

科学研究費助成事業 研究成果報告書

令和元年6月25日現在

機関番号：14301
研究種目：若手研究(B)
研究期間：2016～2018
課題番号：16K18474
研究課題名(和文) 難病解明に向けた大規模HLAシーケンスと多重リスク因子の包括的解析法の確立

研究課題名(英文) Development of association analysis method for rare diseases based on comprehensive HLA typing with multiple risk factors

研究代表者
川口 修治 (Kawaguchi, Shuji)
京都大学・医学研究科・助教

研究者番号：00525404
交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：本研究では、多くの難病において関連が示唆されるHLA遺伝子を網羅的に解析するための技術の確立を目指し、高効率・高精度なHLAのシーケンス・タイピング技術を開発した。また、HLAの情報を複数の因子と絡めて複合的な関連解析するための統計手法の開発を進めた。確立した技術をHTLV-1関連脊髄症(HAM)の収集検体に適用した結果、HAM発症に関わる新たなリスクとなるアミノ酸残基を同定した。また、HLA遺伝子のより高精細な解析に向けてシーケンス結果から遺伝子全長配列を最構成する技術を開発した。

研究成果の学術的意義や社会的意義

従来のHLAアレルのタイピング方法は、高コストでありながら得られる精度や情報が少ないため、解析を行うためのサンプル数の不足や、得られたHLAアレルやアミノ酸残基の頻度に歪みが生じるといった問題が存在した。本研究で確立したHLA解析技術を用いることで、高精度でありながら低コストにHLA遺伝子解析を行うことが可能であり、様々な難治性疾患における発症リスクや診断に有用となるバイオマーカーの発見が期待できる。

研究成果の概要(英文)：We have established novel sequencing and typing technologies for HLA genes that can type more comprehensive HLA alleles with high resolution than conventional typing method. We have also developed a method for association analysis that can derive susceptible or protective amino acid residues on HLA protein with considering various risk factors of disease development. By using the developed methods, we carried out HLA typing of 928 patients of HTLV-1 associated myelopathy (HAM), 2,398 asymptomatic HTLV-1 carriers and 2,553 uninfected controls, and conducted association study between HLA genes and HAM development. As a result, we could identify susceptible and protective amino acid residues for HAM development that have different effect with proviral load.

研究分野：統計科学

キーワード：HLA 次世代シーケンサー タイピング アルゴリズム 遺伝統計学 マルチプレックスPCR

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

HLA 遺伝子は免疫応答をつかさどる重要な遺伝子であるため、様々な自己免疫性疾患やウイルス感染に伴い発症する疾患との関わりが報告されている。特に HLA が関連する疾患には、発症原因が共通するにも関わらずその症状が異なる特徴を持つものが多い。例えば IgG4 関連疾患は、自己免疫性膵炎や涙腺唾液腺炎などにおいて血清 IgG4 濃度高値であることが原因とされるが、症状を呈する臓器や組織が個人毎に異なる。ヒト T 細胞白血病ウイルス (HTLV-1) においては、感染者の多くは無症候性の群であり、そのうち約 5% の人が成人 T 細胞性白血病 (ATL) を約 0.3% の人が HTLV-1 関連脊髄症 (HAM、他地域では熱帯性痲痺性対麻痺: TSP と呼ばれる) を発症する。これらの疾患においては、ゲノムにおいてもとりわけ高多型である HLA のアレルが発症リスクに関連していることが示唆されている。

HLA には特に重要な遺伝子として HLA-A、-B、-C、-DRB1、-DQB1 および -DPB1 の 6 種があり、これらの遺伝子型の違いが様々な疾患の発症リスクに関与していることが知られている。そこで発症群と対照群においてこれらの遺伝子におけるアレルの頻度差を比較することで、疾患の発症リスクに重要な感受性・抵抗性を持つ因子を検出できる。従来、アレルの決定には Luminex® 法に代表される PCR-SSOP 法等が多くの場合用いられて来た。この手法は HLA アレルの様々な配列に対応した蛍光ビーズを用意し、これらと検体の蛍光反応パターンからアレルを決定する。しかしながら、この手法はアレル決定に一部の変位箇所しか用いていないため、全ての HLA アレルのタイピングに対応していない。そのため、アレルやアミノ酸残基における観測値と真の頻度との間に誤差が生じる問題が発生する。

近年、次世代シーケンサー (NGS) の登場により全ゲノム配列を高速に決定できるようになった。研究代表者らは全ゲノムシーケンスや Exome で読まれたリードから、HLA アレルを決定する手法を開発しており、公開データによる既報の手法との比較において最も高い性能を示した。一方、各遺伝子の翻訳領域外側の配列をプライマーとして設計し、PCR で増幅することで、ターゲット領域のみを効率的にシーケンスすることができる。しかしながら、HLA 遺伝子の高多型性と遺伝子間の相同性の高さから、通常の方法では、対象の HLA 遺伝子領域の全てのアレルを包括的かつ特異的に増幅可能なプライマー配列を設計することが困難である。その問題に対して、研究代表者らは主要 HLA 遺伝子に対する新たなマルチプレックス PCR プライマーセットとその増幅技術を開発している。このプライマーセットを用いることで、従来に比べ 6 遺伝子間の増幅効率の差を最小限に抑えた上で、漏れなく HLA アレルを増幅できる。さらに、従来法とは違い遺伝子領域の配列が決定できるために、任意の塩基・アミノ酸残基における疾患群・対照群の比較解析といった高解像度解析がおこなえるようになる。

HLA の関連が示唆される難病の解析には、上記のような高度技術を駆使した上で得られる複雑なデータの解析を必要とされる。そこで、本研究では HLA の高度解析技術の確立とこれらの解析結果と様々な外部因子を用いた難病の複合的な統計解析モデルの開発を目指した。

2. 研究の目的

本研究では高度な HLA 遺伝子シーケンス・タイピング技術の確立とそれらの解析結果とゲノムワイドデータや外部因子を複合的に絡めて解析するための統計モデル開発を目指す。特に次世代シーケンサーを用いて HLA 遺伝子の全長配列を決定し、アミノ酸残基の位置・イントロンにおける変位におけるまで高精度な解析を可能とするモデルを構築する。また他のゲノムワイド観測データと、外部因子を変数として HLA との関係性を見いだすモデルを開発する。

確立したモデルの検証には、研究代表者が所属する京都大学大学院医学研究科附属ゲノム医学センターで収集した難治性疾患と対照群の検体を用いる。これらの検体の DNA サンプルより HLA 遺伝子のシーケンス・タイピングを行い、決定した HLA アレルと疾患の関連解析により疾患の発症に関わるアレルおよびアミノ酸残基を同定する。解析により得られた HLA とその他の因子を用いて発症リスクを推定し、ゲノム診断に適用できる混合マーカの開発に繋げる。

3. 研究の方法

データ選別

全国の共同研究機関で採取し、京都大学大学院医学研究科附属ゲノム医学センターで保管されている DNA 検体の HLA シーケンスおよびタイピングを行う。疾患としては、ヒト T 細胞白血病ウイルス (HTLV-1) 感染における、HTLV-1 関連脊髄症 (HAM/TSP) を対象とする。対照群としては無症候性のキャリア群や高侵淫地域における非感染群の検体を用いた。

次世代シーケンサーによる塩基配列解読

上記検体に対する HLA-A、-B、-C、-DRB1、-DQB1 および -DPB1 の 6 遺伝子の塩基配列決定を実施するため Long-PCR による増幅をおこなう。これには研究代表者らが開発した、マルチプレックス PCR プライマーセットを用いた。PCR 増幅産物を断片化し、検体認識用の index 配列を付与した後、Illumina 社の次世代シーケンサー MiSeq を用いてシーケンスした。MiSeq 標準プロトコルでは、一度のランで実施可能なサンプル数は 384 検体までとなっているため、これ

を超える検体においても識別可能なバーコードの開発を行い、HLA 遺伝子シークエンスの低コスト化を図った。

計算科学的アプローチに基づく HLA タイピングと遺伝子配列決定

次世代シークエンサーで読まれたリードから、IPD-IMGT/HLA データベースに登録されている HLA アレルを判定した。判定には研究代表者らが開発した HLA タイピングプログラム(HLA-HD)を用いる。IPD-IMGT/HLA データベースは定期的に更新が行われるため、HLA-HD に対する HLA アレル情報の自動更新機構の開発を進める。また、データベースに登録されている HLA 遺伝子のアレルの 95%は部分的な cDNA 配列しか登録されていないため、次世代シークエンサーで得られたリードから遺伝子全長の配列予測法の開発を進めた。

解析結果を用いた難治性疾患の解析と外部因子データによる多重リスク混合モデルの開発 HLA におけるアミノ酸残基に加えてその他の因子を説明変数に加えた混合モデルとその最適探索法を開発する。開発した解析手法を本研究で得たデータとゲノム医学センターで取得したデータに適用し、その有効性を検証した。結果の医学的な意味付けについてはサンプル取得機関の共同研究者にアドバイスを求めながら進めた。検証には以下の点を重視して行った。

1. 疾患群と対照群とで HLA アレルもしくはアミノ酸残基の有意な違いが得られているか。
2. 他の遺伝子領域が関連して因子が得られていないか。
3. 急速進行例や重症化等においての特異的な変位の有無があるか。

解析については、本課題でシークエンス・タイピングした検体以外も併せて行った。疾患のデータ解析によって得られた知見をもとに、疾患の発症リスク診断となるマーカーを抽出する。

4. 研究成果

本研究開発によって、HLA 遺伝子を大規模にシークエンス・タイピングする技術の確立が達成できた。主要 HLA 6 遺伝子によるマルチプレックスプライマーについては、高均一・高品質な PCR 増幅プロトコルが完成し、特許出願を行った。また、384 検体を超える同時シークエンスに対応するためのバーコードの開発を行った。開発技術を MiSeq シークエンサーによる HLA 遺伝子シークエンスに適用し、およそ 900 検体の同時シークエンスを達成できた。開発バーコードは設計上 1 万以上の検体に対応しており、他のシークエンス技術にも応用が可能である。高精度かつ包括的な HLA アレル種類に対応したタイピングアルゴリズム HLA-HD (雑誌論文[1]) に関しては、開発完了後も改良による更新を重ねており、IPD-IMGT/HLA のデータベースのリリース更新に伴う自動アップデート機能が付加された(図書[1])。また、タイピングプログラムに関してはホームページを通じて公開しており、2019 年 4 月時点において 30 カ国、計 140 (国内 12、国外 128) の大学・企業・病院・研究施設からの利用申請があった。

確立した技術により、HLA シークエンス・タイピングを行ったによる解析検体数は他プロジェクトも含めて 7,000 検体を超えた。さらに、全ゲノム解析データを用いた HLA-HD によるタイピングも 3,000 検体以上に及ぶため、すでに 10,000 検体以上の HLA タイピング実績が得られた。このうち、928 人は HAM/TSP 患者の検体であり、2,398 人の無症候性キャリア群、2,553 人の高侵淫地域における非感染群における HLA タイピング結果と併せて、HAM/TSP の発症と HLA 遺伝子とにおける関連解析をおこなった。その結果、HAM/TSP 患者と無症候性キャリア群の間で有意に差がある感受性・抵抗性 HLA アレルおよびアミノ酸残基位置とその種類を同定した(学会発表[1])。このアミノ酸に対して HAM/TSP 患者、キャリアの計 2500 名における従来から発症リスクの一つとされてきた末梢血単核細胞中の HTLV-1 感染率(プロウイルス量)測定結果を加えた混合モデルによる解析を行ったところ、HAM/TSP の発症リスクにおいてプロウイルスとは異なる効果を持つことが明らかとなった。この成果は、全国に 100 万人以上存在する、HTLV-1 感染者に対する HAM/TSP 発症リスクを測定する新たなバイオマーカーへの開発を可能とし、HAM/TSP の発症も早期診断に繋がることを期待できる。

HLA 遺伝子の完全長配列リファレンスの作成に関しては、MiSeq で得られた HLA 遺伝子のショートリードシークエンスデータから、遺伝子全長の配列を復元する手法の開発を行った。HLA-HD のタイピング結果を基に、IPD-IMGT/HLA にある遺伝子全長配列データと比較し、アライメントとマッピングを逐次的に更新し、配列を復元するアルゴリズムを開発した。手法の改良には 30 程の少数サンプルについては、PacBio によるロングリードのシークエンス結果も用いた。今後、開発した手法をこれまでの研究で得られたシークエンス検体に適用し、データベースを構築する。

5. 主な発表論文等

[雑誌論文](計 1 件)

[1] Kawaguchi, S., Higasa, K., Shimizu, M., Yamada, R. and Matsuda, F. HLA-HD: An accurate HLA typing algorithm for next-generation sequencing data. *Hum Mutat.* 38:788-797, 2017.

〔学会発表〕(計2件)

[1] 川口 修治, ペノヴァ マリナ, 川口 喬久, 清水 正和, 高橋 めい子, 日笠 幸一郎, 松田 文彦, HTLV-1 高浸淫地域における大規模 HLA 解析による感染および HAM/TSP 発症リスクの関連解析, 第五回日本 HTLV-1 学会学術集会, 2018 年 9 月, 東京

[2] 川口 修治, 日笠 幸一郎, 清水 正和, 山田 亮, 松田 文彦, 次世代シーケンサーと国際データベースを用いた高効率・高精度な HLA 遺伝子群の包括的タイピング技術の確立, 日本人類遺伝学会第 63 回大会, 2018 年 10 月, 横浜

〔図書〕(計2件)

[1] Kawaguchi, S., Higasa, K., Yamada, R., Matsuda, F., Comprehensive HLA Typing from a Current Allele Database Using Next-Generation Sequencing Data. *Methods Mol Biol.*, 2018;1802:225-233.

[2] 川口 修治, 日笠 幸一郎, 山田 亮, 松田 文彦, 次世代シーケンスデータと国際データベースを用いた HLA 遺伝子型の決定, 最新医学 74 巻 2 号, 214-219, 2019

〔産業財産権〕

出願状況 (計2件)

名称: 被験者の HLA 遺伝子のアレルペアを判定する方法

発明者: 松田 文彦, 川口 修治, 日笠 幸一郎, 山田 亮

権利者: 京都大学

種類: 特許権

番号: 特願 2016-257041

出願年: 2016 年

国内外の別: 国内

名称: HLA 遺伝子の PCR プライマーセット及びそれを用いたシーケンス法

発明者: 松田 文彦, 川口 修治, 清水 正和, 日笠 幸一郎

権利者: 京都大学

種類: 特許権

番号: 特願 2017-024397

出願年: 2017 年

国内外の別: PCT 出願済み (PCT/JP2018/4703)

〔その他〕

HLA-HD 利用ホームページ

<https://www.genome.med.kyoto-u.ac.jp/HLA-HD/>

6. 研究組織

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。