

令和 5 年 6 月 12 日現在

機関番号：83901

研究種目：挑戦的研究（萌芽）

研究期間：2021～2022

課題番号：21K19939

研究課題名（和文）免疫単細胞データ解析による改変型T細胞受容体デザイン技術の開発

研究課題名（英文）Development of protein design methods for T cell receptors with immune single cell sequencing data analysis

研究代表者

山口 類（Yamaguchi, Rui）

愛知県がんセンター（研究所）・システム解析学分野・分野長

研究者番号：90380675

交付決定額（研究期間全体）：（直接経費） 4,900,000円

研究成果の概要（和文）：本研究では、T細胞受容体（TCR）遺伝子改変T細胞輸注療法（TCR-T療法）の個別化の実現に向けて、膨大な組み合わせ多様性を持つ改変型TCRと抗原ペプチド・HLA複合体（pHLA）間の結合において、結合強度を増強・制御した、改変型TCRタンパク配列を高精度かつ高効率にデザインする手法の開発を進めた。ここでは上記の高親和性TCRタンパク質デザインの問題を、統計科学におけるサンプリングの問題ととらえ、ベイズ統計の枠組みで親和性予測モデルからの予測結果を基に、親和性で条件づけたTCR配列に対する条件付き確率分布を求め、高親和性TCR配列の候補群を効率よく生成する手法を開発した。

研究成果の学術的意義や社会的意義

免疫細胞であるT細胞の表面にあるT細胞受容体（TCR）タンパク質を改変し、がん細胞の表面に提示されている特異的抗原を標的として、がん細胞を攻撃させる遺伝子改変T細胞輸注療法の開発が進んでいる。そこでは抗原との高い親和性を持つTCRを得ることがカギとなるが、一般に実験的にそのようなTCRを得ることは時間的にもコスト的にも困難な問題である。本研究では、標的抗原に対して高い親和性を持つTCRタンパクの候補を、計算によって探し出す方法を開発した。今後、実験での検証を進めることで、治療法の開発に役立てられることが期待される。

研究成果の概要（英文）：We studied a protein design method for T cell receptors (TCRs) for realizing personalized TCR T cell therapy, in which engineered T cells whose TCRs are replaced with one with an enhanced affinity to its target antigen presented on the surface of cancer cells, are infused to attack the cancer cells. There are several difficulties to identify such TCRs with enhanced affinities for target antigens. To overcome the difficulties, we consider the problem as a statistical sampling problem to generate candidate TCR sequences with high binding affinities. With a Bayesian modeling framework, we developed the sampling method in which a template-based binding affinity prediction model was utilized as a forward model. We also developed a machine-learning model for binding affinity prediction with a multi-modal learning architecture.

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：タンパク質デザイン T細胞受容体 ペプチドHLA複合体 ベイズモデル 機械学習 単細胞計測

1. 研究開始当初の背景

人間の適応免疫における、自己・非自己認識機構は T 細胞の表面にある TCR と、抗原提示細胞が細胞表面に提示する pHLA の結合反応により司られている。細胞傷害性 T 細胞 (CD8 陽性 T 細胞) は、がん細胞が HLA タンパク上に提示するがん特異的抗原ペプチドを非自己として認識し、がん細胞を攻撃する。近年、開発が進む有力ながん免疫療法として、患者由来の TCR に改変を加えて、がん抗原の認識能を高めがん細胞を攻撃させる、TCR 遺伝子改変 T 細胞輸注療法 (TCR-T 療法) およびキメラ免疫受容体 T 細胞輸注療法 (CAR-T 療法) がある。これらの治療法はすでに一部のがんに対して大きな奏功を示しているが、臨床的に意味のある効果を得るためには、TCR 遺伝子の改変により pHLA との結合能 (親和性) を高めることがカギとなる。

しかしながら現在の T 細胞受容体改変療法には、同療法の「個別化」を目指す上で大きな困難がある。それは個々の結合能を高めた改変型 T 細胞受容体の作出に膨大な労力と時間を要することである。なぜならば TCR タンパク中で抗原認識に重要な CDR3 領域のアミノ酸配列のみの改変を考えたとしても 10^{26} 通りを超える膨大な組み合わせ多様性を持ち、なおかつ、超広大な配列探索空間のどこからどの方向へ探索していけば、目的とする pHLA 複合体に対して、より高い結合能をもつ配列を得られそうか経験と勘に頼らざるを得ない状況であるからである。故に、望みの性質をもつ TCR 配列を見出すことには多大な困難を伴う。

その一方で、一人の患者のがん組織から得られるがん特異的抗原 (体細胞変異由来のネオ抗原を含む) は、多種多様かつ個別性が高い。これらの多様ながん抗原を TCR-T 療法の標的とすることができれば、治療方策の幅を大きく広げ将来の患者の福音となる可能性が高い。そのためには上記の困難を克服する新たな方法論の開発が必要不可欠である。

2. 研究の目的

本研究の目的は、近年がん免疫治療において開発が進む T 細胞受容体 (TCR) 遺伝子改変 T 細胞輸注療法 (TCR-T 療法) の個別化の実現に向けて、最先端の「データ科学技術」と「免疫一細胞シークエンス計測技術」の統合により、膨大な組み合わせ多様性を持つ改変型 TCR と抗原ペプチド・HLA 複合体 (pHLA) 間の結合において、結合強度を増強・制御した「改変型 TCR タンパク配列を高精度かつ高効率にデザインする手法」の開発を目指すことである。

3. 研究の方法

本研究では、上述の困難を克服するために、下記の三つの要素技術を統合し、改変型 TCR タンパク配列を高精度かつ高効率にデザインする新規手法の開発を進めることとした。

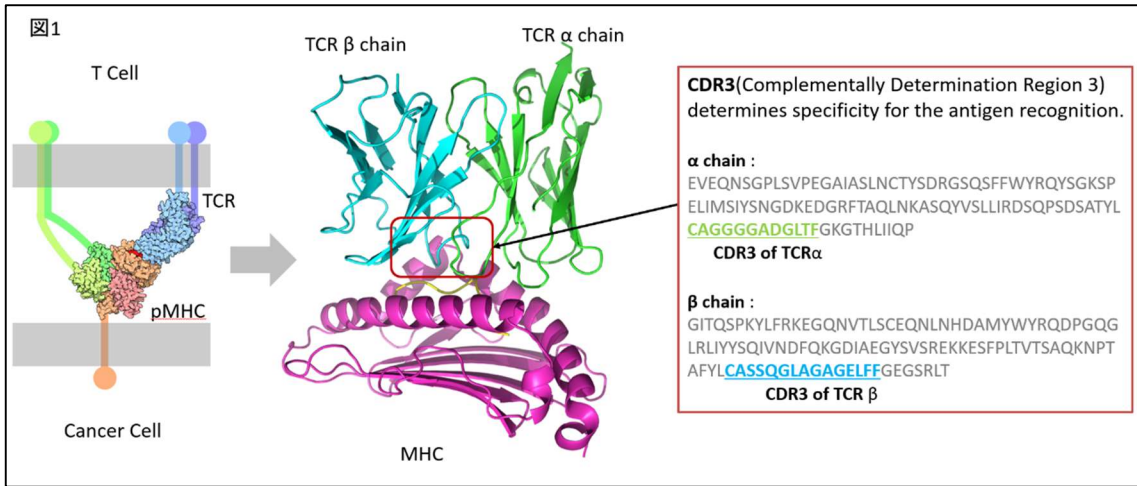
- (1) 免疫一細胞 RNA 計測技術に基づくハイスループット TCR シークエンスおよび遺伝子発現データ解析による実験的結合能評価
- (2) ベイズ統計モデリングに基づく TCR 配列 vs pHLA 結合能 確率的ランドスケープ推定による次期探索 TCR 配列候補自動提案
- (3) 深層ニューラルネットワーク (DNN) 技術およびタンパク立体構造モデリングに基づく数値的結合能予測および確率的 TCR 配列候補群生成

4. 研究成果

本研究においては、上記の研究方法で述べた方針に従って研究を進めた。まず、上記(3)の DNN 学習技術に基づく数値的結合能予測について、既報の数値的結合能予測技術のサーベイを行い、複数の性能の評価を行った。予測モデルの構造に関する知見を得るとともに、特に、予測精度の向上のために必要な訓練用データセットの性質についても知見を得た。次に、(3)の確率的 TCR 配列候補群生成および(2)の次期探索 TCR 配列候補提案に関連して、遺伝的アルゴリズムおよび逐次的モンテカルロサンプリングに基づく生成モデルの構築手法を考案した。提案手法では、超広大な配列候補空間を、高速かつ偏りなく探索することが可能であり、多様な高親和性候補配列の生成を可能としている。また、(1)の実験的結合能の評価に関して、TCR および pHLA の結合能を評価する実験系を構築し、肺がんのサンプルから、がん抗原を認識する T 細胞の TCR 配列を取得し、同 TCR 配列の一部をランダムに改変した配列を作成し、実験的な結合能の測定も行った。また single cell RNA-seq データの取得および解析も進め、TCR 配列および T 細胞の多様性および、がん抗原特異的 T 細胞の性質の解明も進めた。

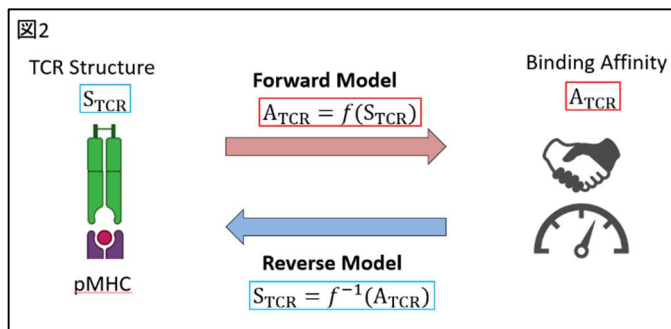
以下、得られた主な成果について詳述する。

まず、逐次的モンテカルロサンプリングに基づく配列生成アルゴリズムの開発について述べる。TCR タンパクは α 鎖 β 鎖からなるヘテロ二量体である (図 1)。それぞれのタンパク鎖には、CDR3 領域と呼ばれるループ状の部位が存在し、これらのループ構造が pHLA 複合体の認識の特異性および結合の親和性にとって重要であることが知られている。基本的に TCR と pHLA の関係は特異的であり、無数の種類の抗原に対応するために、多様な TCR 配列を持つ T 細胞が体内に存在する。この多様性は、各 T 細胞におけるゲノム再編成 (VDJ recombination) によって生み出される。 α 鎖においては V 遺伝子と J 遺伝子の、 β 鎖においては、V 遺伝子、D 遺伝子および J 遺

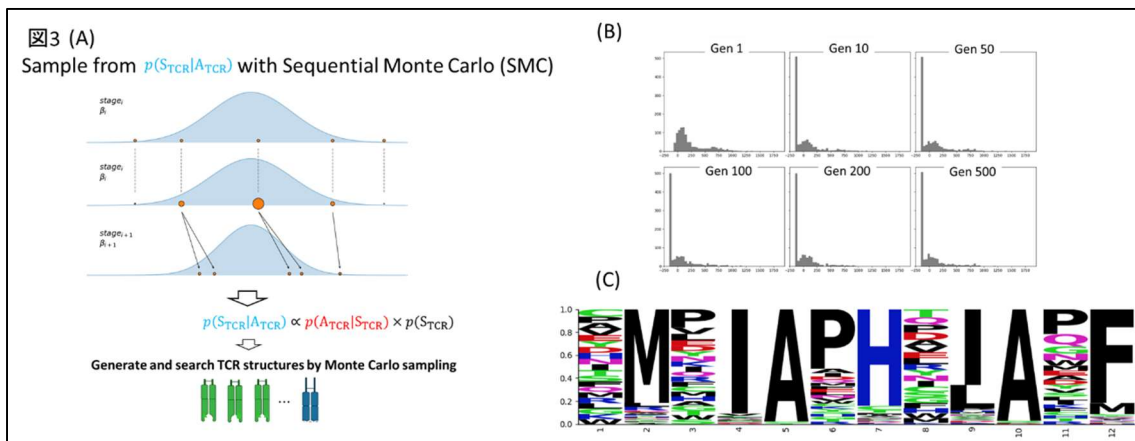


伝子の、ランダムな選択と結合および結合時の遺伝子断片間に生じる、挿入・欠失によるノンテンプレート配列の生成の結果として、超多様な TCR 配列の産生を可能としている。CDR3 は、その中でも超可変領域と呼ばれる領域に位置する。研究背景でも述べたように、CDR3 の配列に限ったとしても 10^{26} 通りを超える膨大な組み合わせ多様性を生じさせる。このような超広大な配列空間の中から、効率よく目的とする pHLA 複合体と親和性の高い候補配列を見出す必要がある、しかしながら、TCR 配列をどのように変化させれば、親和性が向上するかについて明確な指針は存在しない。

この困難を克服するために、我々は、本問題を統計科学におけるサンプリング問題として捉え、ベイズ統計の枠組みを用いることを考えた。図 2 にそのコンセプト図を示す。ここには **Forward model** と **Reverse model** が示されている。前者は、TCR 配列を入力として、pHLA への親和性を予測するモデルであり、後者は、逆に親和性を入力として、TCR 配列・構造を予測するモデルである。本研究の目的としては、**Reverse model** を得ることができれば、そのモデルを基に、高親和性の配列を生成・サンプリングすることができる。しかしながら、**Reverse model** の構成法は自明ではない。一方、**Forward model** については、これまでも様々なモデルや観測法が開発されて来ている。本研究のアイデアは、**Forward model** の出力を基に、結合親和性の条件付き確率分布 $P(A_{TCR}|S_{TCR})$ を求め、ベイズの定理により、TCR 配列についての条件付き確率分布 $P(S_{TCR}|A_{TCR})$ を得ることである。**Forward model** の候補としては、single cell TCR シークエンス等、実際の観測に基づくものや、TCR の配列・構造を入力とし、pHLA との結合能を予測する機械学習モデルに基づくもの、また入力した TCR 配列と類似した TCR タンパクの結晶構造をデータベースから取得しシミュレーションにより構造を改変するテンプレートモデリングに基づく方法などがある。それぞれの **Forward model** の手法について、長所および短所が存在する。我々は、まずトレーニングデータを必要としない、テンプレートモデリングに基づく **Forward model** を用



いて **Reverse model** を構成した。一般に **Reverse model** を表現する、確率分布 $P(S_{TCR}|A_{TCR})$ を陽に得ることは困難である。しかし、この確率分布に従う、配列をモンテカルロ法によりサンプルすることは可能である。我々は、そのために、逐次的モンテカルロサンプラー (SMC: Sequential Monte Carlo sampler) に基づく、配列生成法を構築した (図 3)。SMC では、逐次的に配列を生成するが、ここでは各繰り返しステップを世代と呼ぶ。SMC では、第 N 世代において、第 $N-1$ 世代で得られた P 個の TCR 候補配列に変異を加える、次にそれらの親和性を **Forward** モデルにより予測し、その親和性の大きさに基づいて第 $N+1$ 世代に渡す P 個の候補配列をリサンプルするということを繰り返し、 $P(S_{TCR}|A_{TCR})$ に従う配列を探索・生成する (図 3(A))。図 3(B) は、各世代においてサンプルされた配列のエネルギー値の分布の遷移を示している。図 3(C) は、エネルギー値が低い (親和性が高い) 配列の、アミノ酸配列の分布を示している。図 3(B) に見られるように、エネルギーの分布に、比較的早い段階でエネルギーが低い位置にピークが見取れる、この周辺の配列が最適解である可能性がある一方、局所解にトラップされてしまっている可能性も考えられる。この後に、実際に親和性が高い配列を実験的に検証してさらに絞り込む際ことを考えると、より多様な解を候補として探索しておくことも望ましい。そのため、我々は、より多様な解を効率よくサンプリングするために、クラスタリングに基づくサンプリングアルゴリズムを考案した。そこでは各世代において、配列のリサンプリングを行う前に、配列間の類似度を基に配列をクラスタリングしておき、クラスタごとによりサンプリングを行うことで、多様性を確保した配列のサンプリングを行うことが可能となる。



以上のように、ベイズモデリングに基づく、高親和性候補配列生成法を構築した。一方、配列候補を生成する Reverse model の性能は、Forward model に大きく依存する。上記の実装では、Forward model として、トレーニングデータを必要としないテンプレートモデルを利用していた。その後、既知の高親和性 TCR-pHLA のペアにおいて、TCR β 鎖の CDR3 の中央に位置する、アミノ酸を置換し実験的に親和性 (Luc アッセイ) を計測した。それらの実測値とテンプレートモデルによる親和性の予測値の相関を検討したところ、ピアソン相関係数 0.057 と、低い相関であった。その他 scRNA-seq データ等とも比較を行い、現状のテンプレートモデルでは、対象とする TCR および HLA にもよるが、十分な性能が得られないことが分かった。そこで、代替となる Forward model の候補として、機械学習に基づく親和性予測モデルおよびモデルの学習に必要なトレーニングデータについてサーベイを行い、タンパク質デザインに必要な手法とデータセットについて議論した (Guo and Yamaguchi, Front Bioinform, 2022)。また研究を開始してから、AlphaFold 等のタンパク質立体構造予測手法が発表され、生命科学全体に大きな影響を与えている。我々は、上記の情報も取り込んだマルチモーダル学習に基づく予測手法の開発も進めた。

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Guo Zhongliang, Yamaguchi Rui	4. 巻 2
2. 論文標題 Machine learning methods for protein-protein binding affinity prediction in protein design	5. 発行年 2022年
3. 雑誌名 Frontiers in Bioinformatics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.3389/fbinf.2022.1065703	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2件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 小室裕康、高橋祐介、宮本勇作、松本光善、山本裕崇、白橋幸洋、西田玲奈、宮井まなみ、篠原周一、松井琢哉、岡村文子、福山隆、黒田浩章、松下博和、土井潔、岩田尚
2. 発表標題 HLA-B1501肺癌患者におけるKK-LC-1に対するTCRの取得
3. 学会等名 第38回日本呼吸器外科学会
4. 発表年 2021年

1. 発表者名 小室裕康、篠原周一、福嶋恭啓、杉田裕介、松井琢哉、高橋祐介、岡村文子、村岡大輔、福山隆、浜名洋、岸裕幸、田中雄希、尾上広祐、小野口和英、山下慶子、山口類、黒田浩章、岩田尚、松下博和
2. 発表標題 Single Cell解析を活用した非小細胞性肺癌におけるネオ抗原および癌・精巢抗原特異的なCD8+T細胞集団の同定
3. 学会等名 第26回日本がん免疫学会総会
4. 発表年 2022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
研究分担者	松下 博和 (Matsushita Hirokazu) (80597782)	愛知県がんセンター（研究所）・腫瘍免疫制御TR分野・分野長 (83901)	

6. 研究組織（つづき）

	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	岡村 文子（出町文子） (Okamura Ayako) (10546948)	愛知県がんセンター（研究所）・腫瘍免疫制御TR分野・主任 研究員 (83901)	
研究分担者	吉田 亮 (Yoshida Ryo) (70401263)	統計数理研究所・データ科学研究系・教授 (62603)	
研究分担者	太田 元規 (Ota Motonori) (40290895)	名古屋大学・情報学研究科・教授 (13901)	

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