

平成 30 年 5 月 31 日現在

機関番号：12601

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00399

研究課題名(和文) T細胞受容体シーケンスデータ解析手法の開発と細胞集団多様性の理解

研究課題名(英文) Developments of methods for T-cell receptor sequence data analysis and understanding of diversities in the immune cells

研究代表者

山口 類 (Yamaguchi, Rui)

東京大学・医科学研究所・准教授

研究者番号：90380675

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：免疫細胞の多様性を特徴づける受容体(T細胞受容体(TCR)およびB細胞受容体(BCR))配列を、大量の次世代シーケンスデータから精密かつ高効率に解析するための解析基盤の開発を行った。開発した手法を、がん、クローン病、造血幹細胞移植後のGVHD等の種々の疾病の実データに適用し、これまでに得られていなかった、疾病と免疫環境の関係に関わる情報を抽出することに成功し、新規遺伝子候補配列も発見した。また免疫細胞集団から得られた時系列データからの、動的モデルに基づく集団構造の推定法および予測手法についても開発を進めた。

研究成果の概要(英文)：We have developed methodologies and tools for analyzing Next Generation Sequencing data sets from T-cell or B-cell receptors that characterize diversities of those immune cells. The tools are accurate and efficient for analyzing vast amount of data sets. By applying the tools to data sets from various diseases, e.g., cancers, Crohn's disease, and GVHD after hematopoietic stem cell transplantations, we succeeded to extract novel information representing relationships between those diseases and immune environments. We also could find new exon candidates in those receptor sequences. Furthermore, we proceeded with developments of dynamic models to estimate and predict variations of structure of immune-cell groups from time-course data sets.

研究分野：メディカルバイオインフォマティクス

キーワード：免疫細胞シーケンスデータ解析 免疫細胞受容体レパトア解析 T細胞受容体 B細胞受容体

1. 研究開始当初の背景

近年、シーケンス計測技術の急激な発展により、DNA 配列をはじめ遺伝子発現情報、エピゲノム情報を含む様々なシーケンスデータが高速、安価かつ大規模に産生可能になってきた。またこのことにより個々人のシーケンス情報に基づき、疾病に対する治療法や予防法をカスタマイズする個別化ゲノム医療を目指した研究が世界中で進められている。このような背景の下、個別化医療の実現に向けては、各種計測データから個々人の生体の状態を特徴づける情報抽出のためのデータ解析手法の開発が喫緊の課題である。

世界的には、これまで特にがん研究の分野で、がん細胞のシーケンスデータから各種がんを特徴づける変異を検出する手法の開発が進み、国際がんゲノムコンソーシアム等により、様々な臓器におけるがんの変異の網羅的なカタログが構築されつつあり、がん種特異的遺伝子発現パターン等とともに個別化医療に利用する試みが進んでいる。申請者はこれまで、多くのがん研究者との共同研究を通じて遺伝子発現データからの動的細胞システム推定法や機能的遺伝子モジュール探索法等を開発し、DNA 変異検出を含む実データ解析に携わってきた。特に論文 Yamauchi *et al.*, PLoS ONE, 2011 では、早期肺腺がん患者の予後予測を高精度に行う遺伝子シグネチャーの同定を、申請者等が開発してきた状態空間モデルに基づくベイズ的時系列解析法を応用することで達成し、個別化医療に資する結果を得た。

そのような経験の中で、生体の状態は真に多様であるという認識を得るとともに、がんを含む様々な疾病において免疫系が重要な役割を果たしていることを知り、免疫細胞群の多様性および特異性を特徴づける新たな情報抽出技術の開発が必要であると考えに至った。本研究では特に適応免疫系で主要な働きをする T 細胞について着目する。一つの T 細胞は細胞膜上に特定の抗原を認識する T 細胞受容体 (TCR: T Cell Receptor) を一種類だけ発現している。一方、生体は、未知の抗原に対する認識可能性を高めるため、あらかじめ T 細胞受容体の形状が異なる多種多様な T 細胞クローン群 (TCR レパトア) を用意している (図 1)。T 細胞による免疫応答は、例えばがん組織においては、がん細胞表面にのみ存在するペプチド抗原を特異的に認識する T 細胞が、がん細胞を攻撃することが知られている。また各種感染症や自己免疫疾患および薬疹等の薬剤応答にも深く関わっていることが知られている。故に各種疾病に関わる TCR レパトアに含まれる T 細胞クローンの種類や量を網羅的に特徴づけ、その変化や差異を知ることができれば、疾病の進展や生体システムの応答メカニズムに関して詳細かつ包括的な知見を得ることができると期待された。しかしながら TCR レパト

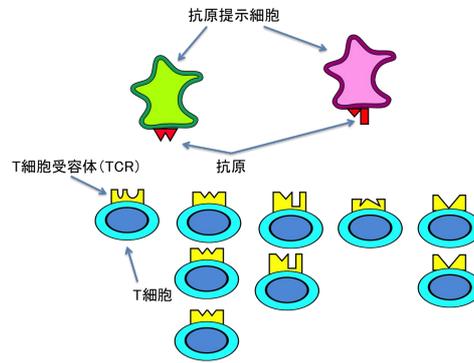


図 1: TCR レパトア。各 T 細胞クローンは特定の抗原を認識する。抗原と TCR は、鍵と鍵穴の関係にある。

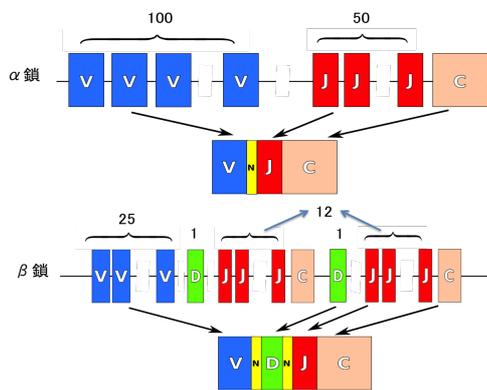


図 2:ゲノム再編成に基づく TCR 多様性獲得過程

アの網羅的解析は、次世代シーケンサー出現以前は不可能であった。何故なら T 細胞受容体は鎖鎖の二量体からなり、各鎖をコードする DNA 配列は、遺伝子断片 (V, (D), J exon) が、DNA の再編成を通じて生成され、特にランダム配列 (N セグメント) も付与されることにより、理論上はその組み合わせは  $10^{15}$  を超える (図 2)。

故に PCR 等による計測では、特定少数の V(D)J の組み合わせの発現量の解析に留まっていた。しかしシーケンス技術の発展により、V(D)J を含む長いリード長 (>150bp) が得られるようになり、T 細胞群の DNA シーケンスまたは RNA シーケンスデータから、TCR レパトアを決定できる道が開かれた。既にいくつか解析手法が提案されている (Zvyagin *et al.*, Bioinformatics, 2013 等)。申請者等も Fang *et al.*, OncoImmunology, 2014 において、ソフトクリップアライメント法に基づく TCR レパトア推定法を提案していた (図 3)。しかしながら T 細胞受容体シーケンスデータ解析手法の開発は端緒にすぎたばかりであり、多様性を正確に理解する上では未だ不十分であった。またこれまで提案されている手法は、単時点の TCR レパトアの同定を目指すものが主であるが、免

疫反応は本質的に動的であり、その本態を理解するためには、TCR レパトア時系列データ解析手法の開発が必要であり、これまで開発してきた高次元短時系列に対するベイズ型時系列解析の手法の経験が有用であるという着想に至っていた。

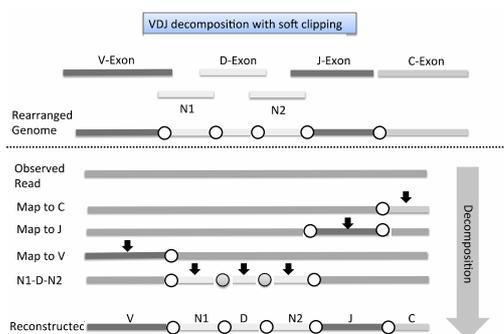


図 3:ソフトクリップ法に基づく V(D)J セグメント同定法

## 2. 研究の目的

適応免疫で主要な役割を果たす T 細胞の各クローンの抗原特異性は、細胞膜上の T 細胞受容体 (TCR) の配列により決定される。本研究では未知数のクローンを含む T 細胞由来の RNA シークエンスデータから、各クローン特有の T 細胞受容体の配列を正確に決定し、クローン種数と各クローンの量 (TCR レパトア) を正確に推定するデータ解析手法を開発し個別化医療に資する情報を抽出することを目指した。また TCR レパトア多様性解析においてこれまで未開拓である、リファレンス配列に含まれない新規 Exon 配列の同定法、典型的ゲノム再編成モデルに従わない配列の同定法および TCR レパトア時系列データからの情報抽出法の開発を行い、複雑な免疫反応の本態に対する新たな認識基盤を与えることを目指した。

## 3. 研究の方法

TCR レパトア多様性解析においてこれまで未開拓である、リファレンス配列に含まれない新規 Exon 配列の同定法、典型的ゲノム再編成モデルに従わない配列の同定法および TCR レパトア時系列データからの情報抽出法を開発を以下のように進めた。まず申請者等が開発を進めてきた TCR レパトア多様性解析法を改良し、既存リファレンスに含まれる Exon 配列に基づく TCR 鎖および鎖の配列決定を、迅速かつ正確に行うことのできる解析基盤を整備する。次に、新規 Exon 同定アルゴリズムおよび非典型ゲノム再編成検出解析アルゴリズムを検討開発し、解析基盤上に実装する。最後に複数時点で作られた T 細胞時系列データを、高次元短時系列とみなし、ベイズ型モデルでモデル化し、時系列変化の背後にあるシステムの情報を抽出する手法および鎖の組を推定する手法の開発を

目指した。

## 4. 研究成果

TCR レパトア多様性解析法の改良:

次世代シーケンサーから得られたシーケンスリードに含まれる TCR 鎖および鎖由来配列を、既存データベースに登録された T 細胞受容体 (TCR) 遺伝子リファレンス配列群に基づき精密決定するためのデータ解析基盤の開発整備を進めた。具体的には申請者がこれまで開発してきた TCR レパトア多様性解析法の改良を進めるとともに高速化を行った。主な解析プロセスは (1) シークエンスリードのリファレンス配列へのアライメント、(2) アライメント結果に基づく、リード中に含まれる V-(D)-J-C 配列に対するセグメント分割、既知配列 ID 同定およびランダムに付加される配列群 (N セグメント) の決定からなる。それぞれのプロセスにおいてパラメータおよびアルゴリズムの最適化を図った。また特に上記 (2) のプロセスを精査することによりボトルネックとなっている処理部分を同定し、当該部分へ並列化を施すことにより約 10 倍の高速化を達成した。また更に解析プログラムへ拡充・改良を施すことにより、マウス由来の TCR 配列データも解析できるようにした。

また既存のリファレンス配列に基づく TCR レパトア精密決定解析基盤を元に、既存リファレンス配列に含まれない新規エクソン配列の探索法を開発を進めた。まず既存の配列に対して、小規模な変異 (一塩基置換および短い欠失挿入変異等) を持つ新規遺伝子配列候補群をターゲットとした探索法を開発を進めた。ここでは上記の解析基盤から得られた、各配列リード中に含まれる V、(D)、J、C 遺伝子セグメント推定情報に加え、リード配列の各リファレンス配列に対するミスマッチ情報を精査することで変異を含む配列群の探索を行う方針で開発を進めた。またシーケンシングエラーによる擬陽性検出の可能性を低減させるために、同一サンプル由来のシーケンスリード群において、候補変異が、異なる V(D)J 遺伝子セグメントの組み合わせでの検出頻度や、また他検体での検出頻度を考慮にいたれたフィルタリング法を検討することで、より精度の高い新規遺伝子セグメント候補群の検出を行うことができるようになった。

更に上記の小規模変異を含む配列よりも、より類似度の低い新規遺伝子配列候補群の探索法を開発も進めた。これに関しては、上記の解析基盤における解析において、V(D)JC 遺伝子セグメントを、部分的に検出出来なかったリード群の中から、新規遺伝子配列候補群を探索する方針で開発を進めた。上記の部分的非決定配列群を高精度に解析する方法を検討し、更に小規模変異を含む配列と同様の方針でフィルタリングを行うことで、有望

な候補群検出を行うことができるようになった。また得られた候補配列の中に含まれる配列群を詳細に検討することで、新規遺伝子配列の生成機序に対する考察も行った。

また TCR レパトア精密決定解析基盤を改良を進めると共に、その基盤を元に B 細胞受容体 (BCR) レパトアの精密決定解析基盤を開発した。

上述の手法を、種々の疾患例、また治療介入前後の実データに適用し、各種疾患における免疫環境に関して様々な知見を得た。

論文<sup>1</sup>では、造血幹細胞移植後の重篤な副作用である GVHD(Graft Versus Host Diseases)を起こした患者と、起こしていない患者の間で、TCR レパトアの多様性を比較した。結果、特に GVHD の患者群では多様性が失われていることを見だし、GVHD 症例においては少数種のクローンが増幅し宿主側の細胞を攻撃している状況にあることが推察された。

論文<sup>2</sup>では、クローン病の患者の患部由来のサンプルにおいて TCR 解析を行い、TCR 配列の精密決定を行うことにより、複数の患者間で共通の CDR3 配列を持つ TCR クローンが増幅していることを見だし、共通の抗原認識が病因に関わる可能性のあることを推察することができた。

論文<sup>3</sup>においては、大腸がん患者にたいする、がんワクチン治療前後の、患部および末梢血の TCR レパトアの時間変化を観測した。結果、末梢血においては治療前後で、TCR レパトアの構造がほとんど変化しないのに対して、患部においては大きく TCR クローンの分布が変化していることをモニターすることが出来た。

論文<sup>4</sup>においては、子宮体がんにおける腫瘍浸潤 T 細胞の T 細胞クローン集団の多様性と免疫関連遺伝子群の発現量との関連を調べ、予後因子としての可能性を見だし報告した。

論文<sup>5</sup>では、BCR 解析においては、食物アレルギー患者に対する経口免疫療法前後の B 細胞クローン集団の多様性を調べ、免疫療法後にある種 IGHα および IGHβ のクローンの有意な濃縮が起きていることを観測した。またそれに加えて開発を行ってきた新規遺伝子配列探索法により新規 IGH-V 遺伝子配列候補も見だし報告した。

更に TCR レパトア時系列データから、クローン集団構造の変化を推定、予測する字形工列モデルの構築も進めた。

更に、通常の RNA-seq 計測により得られた遺伝子発現データから TCR の多様性を推定する方法の開発を進めた。前述の TCR レパトア精密決定解析基盤は、TCR 遺伝子転写産物のみを対象としたターゲットシーケンシングにより得られた比較的長いリード配列中の V(D)JC 遺伝子セグメントを高精度に決定するものである。一方、通常の RNA-seq 計測では得られるリード長は短く、またターゲッ

ト抽出を行わないため TCR 転写産物由来のリード数は限られる。しかしながら TCGA を初めとする巨大レポジトリに蓄積されている多数の RNA-seq データから、ターゲットシーケンスに比べて精度は劣るものの TCR 多様性に関わる情報を抽出することは意味がある。また同時に計測された他の遺伝子群の発現情報を加味することで、腫瘍浸潤細胞群と周辺細胞群の相互作用に関する情報を抽出出来る可能性がある。本年度は TCGA や ICGC の RNA-seq データを対象に、上記の手法の開発を進め予後情報等との関連の探索を行った。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計6件)

<sup>1</sup> Yew PY, Alachkar H, Yamaguchi R, Kiyotani K, Fang H, Yap KL, Liu HT, Wickrema A, Artz A, van Besien K, Imoto S, Miyano S, Bishop MR, Stock W, Nakamura Y, Quantitative characterization of T-cell repertoire in allogeneic hematopoietic stem cell transplant recipients. Bone Marrow Transplantation, 査読有, 50, 2015, 1227-1234.

DOI: 10.1038/bmt.2015.133.

<sup>2</sup> Chapman CG, Yamaguchi R, Tamura K, Weidner J, Imoto S, Kwon J, Marino SR, Miyano S, Nakamura Y, Kiyotani K, Characterization of T-cell Receptor Repertoire in Inflamed Tissues of Patients with Crohn's Disease Through Deep Sequencing. Inflammatory Bowel Diseases, 査読有, 22(6), 2016, 1275-1285.

DOI: 10.1097/MIB.0000000000000752.

<sup>3</sup> Tamura K, Hazama S, Yamaguchi R, Imoto S, Takenouchi H, Inoue Y, Kanekiyo S, Shindo Y, Miyano S, Nakamura Y, Kiyotani K. Characterization of T cell repertoire in tumor tissues and blood in advanced colorectal cancers through deep T cell receptor sequencing. Oncology Letters, 査読有, 11(6), 2016, 3643-3649.

DOI: 10.3892/ol.2016.4465.

<sup>4</sup> Ikeda Y, Kiyotani K, Yew PY, Sato S, Imai Y, Yamaguchi R, Miyano S, Fujiwara K, Hasegawa K, Nakamura Y. Clinical significance of T cell clonality and expression levels of immune-related genes in endometrial cancer. Oncol Rep., 査読有, 37(5), 2017, 2603-2610.

DOI: 10.3892/or.2017.5536.

<sup>5</sup> Kiyotani K, Mai TH, Yamaguchi R, Yew PY, Kulis M, Orgel K, Imoto S, Miyano S, Burks AW, Nakamura Y. Characterization of the B-cell receptor repertoires in peanut allergic subjects undergoing oral immunotherapy. J Hum Genet., 査読有, 63(2), 2018, 239-248.  
DOI: 10.1038/s10038-017-0364-0.

<sup>6</sup> 井元清哉, 山口類, 水野晋一, 中川英刀, HLA タイピング・TCR レパトア解析による免疫ゲノムプロファイリング. 実験医学, 査読無, 35(4), 2017, 536-540  
<https://www.yodosha.co.jp/jikkenigaku/book/9784758101615/index.html>

〔学会発表〕(計 5 件)

<sup>1</sup> Tamura K, Hazama S, Yamaguchi R, Imoto S, Takenouchi H, Inoue Y, Kanekiyo S, Shindo Y, Miyano S, Nakamura Y, Kiyotani K. Characterization of T cell repertoire in advanced colorectal cancers through deep T cell receptor sequencing, 日本癌学会学術総会, 2015 年 10 月 9 日, 名古屋

<sup>2</sup> Yew PH, Alachkar H, Yamaguchi R, Kiyotani K, Liu HT, Wickrema A, Artz A, Imoto S, Miyano S, Bishop MR, Stock W, Nakamura Y, GVHD and GVL effects in allogeneic hematopoietic stem cell transplant through quantitative T cell repertoire analysis, 日本癌学会学術総会, 2015 年 10 月 10 日, 名古屋

<sup>3</sup> 山口類, 免疫レパトアデータ解析技術とその応用, 日本オミックス医療学会シンポジウム, 2016 年 10 月 18 日, 東京

<sup>4</sup> Ikeda Y, Kiyotani K, Yew PY, Sato S, Imai Y, Yamaguchi R, Miyano S, Fujiwara K, Hasegawa K, Nakamura Y, Clinical significance of immunological signatures and TCR repertoire in endometrial cancer, 日本癌学会学術総会, 2016 年 10 月 6 日, 横浜

<sup>5</sup> 清谷一馬, 田村賢司, 山口類, 井元清哉, 竹之内寛子, 宮野悟, 碓彰一, 中村祐輔, 大腸癌患者におけるペプチドワクチン特異的 T 細胞の TCR シークエンス解析, 日本癌学会学術総会, 2016 年 10 月 6 日, 横浜

〔図書〕(計 1 件)

Yamaguchi R, Imoto S, Miyano S. Springer Japan, 2015, Chapter2: A TCR sequence data analysis pipeline: Tcrip. In Immunopharmacogenomics. Nakamura, Y.

(Ed.),

〔産業財産権〕

出願状況 (計 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

取得状況 (計 件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
取得年月日:  
国内外の別:

〔その他〕  
ホームページ等

6. 研究組織

(1) 研究代表者

山口類 (YAMAGUCHI RUI)  
東京大学・医科学研究所・准教授  
研究者番号: 90380675

(2) 研究分担者

( )

研究者番号:

(3) 連携研究者

( )

研究者番号:

(4) 研究協力者

( )