

令和 4 年 6 月 15 日現在

機関番号：12602

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20402

研究課題名（和文）エビデンスに基づくパーソナルゲノム医療・予防実現に向けた統計モデリング技術開発

研究課題名（英文）Statistical Modeling for Evidence based Personalized Medicine

研究代表者

PARK HEEWON (Park, Heewon)

東京医科歯科大学・M&Dデータ科学センター・教授

研究者番号：70756642

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本応募研究では、ゲノムデータ解析からEvidence Based Personalized Anti-Cancer Therapyへの治療のエビデンス提供を目指し、そのための統計モデリング技術開発を行った。特に、個別化医療へのエビデンス抽出のための遺伝子ネットワーク解析技術を開発した。開発された一細胞レベル・患者個々人の情報抽出を可能にする解析技術から、疾患の状況・臨床情報による異なる情報を遺伝子ネットワークから抽出した。癌のメカニズム解明に応用し、関連重要マーカー探索や疾患のメカニズムにおけるそれらの因果関係を抽出した。

研究成果の学術的意義や社会的意義

個別化治療のためには、患者個々のゲノム情報の究明から患者個人に最適化された治療法選択は重要であり、そこで患者個々の治療のエビデンスの抽出は治療効果を向上させるために必須である。しかし、従来の手法に基づいては患者個々の情報を取り出すことができないため、患者個々のゲノム特性が究明可能な新たなデータ解析技術の必要性が強く認識されていた。本研究で行った個別化医療へのエビデンス提供を目指す解析技術開発は、患者個々人の疾患関連情報を可能にするデータ解析技術であり、従来手法の限界を超え、個別化治療へ適切な治療エビデンスの提供を可能にすることで、治療効果を向上に貢献できると期待される。

研究成果の概要（英文）：We developed statistical methodologies for Evidence Based Personalized Anti-Cancer Therapy. Especially, we focused on cell-line specific gene network analysis. We then extracted cancer characteristic specific gene networks by using the developed statistical approaches. We applied the methods to uncover complex mechanism of cancer, and extracted related biomarkers and their regulatory system that involved in mechanism of cancer.

研究分野：統計科学、Bioinformatics

キーワード：遺伝子ネットワーク 個別化医療 統計モデリング

1. 研究開始当初の背景

個別化医療が世界的に注目されていて、分子標的治療等の色々な研究が様々な分野で進んでいる。がん治療での統計・数理科学の役割はがんゲノムデータから治療のエビデンスを取り出すこと（または、取り出せるデータ解析技術の開発）であり、そのため色々な統計手法が開発・活用されている。個別化医療に向けた患者個々の特性を纏めるためには、細胞（または、患者）個々の特徴に従って変化するモデルの構築が極めて重要である。しかしながら、従来の研究では、全体のデータ（すべての細胞・患者集団）から平均的な1つのモデルを構築し、それらの解釈からなる研究であったため、細胞（患者）の多様な特徴によって異なる疾患のメカニズムを精密に解明するには不十分であった。

2. 研究の目的

本研究の目的は、多様なオミクスデータ解析から個別化医療への治療エビデンスを提供することを目指し、そのための統計モデリング技術開発研究の推進することであった。全体のデータから一つのモデルを構築する従来の研究の限界を超えるため、一細胞レベルのモデル（患者個人によって異なるモデル）構築を可能にする統計モデリング技術を開発をメイン研究テーマとして研究を進んで来た。特に遺伝子ネットワーク解析に注目した。がん細胞を含めすべての細胞は、遺伝子同士が複雑なネットワークを形成しており、ある遺伝子が別の遺伝子に指令を出すことで、遺伝子の発現を制御している。したがって、疾患のメカニズムを解明するためには、単に関連している遺伝子を特定するだけでは不十分であり、関連している遺伝子同士がどのようなネットワークを形成しているのか、そのネットワーク構造の中で、特に重要な役割を果たしている遺伝子同士の因果関係はどれなのかなどを明らかにするのは重要である。本研究では、一細胞レベルの遺伝子ネットワーク推定技術を開発し、そこから癌などの疾患のメカニズムの解明を目指して研究を行った。

3. 研究の方法

一細胞レベルのモデル構築を可能にする統計的機械学習法に関する研究を行った。Sample-specific analysis を可能にする varying coefficient model のスパース化に基づいて、一細胞レベルの遺伝子ネットワークを推定し、推定した膨大・大量のネットワークを解釈するための、機械学習・AI モデルに関する研究を行った。開発された技術は人口データに基づく検証を行った。また、がん細胞における遺伝子発現量データ、細胞のがんの特徴値として EMT-score や抗がん剤感受性データを用いて癌の特徴によって変化する遺伝子ネットワークを推定し、その膨大・大量のネットワークの解釈から遺伝子マーカー探索や鍵になる遺伝子間の因果関係を抽出した。

4. 研究成果

一細胞レベルの遺伝子ネットワーク推定

患者個々の heterogeneity によるモデリングのため、sample-specific analysis を想定し、サンプルのある特性によってグルーピングして、似たようながんの特性を持つ患者集団の情報だけに基づいてモデリングする varying coefficient model のスパース化に関する研究を行った。そこから、各細胞、または各患者が持つ独特な疾患のリスクファクターの探索や診断・予防・治療へのエビデンスを遺伝子発現制御関係から抽出した。関する研究成果は以下の論文で発表した。

H Park, K Maruhashi, R Yamaguchi, S Imoto, S Miyano (2020). Global gene network exploration based on explainable artificial intelligence approach, PLOS ONE, 15, e0241508.

抗がん剤の耐性獲得メカニズム解明を目指すネットワーク解析

開発された手法に基づいて、薬剤感受性に従って変化する遺伝子ネットワークを推定し、ある抗がん剤に耐性を持つ細胞と感受性を持つ細胞のネットワークをそれぞれ Tensor データとしてまとめ、2つ Tensor データの異質性検定に基づいた遺伝子マーカー探索やそれらの因果関係を抽出のための統計・数理モデリング技術開発に関する研究を行った。Tensor ネットワーク解析のため、複数のネットワークの共通成分を抽出し、その共通成分を解釈することから、薬剤耐性を持つ細胞と感受性を持つ細胞のネットワークの特徴を取り出し、がん細胞が薬剤耐性・感受性を獲得するメカニズムを遺伝子ネットワークから解明した。また、説明可能な AI 予測モデルを開発し、薬剤耐性・感受性予測の鍵になる遺伝子とその果関係を探索した。関する研究成果は以下の論文で発表した。

K Maruhashi, H Park, R Yamaguchi, S Miyano (2020). Linear Tensor Projection Revealing Nonlinearity, arXiv: 12007.03912.

H Park, K Maruhashi, R Yamaguchi, S Imoto, S Miyano (2020). Global gene network exploration based on explainable artificial intelligence approach, PLOS ONE, 15, e0241508.

H Park, K Maruhashi, R Yamaguchi, S Miyano (2021). 説明可能な AI によるパーソナルゲノム医療・予防の実現, 医学のあゆみ, 276(9): 849-853.

H Park, R Yamaguchi, S Imoto, S Miyano (2022). Uncovering Molecular Mechanisms of Drug Resistance via Network-Constrained Common Structure Identification, Journal of Computational Biology, 29: 257-275.

H Park, R Yamaguchi, S Imoto, S Miyano (2022). Xprediction: Explainable EGFR-TKIs response prediction based on drug sensitivity specific gene networks, PLOS ONE, 17, e0261630.

Sparse Genetic relationship estimation for GWAS

Genetic relationship matrix 推定の精度を高めるため、SNP 選択に基づくランダム効果の分散共分散行列のスパース推定に関する研究を行った。また、そこから固定・ランダム効果の選択とモデルの推定が同時にできるスパース線形混合モデリング (スパース主成分分析技術の改善版) の技術を開発した。関する研究成果は以下の論文で発表した。

H Park, R Yamaguchi, S Imoto, S Miyano (2020). Automatic sparse principal component analysis, Canadian Journal of Statistics, 49: 678-697.

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 4件／うち国際共著 0件／うちオープンアクセス 4件）

1. 著者名 H Park, R Yamaguchi, S Imoto, S Miyano	4. 巻 49
2. 論文標題 Automatic sparse principal component analysis	5. 発行年 2020年
3. 雑誌名 Canadian Journal of Statistics	6. 最初と最後の頁 678 ~ 697
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cjs.11579	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 H Park, K Maruhashi, R Yamaguchi, S Imoto, S Miyano	4. 巻 15
2. 論文標題 Global gene network exploration based on explainable artificial intelligence approach	5. 発行年 2020年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0241508
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0241508	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 H Park, R Yamaguchi, S Imoto, S Miyano	4. 巻 29
2. 論文標題 Uncovering Molecular Mechanisms of Drug Resistance via Network-Constrained Common Structure Identification	5. 発行年 2022年
3. 雑誌名 Journal of Computational Biology	6. 最初と最後の頁 257 ~ 275
掲載論文のDOI（デジタルオブジェクト識別子） 10.1089/cmb.2021.0314	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 H Park, R Yamaguchi, S Imoto, S Miyano	4. 巻 17
2. 論文標題 Xprediction: Explainable EGFR-TKIs response prediction based on drug sensitivity specific gene networks	5. 発行年 2022年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0261630
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0261630	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 H Park, K Maruhashi, R Yamaguchi, S Miyano	4. 巻 276(9)
2. 論文標題 説明可能なAIによるパーソナルゲノム医療・予防の実現	5. 発行年 2021年
3. 雑誌名 医学のあゆみ	6. 最初と最後の頁 849-853
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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 K Maruhashi, H Park, R Yamaguchi, S Miyano	4. 巻 12007.03912
2. 論文標題 Linear Tensor Projection Revealing Nonlinearity.	5. 発行年 2020年
3. 雑誌名 arXiv	6. 最初と最後の頁 1-13
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
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5件(うち招待講演 1件/うち国際学会 0件)

1. 発表者名 Park Heewon, Konishi Sadanori
2. 発表標題 Sparse Tensor Subspace method for biological modulator selection
3. 学会等名 科研費シンポジウム「高次元複雑データの統計モデリング」
4. 発表年 2020年

1. 発表者名 Park Heewon
2. 発表標題 XAIに基づく大規模遺伝子ネットワークの総合的解析
3. 学会等名 規模データ解析と人工知能技術によるがんの起源と多様性の解明 シンポジウム
4. 発表年 2021年

1. 発表者名 Park Heewon
2. 発表標題 大規模遺伝子ネットワーク解析に基づく抗がん剤耐性マーカー探索
3. 学会等名 スーパーコンピュータ「富岳」 成果創出加速プログラムシンポジウム
4. 発表年 2021年

1. 発表者名 Park Heewon
2. 発表標題 データサイエンス・説明可能なAIによる個別化医療
3. 学会等名 第16回四大学連合文化講演会
4. 発表年 2021年

1. 発表者名 Park Heewon, Maruhashi Koji, Yamaguchi Rui, Imoto Seiya, Miyano Satoru
2. 発表標題 Comprehensive gene regulatory network analysis based on explainable AI
3. 学会等名 第80回日本癌学会学術総会（招待講演）
4. 発表年 2021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
---------	---------