakaue S, Hirata J, Maeda Y, **Kawakami E**, Nii T, Kishikawa T, Ishigaki K, Terao C, Suzuki K, Akiyama M, Suita N, Masuda T, Ogawa K, Yamamoto K, Saeki Y, Matsushita M, Yoshimura M, Matsuoka H, Ikari K, Taniguchi A, Yamanaka H, Kawaji H, Lassmann T, Itoh M, Yoshitomi H, Ito H, Ohmura K, R Forrester AR, Hayashizaki Y, Carninci P, Kumanogoh A, Kamatani Y, de Hoon M, Yamamoto K, Okada Y., Integration of genetics and miRNA-target gene network identified disease biology implicated in tissue specificity., *Nucleic acids research*, 46, 11898-11909, 2018
- (5). Ise W, Fujii K, Shiroguchi K, Ito A, Kometani K, Takeda K, **Kawakami E**, Yamashita K, Suzuki K, Okada T, Kurosaki T., T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling Germinal Center Cell Fate., *Immunity*, 48, 702-715, e4, 2018
- (6). Fujimoto H, Saito Y, Ohuchida K, **Kawakami E**, Fujiki S, Watanabe T, Ono R, Kaneko A, Takagi S, Najima Y, Hijikata A, Cui L, Ueki T, Oda Y, Hori S, Ohara O, Nakamura M, Saito T, Ishikawa F., Deregulated Mucosal Immune Surveillance through Gut-Associated Regulatory T Cells and PD-1+ T Cells in Human Colorectal Cancer., *The Journal of Immunology*, 200, 3291-3303, 2018
- (7). Miyai T, Takano J, Endo TA, **Kawakami E**, Agata Y, Motomura Y, Kubo M, Kashima Y, Suzuki Y, Kawamoto H, Ikawa T., Three-step transcriptional priming that drives the commitment of multipotent progenitors toward B cells., *Genes Dev.*, 32, 112-126, 2018
- (8). Kochi Y, Kamatani Y, Kondo Y, Suzuki A, **Kawakami E**, Hiwa R, Momozawa Y, Fujimoto M, Jinnin M, Tanaka Y, Kanda T, Cooper RG, Chinoy H, Rothwell S, Lamb JA, Vencovský J, Mann H, Ohmura K, Myouzen K, Ishigaki K, Nakashima R, Hosono Y, Tsuboi H, Kawasumi H, Iwasaki Y, Kajiyama H, Horita T, Ogawa-Momohara M, Takamura A, Tsunoda S, Shimizu J, Fujio K, Amano H, Mimori A, Kawakami A, Umehara H, Takeuchi T, Sano H, Muro Y, Atsumi T, Mimura T, Kawaguchi Y, Mimori T, Takahashi A, Kubo M, Kohsaka H, Sumida T, Yamamoto K., Splicing variant of WDFY4 augments MDA5 signalling and the risk of clinically amyopathic dermatomyositis., *Annals of the rheumatic diseases*, annrheumdis-2017-212149, 2018
- (9). **Kawakami E**, Adachi N, Senda T, Horikoshi M., Leading role of TBP in the Establishment of Complexity in Eukaryotic Transcription Initiation Systems., *Cell Reports*, 21, 3941-3956, 2017
- (10). Tenno M, Shiroguchi K, Muroi S, **Kawakami E**, Koseki K, Kryukov K, Imanishi T, Ginhoux F, Taniuchi I., Cbfb2 deficiency preserves Langerhans cell precursors by lack of selective TGFβ receptor signaling., *J. Exp. Med.*, 214, 2933-2946, 2017
- (11). Hayatsu N, Miyao T, Tachibana M, Murakami R, Kimura A, Kato T, **Kawakami E**, Endo TA, Setoguchi R, Watarai H, Nishikawa T, Yasuda T, Yoshida H, Hori S., Analyses of a Mutant Foxp3 Allele Reveal BATF as a Critical Transcription Factor

in the Differentiation and Accumulation of Tissue Regulatory T Cells., *Immunity*, 47, 268-283.e9, 2017

- (12). Kiso M, Iwatsuki-Horimoto K, Yamayoshi S, Uraki R, Ito M, Nakajima N, Yamada S, Imai M, **Kawakami E**, Tomita Y, Fukuyama S, Itoh Y, Ogasawara K, Lopes TJS, Watanabe T, Moncla LH, Hasegawa H, Friedrich TC, Neumann G, Kawaoka Y., Emergence of oseltamivir-resistant H7N9 influenza viruses in immunosuppressed cynomolgus macaques., *The Journal of infectious diseases*, 216, 582-593, 2017
- (13). **Kawakami E\***, Nakaoka S, Ohta T, Kitano H., Weighted enrichment method for prediction of transcription regulators from transcriptome and global chromatin immunoprecipitation data., *Nucleic Acids Res.*, 44, 5010-21, 2016 (\*Corresponding author)

〔学会発表〕 (計 24 件)

#### シンポジウム

- (1). Evaluate Cellular Stability and Transition Based on Single-cell Transcriptome Data **Eiryo Kawakami**, Single Cell Science Symposium 2018 (東京) (2018)
- (2). 機械学習・数理科学で疾患の多様性と個別性に迫る 川上英良, 第 30 回高遠・細胞分子生物学シンポジウム (長野) (2018)
- (3). 機械学習・数理科学で疾患の多様性と個別性に迫る 川上英良, Society 5.0 時代のヘルスケアの在り方 (東京) (2018)
- (4). Discovery of combinatorial biomarkers for personalized medicine based on machine learning 川上英良, 医科学イノベーションハブ第一回国際シンポジウム (東京) (2017)
- (5). 公共 ChIP データを活用した遺伝子制御ネットワーク構築と転写因子予測 川上英良, トーゴの日シンポジウム(東京) (2016)

#### 国際学会、国外発表

- (1). Data-driven systems medicine to elucidate and predict heterogeneous disease **Eiryo Kawakami**, South Lake Workshop in Bioinformatics and System Biology (Wuhan, China) (2018)
- (2). Approaches based on machine learning and mathematical science towards precision/individualized medicine **Eiryo Kawakami**, Joint RIKEN-Luxembourg Symposium (Luxembourg, Luxembourg) (2018)
- (3). Dynamic transition of the human skin microbial patterns associated with atopic dermatitis pathology **Eiryo Kawakami**, Hiroshi Kawasaki, Shoko Obata, Aki Honda, Naoko Mochimaru, Ayano Fukushima, Fumiyo Yasuda-Sekiguchi, Takashi Sasaki, Wataru Suda, Kenya Honda, Tamotsu Ebihara, Masayuki Amagai Cell Symposia Next Gen Immunology (Rehovot, Israel) (2018)

#### 国内における一般講演

- (1). 機械学習・数理科学にもとづく疾患の層別化と予測 川上英良, 千里ライフサイエンスセミナーM5 (大阪) (2019)
- (2). 予測・個別化医療に向けたトランスオミクス数理解析 川上英良, 第 41 回日本分子生物学会年会 (神奈川) (2018)
- (3). A machine learning approach based on preoperative blood makers for diagnostic and prognostic prediction of epithelial ovarian cancer 川上英良, 第 60 回日本婦人科腫瘍学会学術講演会 (京都) (2018)
- (4). 医学生物学研究におけるデータ駆動型アプローチ 川上英良, 第 9 回数理解デザイン道場 (CREST 生命動態) (沖縄) (2017)
- (5). ウイルス・細菌感染における宿主転写制御 川上英良, 第 90 回 日本細菌学会総会 (宮城) (2017)

- (6). 医学生物学研究におけるデータ駆動型アプローチ  
川上英良, 第三回生命情報研究会 (京都) (2017)
- (7). 遺伝子をまとめて見る～遺伝子発現データの Enrichment 解析  
川上英良, 第二回生命情報研究会 (京都) (2017)
- (8). 複合体が生み出すシグナル伝達確率的ダイナミクスの制御機構  
川上英良, 第 39 回 日本分子生物学会年会 (神奈川) (2016)
- (9). 生物ネットワークに潜む情報を探る: タンパク質相互作用ネットワークに基づく  
シグナル伝達パスウェイ網羅的再構成  
川上英良, 第 89 回 日本生化学会大会 (宮城) (2016)

他 7 件

〔図書〕 (計 3 件)

- (1). wPGSA 法 –公共 ChIP-seq データを用いて転写因子の影響を推定する –  
川上英良\*, 梶下紘貴  
遺伝子医学 MOOK Vol.33 201-207 頁 (2018)
- (2). アトピー性皮膚炎のリバーストランスレショナルリサーチ  
川崎洋\*, 川上英良, 天谷雅行, 古関明彦  
実験医学 Vol.35 No.1 2-29 頁 (2017)
- (3). 公共ハイスループット ChIP データを用いた遺伝子制御ネットワークの構築とその活用法  
川上英良\*, 大田達郎  
実験医学増刊 Vol.35 No.5 33-39 頁 (2017)

〔産業財産権〕

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

〔その他〕

ホームページ等

- (1). wPGSA online (大規模 ChIP-seq データに基づく転写制御予測)  
<http://wpgsa.org/>
- (2). 機械学習で卵巣腫瘍の特性を術前予測 (理化学研究所プレスリリース)  
[http://www.riken.jp/pr/press/2019/20190415\\_1/](http://www.riken.jp/pr/press/2019/20190415_1/)

## 6. 研究組織

### (1)研究分担者

研究分担者氏名: 中岡慎治 (2017/3/7 まで ※さきがけ専任研究員として採用されたため削除)

ローマ字氏名: Nakaoka Shinji

所属研究機関名: 東京大学

部局名: 生産技術研究所

職名: 特任助教

研究者番号 (8 桁): 30512040

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。