Human leukocyte antigen susceptibility map for SARS-CoV-2

Austin Nguyen, Julianne K. David, Sean K. Maden, Mary A. Wood, Benjamin R. Weeder, Abhinav Nellore, Reid F. Thompson

DOI: 10.1128/JVI.00510-20

SARS-CoV-2のヒト白血球抗原感受性マップ

ABSTRACT

3つの主要組織適合性複合体（MHC）クラスI遺伝子（ヒト白血球抗原[HLA]A、B、およびC）にわたる遺伝的変異は、コロナウイルス疾患2019（COVID-19）の原因ウイルスであるSARS-CoV-2への感受性および重症度に影響を与える可能性がある。我々は、すべてのSARS-CoV-2ペプチドに対する145のHLA -A、-B、および-C遺伝子型にわたるウイルスペプチド-MHCクラスI結合親和性の包括的なin silico解析（コンピュータ上での解析）を実行します。我々はさらに、4つの一般的なヒトコロナウイルスへの事前曝露によって付与される交差防御免疫の可能性を探る。SARS-CoV-2プロテオームは正常にサンプリングされ、HLA対立遺伝子の多様性によって提示される。しかしながら、HLA-B\*46:01はSARS-CoV-2に対する結合ペプチドの予測値が最も少ないことがわかった。これは、この対立遺伝子を持つ個体は、以前にSARSに対して示されたように、COVID-19に対して特に脆弱である可能性を示唆している。逆に、HLA-B\*15:03は、共通のヒトコロナウイルス間で共有されている高度に保存されたSARS-CoV-2ペプチドを提示する最大の能力を示し、これは交差防御的なT細胞ベースの免疫を可能にする可能性があることを示唆していることを発見した。最後に、現在のパンデミックの状況において、疫学的な影響を及ぼす可能性のあるHLA型の世界的な分布を報告する。

Importance

個人の遺伝的変異は、集団全体のウイルスに対する異なる免疫反応を説明するのに役立つかもしれません。特に、HLAの変異がCOVID-19の経過にどのように影響するかを理解することは、COVID-19のリスクが高い個人を特定するのに役立つだろう。HLAタイピングは、迅速かつ安価に行うことができる。HLAタイピングとCOVID-19検査を組み合わせることが可能であれば、集団におけるウイルスの重症度の評価を向上させることができる。COVID-19の原因となるウイルスであるSARS-CoV-2に対するワクチンが開発された後、リスクの高いHLA型を持つ人々を優先的にワクチン接種の対象とすることが考えられる。

Introduction

近年，新型のベータコロナウイルス（重症急性呼吸器症候群コロナウイルス2（SARS-CoV-2））が世界的な病原体として出現し，2020年1月に世界保健機関（WHO）は国際公衆衛生緊急事態を宣言した．SARS-CoV-2は、エンベロープ型陽性鎖RNAウイルスで構成される大型コロナウイルスファミリーの中で7番目に遭遇した株であり、軽度の感冒から重度の人獣共通感染症である中東呼吸器症候群（MERS-CoV）や重症急性呼吸器症候群（SARS-CoV）に至るまでの呼吸器疾患を引き起こします。2020年4月現在、コロナウイルス疾患19（COVID-19）の推定または確定症例は世界で100万人を超え、総死亡者数は5万人を超えています。COVID-19の重症度と死亡率は、年齢や心血管疾患や肺疾患を含む多くの併存疾患によって上昇するようですが、感染者の約80％は軽度の症状を呈しています。SARS-CoVやMERS-CoVと同様に、小児は本疾患への感受性が低いようである；成人と同様の感染率にもかかわらず、重症化または重症化した小児症例はわずか5.9％であり、これはおそらく小児ではACE2受容体の結合能が低いか、あるいは一般的に抗ウイルス抗体のレベルが高いためであろう。SARS-CoV-2と他のコロナウイルス、特にSARS-CoVやMERS-CoVとの間には、ゲノム反応や免疫系反応などの類似点があり、現在進行中の研究テーマとなっています。

ウイルス抗原提示経路の重要な構成要素であるヒト白血球抗原（HLA）対立遺伝子は、これまでの研究で、ウイルス感受性と疾患の重症度に差があることが示されています。例えば、近縁種の SARS-CoV に起因する疾患は、HLA-B\*46:01 遺伝子型の人で重症度が高くなることが示されています。HLA遺伝子型と疾患の重症度との関連は、他のいくつかの無関係なウイルスにも広く及んでいる。例えば、ヒト免疫不全ウイルス1（HIV-1）では、特定のHLA型（HLA-A\*02:05など）は血清転換のリスクを低下させる可能性があり、デングウイルスでは、特定のHLA対立遺伝子（HLA-A\*02:07、HLA-B\*51など）は、タイの民族間での二次疾患の重症度の増加と関連している。

COVID-19パンデミックの詳細な臨床像が明らかになりつつある一方で、SARS-CoV-2に対する免疫応答における個人の遺伝的変異の役割については未回答のままである。我々は、個々のHLA遺伝子型がT細胞を介した抗ウイルス反応を異なる形で誘導し、疾患の経過や感染経路を変化させる可能性があると仮説を立てている。本研究では、SARS-CoV-2プロテオーム全体について、145の異なるHLAタイプのウイルスペプチド-MHCクラスI結合親和性の包括的なin silico解析を行った。

Result

あるHLA対立遺伝子が抗ウイルス反応を引き起こす可能性を調べるために、SARS-CoV-2プロテオーム（n=48,395個のユニークなペプチド）から可能性のあるすべての8～12-merのHLA結合親和性を評価した。その後、プロテアソーム切断を介してMHCクラスI抗原処理経路に入ると予測されなかった16,138個のペプチドを、さらなる検討から除外した。残りの32,257個のペプチドについて、合計145種類の異なるHLAタイプについて結合親和性予測を繰り返し、ここでは、アレルあたりのプロテオーム提示のSARS-CoV-2特異的な分布を示す（予測結合親和性閾値＜500nM、図1、補足表S1）。重要なことに、我々は、SARS-CoV-2抗原提示のために考えられる能力は、集団におけるHLA対立遺伝子頻度とは無関係であることに注意してください。我々は、SARS-CoV-2に対する結合ペプチドの予測数が最も少ないHLA対立遺伝子としてHLA-B\*46:01を同定した。我々は、密接に関連するSARS-CoVプロテオームについても同様の解析を行ったが、同様に、HLA-B\*46:01が最も少ないSARS-CoVペプチドを提示することが予測され、この対立遺伝子が重篤な疾患と関連する以前の臨床データと一致していることに注目した。

一般的なヒトコロナウイルス（HKU1、OC43、NL63、および229Eなど）への事前曝露によってもたらされる交差防御免疫の可能性を評価するために、次に、保存性の高い線状エピトープを同定するために、多様なコロナウイルス亜属にまたがるSARS-CoV-2プロテオームの保存性を特徴づけることを試みた。既知のすべてのヒトコロナウイルスを含む34の異なるアルファおよびベータコロナウイルスの5つの必須ウイルスコンポーネント（ORF1ab、S、E、M、およびNタンパク質）の参照プロテオーム配列データをアラインメントした後、48の高度に保存されたアミノ酸配列スパンを同定しました（補足データファイル1）。密接に関連したペプチド間で交差保護免疫を推論することは困難であることを認識し、我々は同一のペプチドの一致にのみ注意を払った。保存された配列の中で、44個のSARS-CoV-2配列はそれぞれ、少なくとも1つの他の一般的なヒトコロナウイルス内にも存在する少なくとも1つの8〜12merの線形ペプチドエピトープを生成すると予想されます（補足表S2、図2）。合計で564個のそのような8〜12-merペプチドが、対応するOC43、HKU1、NL63、および229E配列（それぞれ467、460、179、および157ペプチド）と100％の同一性を共有することが見出された（補足表S3）。

MHCクラスI抗原処理経路を介して生成されると予想されるこれらの潜在的な交差保護ペプチドのサブセットについて、145の異なるHLA-A、-B、および-C対立遺伝子にわたって結合親和性予測を行った。上記のように、我々は、これらの保存されたペプチドのアレルあたりの提示のSARS-CoV-2-特異的な分布を実証する。我々は、対立遺伝子HLA-A\*02:02、HLA-B\*15:03、およびHLA-C\*12:03が保存ペプチドのトップ提示者であることを見出した。逆に、56種類のHLA対立遺伝子は、SARS-CoV-2保存ペプチドのいずれに対しても評価できる結合親和性（<500nM）を示さなかったことから、他のヒトコロナウイルスからの交差防御免疫の可能性が同時に欠如していることが示唆された。我々は、特にHLA-B\*46:01がこれらの対立遺伝子の一つであることに注目している。また、保存されたペプチド提示能力は、集団におけるHLA対立遺伝子頻度とは無関係であることにも注意してください（図3）。さらに、我々は、SARS-CoV-2プロテオームの保存性とその予測されるMHC結合親和性との間に評価できるほどのグローバルな相関関係は見られず、コロナウイルスエピトープを提示する能力に対する選択的圧力の欠如を示唆している（p=0.27 [フィッシャーの正確検定]、補足図S2）。

さらに、SARS-CoV-2プロテオームの特定の領域がMHCクラスI経路によって異なる発現を示しているかどうかにも興味を持っていた。そこで、我々はプロテオーム全体の抗原提示能の分布を調査し、最も保存されているペプチド配列を強調した（図4）。プロテオーム全体を通して、HLA-AおよびHLA-C対立遺伝子は、それぞれSARS-CoV-2抗原提示能が相対的に最大および最小であった。しかし、3つの主要なクラスI遺伝子のそれぞれは、プロテオーム全体でペプチド提示の非常に類似したパターンを示した（補足図S3）。

我々 はさらに、ペプチド提示は、ウイルスのライフサイクル中のペプチド生成の推定時間に依存しないようであることに注意してください、早期および後期の両方の SARS-CoV-2 ペプチドのペプチド提示の区別がつかないレベルで（補足図S4）。

現在のCOVID-19パンデミックの世界的な性質を考慮して、我々は、T細胞ベースの免疫応答をサポートするSARS-CoV-2エピトープのレパートリーを生成するのに最も適した（または最も適していない）HLA対立遺伝子の集団レベルでの分布を記述することを求めた。ここで調査した145種類の対立遺伝子の全セットについて、個々のHLA対立遺伝子頻度のグローバルマップを提示する（補足データファイル2）が、我々は特に、3つの最高提示型（A\*02:02、B\*15:03、C\*12:03）と3つの最悪提示型（A\*25:01、B\*46:01、C\*01:02）HLA-A、-B、および-C対立遺伝子のグローバルな分布を強調している（図5）。すべての対立遺伝子頻度は国別に集計されているが、対立遺伝子頻度ネットデータベースで入手可能なHLAデータの分布を暗黙的に反映していることに注意。

最後に、我々は、ほぼすべての個体が2つのHLA-A/B/Cハプロタイプを有しており、3つのHLA-A/B/Cハプロタイプは少なくても6つの異なる対立遺伝子を構成していることを認識している。我々は、SARS-CoV-2の提示における対立遺伝子特異的変動がHLAハプロタイプ全体にも、また個々のHLA遺伝子型全体にも及んでいるかどうかについて検討した。SARS-CoV-2エピトープ提示能力が最も高く（HLA-A\*02:02、HLA-B\*15:03、HLA-C\*12:03）、最も低い（HLA-A\*25:01、HLA-B\*46:01、HLA-C\*01:02）と予測された6つの代表的な対立遺伝子については、異なるハプロタイプ間で提示能力に大きなばらつきがあるにもかかわらず、これらの差はハプロタイプレベルで有意であることが示された（図6）。145のすべての対立遺伝子についてのハプロタイプレベルのデータは、補足図S5および補足データファイル2に含まれている。次に、完全なHLA遺伝子型データを有する3,382人を同定し、SARS-CoV-2プロテオームからのペプチド提示能力に大きなばらつきがあることを指摘した。

Discussion

この研究は、我々の知る限りでは、MHC-ペプチド結合親和性予測因子を用いて、広範囲のHLA対立遺伝子にまたがるウイルスプロテオーム提示をアレル単位で評価した初めての研究である。

この研究はまた、コロナウイルスの配列保存性とMHCクラスI抗原提示との関係を紹介する。我々は、個々のHLA、ハプロタイプ、および完全な遺伝子型の可変性がSARS-CoV-2感染に対応する能力に影響を与える可能性が高いことを示し、我々は、以前にSARS-CoVで示されたように、より重篤な感染に関連する可能性のある特定の対立遺伝子（例えば、HLA-B\*46:01）に注目している（49）。実際、我々はさらに、SARS-CoVとSARS-CoV-2のペプチド提示を比較し、HLAタイプ全体で2つの間の高度な類似性に注意してください。最後に、これは、現在のパンデミックの状況下での潜在的な疫学的な影響を持つ HLA 型とハプロタイプのグローバルな分布を報告する最初の研究です。

我々は一般的に、集団におけるHLA対立遺伝子頻度とSARS-CoVまたはSARS-CoV-2ペプチドを結合する対立遺伝子の能力との間には、ウイルス複製サイクル中のペプチド産生の推定時期にかかわらず、相関関係がないことを発見した。

ヒトコロナウイルスエピトープの豊富さと免疫応答との間の関係を明示的に報告した研究には気づいていないが、初期のペプチド抗原はCD8+ T細胞応答を生成する可能性が高く、一方、抗体およびCD4+ T細胞応答は、ウイルス中のより高いペプチドの豊富さを有する後期のmRNA発現を標的とする可能性が高いことを示唆するデータが、ワクシニアウイルスに存在する。

しかしながら、我々は、我々の研究にはいくつかの制限があることに注意してください。まず第一に、我々の結合親和性予測の一握りが実験的に検証されたSARS-CoVペプチド（補足表S4）で実証されたことに留意しながらも、これは完全にインシリコ研究であることを認識している。現時点では、実際のCOVID-19集団について、個人レベルのHLAタイピングおよび臨床転帰データを得ることができないため、提示されたデータは理論的なものであり、それがベースとなっているMHC結合親和性予測ツール（複数可）に暗黙のうちに同じ制限の多くを受ける。そのため、年齢や臨床合併症などの既知の疾患修飾危険因子と比較して、HLA型の相対的な重要性を評価することはできない。さらに、ペプチド-MHC結合親和性は、その後のT細胞応答の予測因子としては限定的であり、我々はここではT細胞応答を研究していないことに注意してください。このように、我々は元々の抗原性の罪のような現象を探求するには不十分であり、密接に関連した感染症への事前の曝露が、新規感染症の設定でT細胞のアレルギーや免疫原性発生の引き金となる可能性がある。我々は、よく研究されている145のHLA対立遺伝子の限られたセットのみを調査したが、この解析は、より広範な遺伝子型にわたって実施できることに留意した。さらに、我々は遺伝子型の不均一性やSARS-CoV-2のin vivoでの進化を評価していない。これは、提示されるウイルスエピトープのレパートリーを変更したり、HLAに依存しない方法で病原性を調節したりする可能性がある（https://nextstrain.org/ncov）。また、他のタンパク質（例えば、アンジオテンシン変換酵素2[ACE2]や膜貫通型セリンプロテアーゼ2[TMPRSS2]など、SARS-CoV-2のプライミングおよび細胞侵入に不可欠な宿主タンパク質で、宿主と病原体の界面を調節するための個体レベルの遺伝的変異の可能性についても、我々は言及していない。

今回発表した所見は、臨床的に検証されるまでは、いかなる臨床目的にも使用すべきではない。しかし、この時点では、HLA検査を臨床試験に統合し、HLAタイピングとCOVID-19検査を組み合わせることで、集団におけるウイルス重症度の予測因子をより迅速に開発し、展開し、将来のワクチン接種戦略を遺伝子型的にリスクの高い集団に合わせて調整することを推奨している。このアプローチは、他の広範なウイルスの管理にも影響を及ぼす可能性があります。

METHOD AND MATERIALS

配列検索とアラインメント

既知のすべてのヒトコロナウイルス（SARS-CoV、SARS-CoV-2、MERS-CoV、HKU1、OC43、NL63、および229E）を含む、広範な属および亜属分布から、34の異なるが代表的なアルファおよびベータコロナウイルスのそれぞれについて、完全なポリタンパク質1ab（ORF1ab）、スパイク（S）タンパク質、膜（M）タンパク質、エンベロープ（E）タンパク質、およびヌクレオカプシド（N）タンパク質の配列が得られました。FASTAフォーマットのタンパク質配列データ（完全なアクセッション番号リストは補足表S5に掲載）は、国立生物工学情報センター（NCBI）から取得した。各タンパク質クラス（すなわち、ＯＲＦ１ａｂ、Ｓ、Ｍ、Ｅ、Ｎ）について、３４個の全てのコロナウイルス配列をＣｌｕｓｔａｌ Ｏｍｇｅｍａｖｉｃａｌ ｖ１． 4 マルチシーケンスアライナーツールClustal Omega v1.2.2.4を用いて、配列タイプ[Protein]、出力アラインメントフォーマット[clustal\_num]、delign[false]、mBed型クラスタリングガイドツリー[true]、mBed型クラスタリング反復回数[true]、結合反復回数[0]、最大ガイドツリー反復回数[-1]、最大HMM反復回数[-1]を設定してアラインメントを行った。ウイルスペプチドの産生時期を推定する目的で、ORF1aおよびORF1bペプチドを "早期 "に分類し、サブゲノムmRNAによって産生される他のすべてのペプチドを "後期 "に分類した。

保存ペプチドの評価

アライメントされた配列を Jalview v. 2.1.1 (71) にインポートし、以下のアラインメントアノテーションを自動生成した。1）配列コンセンサス（カラムあたりのモード残基の割合として計算）、2）配列保存性（0-11）、3）アライメント品質（0-1）、4）占有率（各位置の全残基のBLOSUM62比の正規化された和として計算）、各位置のアライメントされた残基の数（ギャップを含まない）として計算。すべての場合において、ヒトコロナウイルスのみ（ｎ＝７）、すべてのベタコロナウイルス（ｎ＝１６）、およびα-およびベタコロナウイルスの複合配列（ｎ＝３４）の３つのグループのそれぞれについて配列保存性を評価した。アラインメントされたSARS-CoV-2配列およびすべてのアノテーションは、その後の解析のために手動でエクスポートされた。保存されたヒトコロナウイルスペプチドは、SARS-CoV-2および4以上の他のヒトコロナウイルス配列とコンセンサス配列が一致する、長さ8以上の連続したアミノ酸を有するペプチドとして定義した（補足表S2）。また、これらの保存ペプチドについて、SARS-CoV-2と他の4つの共通ヒトコロナウイルス（すなわち、OC43、HKU1、NL63、229E）との間で同一のアミノ酸配列を共有する8～12-merの構成要素数を評価した（補足表S3）。すべてのペプチドについて、ヒト、ベータ、および複合保存性スコアを、カスタムR v.3.6.2スクリプトを用いて、配列保存性の平均値（関連する場合はギャップペナルティを差し引いた値）として得た（https://github.com/pdxgx/covid19 を参照）。

ペプチド-MHCクラスI結合親和性予測

SARS-CoV-2およびSARS-CoVプロテオーム全体のFASTAフォーマットの入力タンパク質配列をNCBI RefSeqデータベースからアクセッション番号NC\_045512.2およびNC\_004718.3で取得した。これらの配列のそれぞれを８〜１２マーにクマー化し、プロテオーム全体でのＭＨＣクラスＩペプチド結合親和性を評価した。MHCクラスIの結合親和性予測は、以前に記載したようにグローバルな対立遺伝子頻度データが利用可能であった145の異なるHLA対立遺伝子を用いて実施した（補足表S5を参照）netMHCpan v4.0で、結合親和性予測を含めるために'-BA'オプションを使用し、長さ8〜12のペプチドを指定するために'-l'オプションを使用した（補足表S1）。文字'|'をその配列に含むペプチドについては、結合親和性は予測されませんでした。追加のMHCクラスI結合親和性予測を、MHCnuggets 2.3.2およびMHCflurry 1.4.3の両方を使用して、66個のMHCflurry支持対立遺伝子（--list-支持された対立遺伝子、補足表S6）すべてについて行った（補足表S7およびS8、補足図S7〜S10）。我々はさらに、これらのペプチドのリストを、Immune Epitope Database（補足表S4）に存在する既存の実験的に検証されたSARS-CoVエピトープと相互参照した。次に、3つのツールで共有されている66のサポートされている対立遺伝子について、対立遺伝子の連合セットを取り、対立遺伝子の連合セットに一致するペプチド-対立遺伝子ペアをフィルタリングすることにより、コンセンサス結合親和性予測を実行した。アレルごとのプロテオーム提示のSARS-CoVおよびSARS-CoV-2特異的分布については、500nM以上の結合予測値を有するすべてのペプチド-アレルペアを除外した。すべての場合において、我々は、プロテアソーム切断（ペプチドのC末端の）を介して正規のMHCクラスI抗原処理を受けることが予測されなかったペプチドをさらに除去するために、0.1の切断閾値を有するnetchop v3.0 "C-term "モデルを使用した。

全世界のHLA対立遺伝子およびハプロタイプ頻度

HLA-A, -B, -C対立遺伝子およびハプロタイプの頻度データは、Alle Frequency Net Databaseから、101カ国805の異なる集団、2628の異なるメジャー/マイナー（4桁）対立遺伝子、20,478の異なるハプロタイプに対応するデータを取得した（https://github.com/pdxgx/covid19）。また、HLA型が145のHLA対立遺伝子に限定されている3,382人のHLA遺伝子型データも確認した。母集団の対立遺伝子およびハプロタイプ頻度データは、全構成集団の対立遺伝子またはハプロタイプ頻度を母集団のサンプルサイズで加重した平均値として国別に集計したが、母集団の代表的な民族人口統計学的サイズは考慮していない。全世界の対立遺伝子頻度マップは、rworldmap v1.3-6パッケージを使用して作成され、全世界の対立遺伝子頻度とハプロタイプ頻度の推定値は、国ごとの対立遺伝子頻度とハプロタイプ頻度の平均値として計算され、2005年の各国の人口で加重された。

データの利用可能性

ソースコードは MIT ライセンスのもと https://github.com/pdxgx/covid19 で入手可能です。

補足データファイル4は

https://github.com/pdxgx/covid19/blob/master/supporting\_data/Appendix\_4.zip にあります。

謝辞

Christopher Loo博士とJeffrey Barnet博士には、原稿の批評的な読解に感謝する。

Jonah Sacha博士とPaul Spellman博士の有益な議論に感謝します。

免責事項

内容は、米国退役軍人局や米国政府の見解を示すものではありません。

資金調達

RFTは、米国退役軍人局の支援を受けています。

1IK2CX002049-01、Sunlin & Priscilla Chou財団。

参考文献

（以下、省略）