

令和 3 年 5 月 27 日現在

機関番号：12601

研究種目：基盤研究(B) (一般)

研究期間：2018～2020

課題番号：18H03328

研究課題名(和文) ベイズモデルによるがんと免疫のシステムの統合理解と免疫逃避機構の解明

研究課題名(英文) Bayesian Modeling for Understanding Systems of Immune Responses and Revealing Immune Escape Mechanisms

研究代表者

井元 清哉 (Imoto, Seiya)

東京大学・医科学研究所・教授

研究者番号：10345027

交付決定額(研究期間全体)：(直接経費) 12,700,000円

研究成果の概要(和文)：がんの免疫監視を行う複雑なヒト免疫システムの理解のためのデータ解析技術開発に関して次の成果を得た。全ゲノムシーケンスデータからのHLA遺伝子型の決定は、精度が極めて低い状態であったが、新たなベイズ型統計モデルを構築することで98%の精度を達成した。更にデータベースに登録のない新規のHLA遺伝子型やHLA遺伝子の体細胞変異も正確に同定できる方法へ発展させた。更に、HLAに結合しT細胞に提示されるネオ抗原の同定のため、Neoantimonアプリケーションを開発して公開できた。これらの技術を元に、各患者の癌組織が免疫システムからどのように選択圧を受けてきたかを解析出来る新しい指標IEIを開発した。

研究成果の学術的意義や社会的意義

開発したデータ解析技術群は、大型計算機環境を必要とするが、研究代表者が所属するヒトゲノム解析センターのスパコンSHIROKANE上に実装されているため、多くの研究者が免疫ゲノム解析を行うことが可能となり、研究分野の活性化に繋がる。開発した技術群は、患者の癌の免疫学的と級長を炙り出すことが出来る。免疫細胞が浸潤した体細胞変異の多い癌は免疫チェックポイントインヒビターなどの免疫療法が有効であることが知られている。一方、そのような免疫療法が有効ではない癌の特徴を開発した方法を駆使しその共通性を見出すことで、難治癌に対する新しい免疫学的な治療法を開発するシーズになることが期待される。

研究成果の概要(英文)：The following results were obtained regarding the development of data analysis techniques for understanding the complex human immune system that monitors cancer immunity. Although the accuracy of HLA genotyping from whole genome sequencing data was extremely low, we achieved 98% accuracy by constructing a new Bayesian statistical model. Furthermore, we have developed a method to accurately identify new HLA genotypes and somatic mutations in HLA genes that are not registered in the database. In addition, we have developed and released an application named Neoantimon for the identification of somatic neoantigens that bind to patient HLA and are presented to T cells. Based on these technologies, we have developed a new index, IEI, which can be used to analyze how each patient's cancer tissue has been subjected to selective pressure from the immune system.

研究分野：バイオインフォマティクス

キーワード：ベイズモデル 免疫システム HLA ネオ抗原

## 1. 研究開始当初の背景

次世代シーケンサーの発展に伴い、細胞内の DNA や RNA の網羅的な計測が可能となった。ゲノムに蓄積したさまざまな後天的変異から生じるがんの分子病態の理解は飛躍的に進んでいる。特に、分子標的薬の効果予測や開発に関して、ゲノムの変異情報は重要な情報となりさまざまな成果が上げられていた。

一方、免疫システムによるがん抑制機構については、T 細胞応答を抑制もしくは停止させる免疫チェックポイント分子 (PD-1 や PD-L1、CTLA-4 など) を標的としたオプジーボを代表とする免疫チェックポイント阻害薬が一部のがんで有効性を示したことから近年大きな注目を集めていた。当時、がん組織における免疫学的な特性を理解するために、免疫関連分子であるヒト白血球型抗原 (human leukocyte antigen: HLA) や T 細胞受容体、B 細胞受容体などについて、シーケンス技術を駆使した研究が進められていた (Shukla SA, et al. Nature Biotechnol., 2015; Li B, et al. Nature Genet., 2016)。

これら免疫反応に関連する分子は、多様な「非自己 (細菌・ウイルス)」に対応するために、ヒトの長い歴史の中で突然変異が蓄積され高い多様性を有している。具体的には、HLA class I に属する HLA-A 遺伝子は、遺伝子上の多型のパターンにより 3,000 以上の型に分類される。HLA-B や HLA-C 遺伝子にもそれぞれ約 4,000、2800 の型が知られている。また、T 細胞受容体についても、 $\alpha$  鎖、 $\beta$  鎖のヘテロダイマーとして 10 の 15 乗以上の多様性を有している。このように、極めて多様性の高い複雑な免疫システムを理解し、例えば免疫療法の有効性の予測を行うためには、膨大なゲノム関連データ、および臨床情報が必要となり、免疫反応をシステムとして解析出来る新たなデータ解析技術の開発が必要不可欠であった。

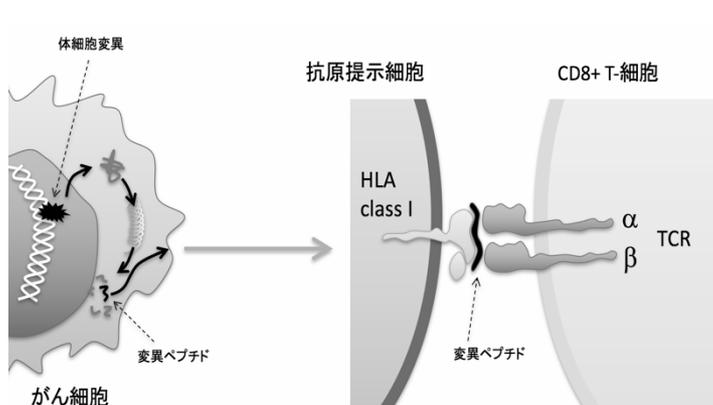


図1：体細胞変異からのがん抗原の生成から HLA クラス I 遺伝子による抗原提示、T 細胞による認識に至るプロセスの概念図

がん細胞のゲノム上、翻訳領域に蓄積した後天的変異の多くは、転写・翻訳されることによりアミノ酸置換を伴う変異ペプチドを生成する (図1)。この変異ペプチドが HLA 分子により提示され (ネオ抗原と呼ばれる)、細胞障害性 T 細胞に認識されると T 細胞はがん細胞を「非自己」と認識し攻撃する。これが免疫システムによるがん抑制機構である。上述した HLA や T 細胞受容体の研究は、このシステムのコンポーネントをなすものである。これらのコンポーネントを統合してシステムとして理解することが必

要となっていた。しかしながら、各コンポーネントを解析する技術の開発は徐々に進んできてはいるが、統合的なシステムとして解析する情報技術はまだ十分な研究はなかった。

また、がん組織は、上述した免疫システムからがん抑制機構によるプレッシャーを受けながらも、そのプレッシャーから逃れ構築されたと考えられる。すなわち、がん組織が形成される過程で、がんと免疫システムのせめぎ合いがあり、各がん組織が受けてきた免疫プレッシャーには、その強さ・期間に違いがある。例えば、ある患者のがん組織は長い期間免疫プレッシャーからの抑制を受け続けてきた。一方、別の患者のがん組織は、がん細胞を取り巻くさまざまな要因 (がんの微小環境) により、変異が蓄積していく比較的初期の段階で免疫プレッシャーからのエスケープを達成し構築されたという違いである。がん組織が受けてきた免疫プレッシャーの“ヒストリー”が分かれば、免疫プレッシャーからエスケープするためのがんの“免疫逃避機構”が明らかとなり、この免疫逃避機構を標的とした新たな治療法へと繋がる可能性がある。ヒストリーの違いの一端は、蓄積している変異数にも表れていると考えられるが、組織による違い、個人差もあり明確に各がん組織が受けてきた免疫プレッシャーの強さを測る統計量は存在しなかった。

## 2. 研究の目的

個々には進みつつある免疫反応のコンポーネント「HLA 遺伝子の多様性・変異」、「ネオ抗原の予測」、「がん組織に浸潤している免疫細胞の解析」、「T 細胞・B 細胞受容体レパトアの解析」について、それらを統融合した情報解析技術を開発し、免疫反応のシステム的理解を目指す。また、がん組織が受けてきた免疫プレッシャーの“ヒストリー”、およびそこから明らかとなる免疫逃避機構について解明できる情報解析技術の開発を行う。

### 3. 研究の方法

研究チームは、これまでに国際がんゲノムコンソーシアム ICGC/TCGA PCAWG プロジェクトにて約 2,800 名分の全ゲノムシーケンスデータ、RNA シーケンス、臨床情報合わせてを免疫学的観点から解析した経験を有し、この大規模ゲノムデータを活用した研究開発を推進する。大きく分けて以下の2つの研究を推進する。

1. HLA 遺伝子 (A、B、C、DP、DQ、DR) の型、および非翻訳領域までを含むレアバリエント・体細胞変異情報、SNV、Indel、Fusion などさまざまな体細胞変異から同定されるネオ抗原、浸潤免疫細胞、T 細胞受容体・B 細胞受容体レパトアの情報を統合し、予後や免疫療法の有効性など臨床的な指標を予測するベイズモデルの開発。
2. がん組織ががん化の過程で免疫システムから受けてきた免疫プレッシャーの“ヒストリー”を解析し、がんの免疫逃避機構を解明する情報科学技術。

個々の解析モデルを階層ベイズ的に統合し、臨床的な指標に相関する免疫システムの特徴的挙動を見いだす。また、ヒストリーの違いについての情報として、がん種横断的な研究が必要となる。基本的なアイデアとして、ネオ抗原の数は一つの指標になると考えているが、単に数のみではどれほど強いプレッシャーを受けてきたか判断ができない。そこで、免疫プレッシャーを受けていないと推察される指標を構築し、コントロールとして比較する必要がある。

### 4. 研究成果

以下の成果を得た。

(1) HLA 遺伝子の解析技術について開発研究を進め、全ゲノムシーケンスデータから高精度に HLA 遺伝子型を決定することのできるベイズ統計学に基づく数理モデル ALPHLARD を構築した (図2)。この情報解析技術を用いることにより、これまでは50%程度 (Bauer et al., Briefings in Bioinformatics, 2016) であった全ゲノムシーケンスデータからの HLA 遺伝子型の決定について、98%以上の精度を達成するという革新的な結果を得た。また、開発した技術は、HLA 遺伝子型について、これまで知られていない型も予測できるよう、HLA 遺伝子の配列を予測する機能も備えており、新たな型の発見や未だ多くの型で決定されていないプロモーターやイントロンなど非翻訳配列の DNA 配列の決定も行える特徴を有している。

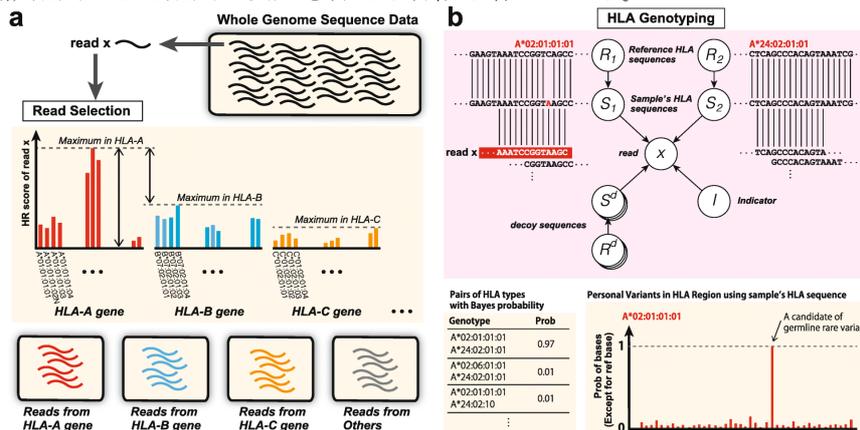


図2：ALPHLARDによるHLA遺伝子シーケンス解析の流れ

(2) がん細胞から取得したシーケンスデータと生殖系細胞ゲノムのシーケンスデータを合わせて HLA 遺伝子に生じた後天的ゲノム変異を高精度に同定する手法の開発を行った。生殖系細胞のゲノムシーケンスはがん細胞からのゲノムシーケンスよりも深度は浅めで読まれることが多い。すなわちがん細胞からのシーケンスデータの方が HLA 遺伝子型の決定には有利である。しかしながら、がん細胞における HLA 遺伝子には変異が入っている可能性がある。そこで両方のデータを同時に使用して2つの解析を同時に実行し、2つの解析の精度を相乗効果によって高める新たなデータ解析技術である ALPHLARD-NT を開発した。その結果、上記 class I および class II の遺伝子において、1st field で100%, 2nd field で99.2%, 3rd field で98.5%と世界最高精度を示すことが出来た。また、ALPHLARD を用いて国際がんゲノムコンソーシアムの PanCancer Analysis of Whole Genomes (PCAWG) プロジェクトにおいて約2,800のがんサンプルの HLA 遺伝子型を解析した。この結果は、PCAWG プロジェクトの成果の一部として Nature, Feb6;578(7793):82-93(2020)において発表された。

(3) がんゲノムの体細胞変異によって生じる変異型ペプチドが HLA 分子によって免疫細胞に抗原提示されるか否かを予測するソフトウェア Neoantimon を開発し、論文 (Bioinformatics, 2020) にて発表すると共に、ソフトウェアは GitHub において公開した。

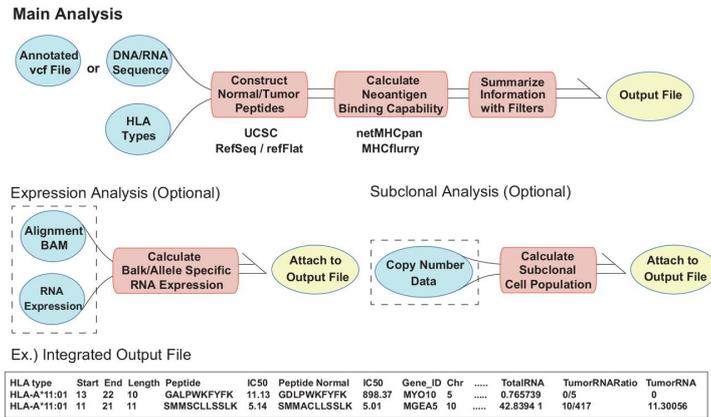


図3：Neoantimonの解析の流れ

(4) RNA シークエンスデータに基づき、浸潤免疫細胞を詳細に解析することで同じがん種で有りながらがん細胞の周辺に存在する免疫細胞の違いを微小環境と捉え、機能的パスウェイの違いとして抽出する解析技術を構築し、ICGC/TCGA PanCancer Analysis of Whole Genomes のデータに適用した。その結果、免疫細胞の浸潤の多い症例については、癌腫横断的に共通の免疫関連のパスウェイが活性化している一方、浸潤の少ない症例はがん種毎に活性化しているパスウェイに顕著な差が見られた。これは、所謂 cold tumor に相当し、免疫細胞の浸潤が少ないため免疫療法の有効性があまり高くなく、新たな治療法が求められるグループであると考えられる。

(5) がん組織が受けてきた免疫プレッシャーの“ヒストリー”を解析するための Immune Editing Index (IEI) を開発した。これは、エキソン領域中の Nonsynonymous 変異が免疫監視によって選択圧を受ける程度の定量を狙ったものである。アイデアは、Pseudogene 領域に生じた変異はタンパク質に翻訳されないため選択圧を受けないと考えられるため、Pseudogene 領域上の変異とエキソン領域上の変異の密度を比較するものである。この IEI は、Pseudogene 領域の情報を用いるため、全ゲノムシークエンスデータが必要となる。上述した PanCancer の全ゲノムシークエンスデータを活用することで、がん種による免疫プレッシャーの違い、同じがん種であっても個人ごと異なる免疫プレッシャーが予後などにどのように関わるかを解析した。この成果は現在国際科学誌に投稿中である。

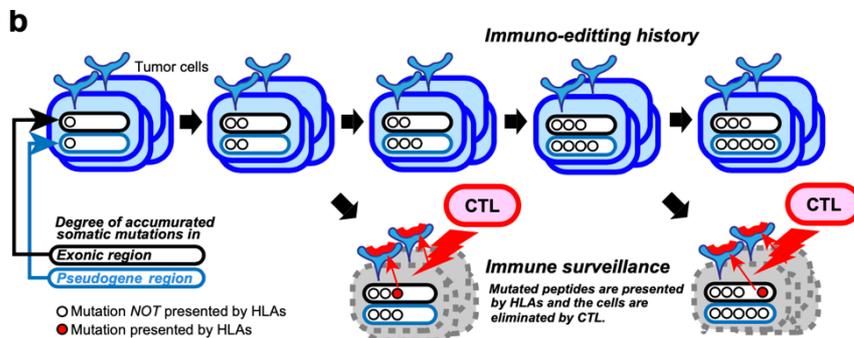


図3：Immune Editing Index (IEI) の概念図

5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 1件/うちオープンアクセス 6件）

1. 著者名 Fujimoto A, Fujita M, Hasegawa T, Wong J-H, Maejima K, Oku-Sasaki A, Nakano K, Shiraishi Y, Miyano S, Yamamoto G, Akagi K, Imoto S, Nakagawa H	4. 巻 30
2. 論文標題 Comprehensive analysis of indels in whole-genome microsatellite Regions and microsatellite instability across 21 cancer types	5. 発行年 2020年
3. 雑誌名 Genome Research	6. 最初と最後の頁 334-346
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/gr.255026.119.	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 The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium	4. 巻 578(7793)
2. 論文標題 Pan-cancer analysis of whole genomes	5. 発行年 2020年
3. 雑誌名 Nature	6. 最初と最後の頁 82-93
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41586-020-1969-6	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 該当する

1. 著者名 Fujita M, Yamaguchi R, Hasegawa T, Shimada S, Arihiro K, Hayashi S, Maejima K, Nakano K, Fujimoto A, Ono A, Aikata H, Ueno M, Hayami S, Tanaka H, Miyano S, Yamaue H, Chayama K, Kakimi K, Tanaka S, Imoto S, Nakagawa H	4. 巻 53
2. 論文標題 Classification of primary liver cancer with immunosuppression mechanisms and correlation with genomic alterations	5. 発行年 2020年
3. 雑誌名 EBioMedicine	6. 最初と最後の頁 102659
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.ebiom.2020.102659	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 Yang F, Kim D-K, Nakagawa H, Hayashi S, Imoto S, Stein L, Roth FP.	4. 巻 15(7)
2. 論文標題 Quantifying immune-based counterselection of somatic mutations	5. 発行年 2019年
3. 雑誌名 PLoS Genet	6. 最初と最後の頁 e1008227
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pgen.1008227	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 Konishi H, Komura D, Katoh H, Atsumi S, Koda H, Yamamoto A, Seto Y, Fukayama M, Yamaguchi R, Imoto S, Ishikawa S.	4. 巻 20(1)
2. 論文標題 Capturing the differences between humoral immunity in the normal and tumor environments from repertoire-seq of B-cell receptors using supervised machine learning	5. 発行年 2019年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 267
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-019-2853-y	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 Hayashi S, Moriyama T, Yamaguchi R, Mizuno S, Komura M, Miyano S, Nakagawa H, Imoto S	4. 巻 26(9)
2. 論文標題 ALPHLARD-NT: Bayesian method for HLA genotyping and mutation calling through simultaneous analysis of normal and tumor whole-genome sequence data	5. 発行年 2019年
3. 雑誌名 Journal of Computational Biology	6. 最初と最後の頁 923-937
掲載論文のDOI (デジタルオブジェクト識別子) 10.1089/cmb.2018.0224	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 Hayashi Shuto, Yamaguchi Rui, Mizuno Shinichi, Komura Mitsuhiro, Miyano Satoru, Nakagawa Hidewaki, Imoto Seiya	4. 巻 19
2. 論文標題 ALPHLARD: a Bayesian method for analyzing HLA genes from whole genome sequence data	5. 発行年 2018年
3. 雑誌名 BMC Genomics	6. 最初と最後の頁 790
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12864-018-5169-9	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 Muraoka Daisuke, Seo Naohiro, Hayashi Tae, Tahara Yoshiro, Fujii Keisuke, Tawara Isao, Miyahara Yoshihiro, Okamori Kana, Yagita Hideo, Imoto Seiya, Yamaguchi Rui, Komura Mitsuhiro, Miyano Satoru, Goto Masahiro, Sawada Shin-ichi, Asai Akira, Ikeda Hiroaki, Akiyoshi Kazunari, Harada Naozumi, Shiku Hiroshi	4. 巻 129
2. 論文標題 Antigen delivery targeted to tumor-associated macrophages overcomes tumor immune resistance	5. 発行年 2019年
3. 雑誌名 Journal of Clinical Investigation	6. 最初と最後の頁 1278 ~ 1294
掲載論文のDOI (デジタルオブジェクト識別子) 10.1172/JCI97642	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 -

〔学会発表〕 計2件（うち招待講演 2件 / うち国際学会 1件）

1. 発表者名 Seiya Imoto
2. 発表標題 Immuno-genomics pancancer landscape reveals diverse immune escape mechanisms and immune-editing histories
3. 学会等名 Biomedical Sciences in the Era of Big Data (招待講演) (国際学会)
4. 発表年 2018年

1. 発表者名 Seiya Imoto
2. 発表標題 Cancer Immunogenomics Analysis of ICGC/PCAWG
3. 学会等名 The 77th Annual Meeting of the Japanese Cancer Association (招待講演)
4. 発表年 2018年

〔図書〕 計1件

1. 著者名 井元清哉, 長谷川嵩矩, 山口類	4. 発行年 2018年
2. 出版社 メディカルドゥ	5. 総ページ数 5
3. 書名 T細胞受容体レパトア解析	

〔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
研究分担者	中川 英刀  (Nakagawa Hidewaki)  (50361621)	国立研究開発法人理化学研究所・生命医科学研究センター・チームリーダー   (82401)	

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
---------	---------