

Twist Comprehensive Viral Research Panel を用いた新規ウイルス検出

要約

新しいウイルス種およびウイルス株からの感染は世界の公衆衛生に深刻で繰り返し発生する脅威をもたらします。これらの脅威への対応は、ウイルス病原体の同定及びアッセイの開発能力に大きく依存します。これはヒトウイルスの多様かつ急速に進化する配列空間を考慮すると、重大な課題となります。この問題は、動物ウイルスからの新規感染によってさらに悪化し、既知のヒト感染種との配列類似性がほとんどない可能性があります。理論的には次世代シーケンス (NGS) アプローチにより、完全に新しいウイルス病原体を検出できます。しかし、その適用は、ホストの配列の混入によって制限されており、ウイルステンプレートを完全にカバーするにはかなりのシーケンス深度が必要です。ここでは、非常に多様なウイルス株を検出できるハイブリダイゼーションベースのターゲットエンリッチメントソリューションである Twist Comprehensive Viral Research Panel について解説します。今回の結果から、Twist Comprehensive Viral Research Panel が、新規ウイルスの発見と監視の汎用的なソリューションとなることが浮き彫りになりました。

はじめに

近年では、人口の拡大と自然生息環境との相互作用及び自然生息環境の改変に伴ってウイルス流行の頻度が高くなっています (Woolhouse & Gowtage-Sequeria, 2005; Chomel, Belotto, & Meslin, 2007)。新しいパンデミックウイルスへの不安に対し、新規及び再発現中のウイルス病原体の発見及び監視用のツールを改善する必要があります。

新たに発現したウイルスの検出は、変異、組換え及び再配列の確率が高いため課題があります (Holland et al., 1982)。リアルタイム定量ポリメラーゼ連鎖反応 (qRT-PCR) などの一般的な診断アッセイによる検出の成否は、感染性病原体の演繹的な知識に依存します。対照的に、次世代シーケンス (NGS) などの仮説のない検出法を用いて、リファレンス株と大きく異なるウイルス病原体を同定することができます。NGS の偏りのないアプローチは、明確な診断上の利点をもたらす一方で、患者検体における高濃度のヒト宿主核酸の混入により分析感度が低下し影響を及ぼすため、新規ウイルスの発見への適用において大きな障害となります (Chiu 2015)。したがって、ウイルス濃縮の戦略が、新規ウイルス発見に対する NGS の分析感度を改善すると考えられます。

当社のソリューション Twist Comprehensive Viral Research Panel は、ハイブリダイゼーションベースのターゲットエンリッチメントを用いた NGS アプローチであり、新たなウイルス病原体の発見と特性評価を容易にします。このキャプチャシステムを、最近シンガポールで発見された Rousettus コウモリコロナウイルス GCCDC1 (RoBat-CoV GCCDC1) (Paskey et al., 2020)、及び 2020 年 6 月のブタでの

アウトブレイクから分離された新規の H1N1 型インフルエンザ株 (Sun et al., 2020) を含む、Twist Comprehensive Viral Research Panel のデザイン後に出現したウイルス株に由来する合成コントロールに適用しました。また、既知の単一塩基レベルで置換された合成コントロールを作製し、パネルの濃縮効率を検討しました。このアプリケーションノートでは、以下について説明します。

1. リボソーム除去後の既知のウイルス種の高感度検出 (ウイルス数 1,000 コピー未満)。
2. RoBat-CoV GCCDC1 及び H1N1 プタインフルエンザから得られた新規ウイルス配列の濃縮及び全配列の特性解析。
3. 合成改変された HH1N1 ヘマグルチニン (HA) フラグメントのセットを用いたハイブリッドキャプチャシステムのミスマッチ許容性。
4. One Codex プラットフォームを用いた、菌株の演繹的な知識なくシミュレートされた感染における新規 H1N1 プタインフルエンザ配列の同定。

結果

Twist Comprehensive Viral Research Panel には、ウイルス配列のハイブリダイゼーション キャプチャ用の 1,052,421 の固有の 120 塩基対 (bp) ピオチン化プローブが含まれています。Twist Comprehensive Viral Research Panel の設計では、ヒト感染性ウイルス 3,153 株及び 15,488 の異なる株 (一本鎖 [ss]DNA、二本鎖 [ds]DNA、dsRNA 及び ssRNA ウイルスを含む) の多様性を組み入れ、NGS による検出及び特性評価のため広範なウイルス配列の濃縮を可能にしました。この探索的パネルの性能を、自然及び人工の配列変異源に対して評価し、出現したウイルスを検出する能力を用いて評価しました。

Twist Comprehensive Viral Research Panel を用いて、まず、同時感染サンプルとして SARS-CoV-2 (PN 102024) 及びヒトライノウイルス 89 (PN 103006) のリファレンス配列に由来する 2 つの RNA コントロールについて検討しました。NEBNext rRNA Depletion Kit v2 (NEB #E7400, #E7405) を用いた rRNA 除去後、[Twist Total Nucleic Acids Library Preparation Kit for Viral Pathogen Detection and Characterization protocol](#) を用いて、ヒトキャリア RNA バックグラウンドでウイルス力価 1,000 及び 1,000,000 コピーの TruSeq ライブラリを調製しました。その後、後述の [Twist Standard Target Enrichment のワークフロー](#) に記載の材料及び方法に従い Twist Comprehensive Viral Research Panel を用いてハイブリッドキャプチャを実施しました。高力価 (1,000,000 コピー) では、両ウイルステンプレートに対して 99.9% を超えるカバレッジが得られ、シーケンスリードの 15% 以上がウイルステンプレートにマッピングされました。低力価 (1,000 コピー) でも、ウイルステンプレート配列の 90% 以上が回収され、リードの約 0.1% がウイルステンプレートにマッピング

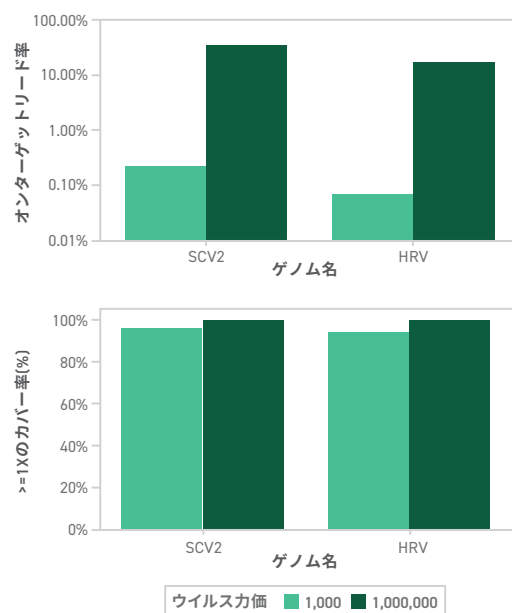


図 1: 同時感染サンプルとしてリボソームの除去後にキャプチャした合成 SARS-CoV-2(SCV2) 及びヒトライノウイルス (HRV) RNA コントロールのオンターゲット率(上) および 1x 以上のカバー率(下)。キャプチャをウイルス力価 1,000 及び 1,000,000 で示す。

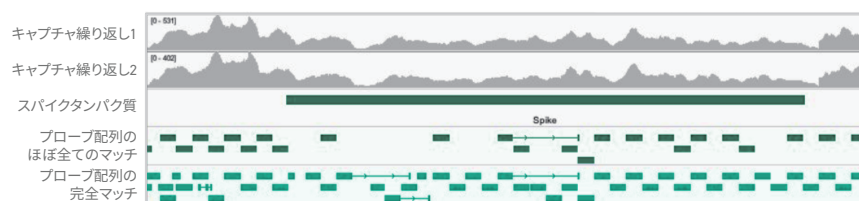


図 2: ゲノムブラウザビューは、RoBat-CoV GCCDC1 ゲノムのスパイクコード領域周囲のリード密度を示している。「クローズプロブのマッチ」には約 110 以上のヌクレオチドマッチと 5 個未満のミスマッチが含まれる。

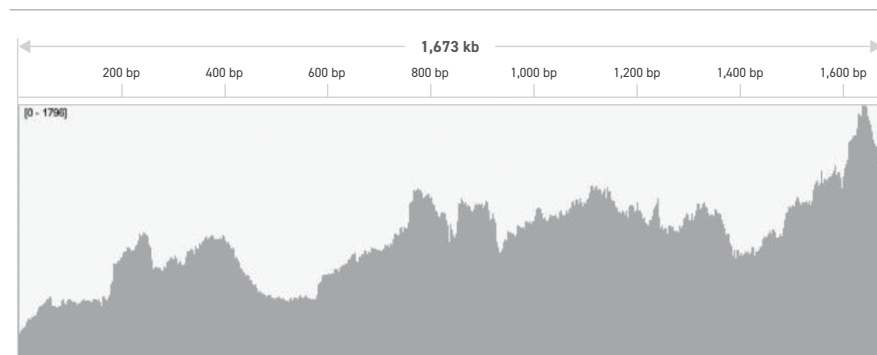


図 3: H1N1 インフルエンザ由来の HA 変異体 (Genbank Accession 番号: MN416597) からのリード密度のゲノムブラウザの表示。ほとんどの塩基において、HA 変異体は高深度(塩基の >200x のカバレッジが 100%)であった。

されました(図 1)。これらの結果から、Twist Comprehensive Viral Panel は、ターゲット領域が広くオフターゲットキャプチャを大量に生成するにもかかわらず、高感度のウイルス検出能力があることが明確になりました。

既知の配列の効率的なキャプチャが確立されていることから、Twist Comprehensive Viral Research Panel がパネルデザインに明示的に含まれていないウイルス配列をキャプチャする能力を検証しました。この目的のため、最近報告された RoBat-CoV GCCDC1 シンガポール株の完全なゲノム (Paskey et al. 2020 年、Genbank Accession 番号: MT350598)、ルーセットオオコウモリに感染する新規ベータコロナウイルスを合成しました。キャプチャは、材料及び方法の記載に従って、リボソーム除去のないウイルスゲノム 1,000,000 コピーで実施しました。

図 2 は、リボソーム RNA の事前の除去のない、シンガポールでの RoBat-CoV GCCDC1 ゲノムのスパイクコード領域の高感度キャプチャを示したものです。この領域はパネルデザインにおいてプローブが完全に網羅されておらず、特に注目に値します(図 2)。全体では、ほぼ全てのゲノム (99.8%) が少なくとも 1x の深度までカバーされました。これらの結果は、既存のコロナウイルスの広く多様な配列空間を効果的にターゲットとするパネルデザイン、ミスマッチの大きな広がり許容性のあるハイブリッドキャプチャシステム、そして、プローブのターゲットの配列空間を超えて拡張する配列のフラグメントをキャプチャするキャプチャプローブの能力に起因するものです。

次に、Twist Comprehensive Viral Research Panel が、ヒトに感染することが知られている新たな変異株をキャプチャする能力について、

ヒトへの感染の可能性が示された最近の H1N1 プタインフルエンザ株を用いて検討しました。最近の監視により、2009 年のパンデミックウイルスに関連する新しいヒト感染性 H1N1 株がブタと接する労働者で特定されました (Sun et al. 2020)。Sun ら (2020) が報告した配列から、それぞれ 4 つの HA 及び 4 つのノイラミニダーゼ (NA) セグメントを合成しました。これらの HA と NA のセグメントは H1N1 ウイルス全体で高い配列多様性を示すため選択されました (Kosik et al, 2019)。Twist Comprehensive Viral Research Panel によるハイブリダイゼーションキャプチャ後、8 つの配列すべてが、100% のゲノムカバレッジで 1x 以上でカバーされました(表 1)。図 3 は、インフルエンザ株 A/swine/Hebei/0116/2017(H1N1) の HA セグメントのリード密度を示しています (GenBank Accession 番号: MN416597)。これらのデータによって、Twist Comprehensive Viral Research Panel が最近進化したウイルス配列を十分にキャプチャできることが浮き彫りになりました。

Twist Comprehensive Viral Research Panel のミスマッチ許容性をより明確に定義するため、一連の HA セグメントを合成し、パンデミックインフルエンザ株 A/California/07/2009(H1N1) の HA セグメントから 5% ~ 30% の範囲であらかじめ定めたレベルで人工変異を含めました (Genbank Accession 番号: NC_026433)。図 4 に示すようにリファレンス配列から最大 10% の変異を含む HA セグメントについて、100 万の総マッピングリードで 1x 以上の完全なカバレッジを得ました。ミスマッチが 10% を超す検体では、総キャプチャ量が減少しました。それにもかかわらず、15% と 20% のミスマッチを含む HA セグメントのテンプレートの 70% と 55% がそれぞれ、100 万の総マッピングリードで 1x 以上の深度でキャプチャされました(図 4)。RoBat-CoV GCCDC1 Singapore のキャプチャデータと 2009 H1N1

菌株データと組み合わせると、これらの所見から、Twist Comprehensive Viral Research Panel には、異なるウイルス配列を十分にキャプチャするために必要な高いミスマッチ許容性があることが示唆されます。

上記の図は、Twist Comprehensive Viral Research Panel が、既知のヒト感染性ウイルス、人獣共通感染の可能性のある新規ウイルス株、及び既知のウイルスの高度変異株をキャプチャする能力を示したものです。これらのすべての症例において、バイオインフォマティクス解析からのバイアスによってキャプチャ効率の解析が曖昧になるのを避けるため、使用した正確な菌株との直接的な配列アラインメントにより解析を実施しました。しかしながら、このアプローチはウイルス配列の過去の知見に依存するため、本当に新しいウイルスまたは未知のウイルスを含む分離株の配列決定結果を解析することは不可能です。これらの配列が事前学習なしで検出できるかどうかを試験するために、さらに、K-mer 検索を介して新規のブタインフルエンザ株の1つを、One Codex プラットフォームを用いて分類学データベースで解析しました。図 5 に解析結果を示します。新規の配列内容にもかかわらず、One Codex により、検体がインフルエンザ A/H1N1 であり、これらの配列がブタインフルエンザ株に由来する可能性が高いことが示されました。

サマリ

NGS により、未知の菌種や非常に多様な菌株など偏りのない感染菌種の同定が可能となりますが、診断検体におけるヒト宿主核酸の存在率が高いため、このアプローチの分析感度が低下します。Twist Target Enrichment のワークフローと NGS を組み合わせることで、Twist Comprehensive Viral Research Panel は、未知分析のウイルスの広範な検出および監視のための汎用的なソリューションを提供し、パネルデザインに明示的に組み込まれているウイルス配列の検出を超える検出を可能にします。

材料及び方法

TruSeq ライブラリは、Twist 合成 RNA ウイルスコントロール並びに [Twist Total Nucleic Acids Library Preparation Kit for Viral Pathogen Detection and Characterization protocol](#) を用いて作製しました。簡潔に述べると、RNA ウイルスコントロールを 50 ng のヒトリファレンス RNA (Agilent) のバックグラウンドに添加しました。得られた検体を、Protoscript II First Strand cDNA Synthesis キット (E6560S) 及び New England Biolab の Random Primer 6 (S1230S) を用いて cDNA に変換しました。次に、NEBNext Ultra II Non-Directional RNA Second Strand Synthesis キット (E6111S) を用いて、一本鎖 cDNA を dsDNA に変換しました。次に、Twist Library Preparation Kit with Enzymatic Fragmentation (PN 101059 及び 100401) 及び Unique Dual Indices (UDI) (PN 101307) を用いて、Illumina TruSeq 対応のライブラリを調製しました。最終的に、ウイルス力価 10^6 のライブラリが作製されました。

ハイブリダイゼーションのキャプチャは、Twist Comprehensive Viral Research Panel (PNs 103545, 103547, 103548) 及び [Twist Standard Target Enrichment](#) のワークフローを用いて実施しました。

合成ウイルス	HA Accession 番号	NA Accession 番号	HA の 1x 以上のカバー率 (%)	NA の 1x 以上のカバー率 (%)
1	MN416597	MN416693	100%	100%
2	KP735714	MN416700	100%	100%
3	MN416610	MN416718	100%	100%
4	MN416620	MN416735	100%	100%

表 1: 新規の H1N1 ウイルスからの新規の NA 及び HA 変異の検出 (Sun et al., 2020)。

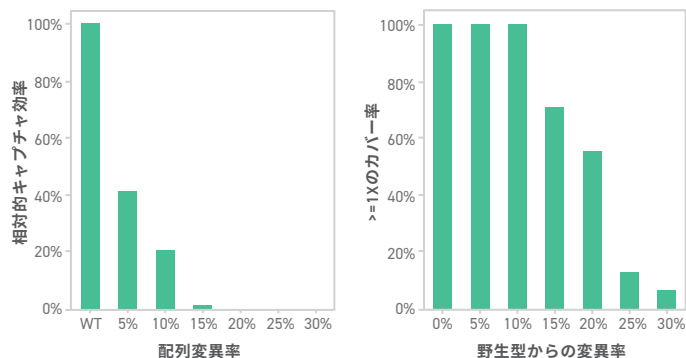


図 4: 野生型からの変異率が及ぼすキャプチャ効率およびカバー率の例。

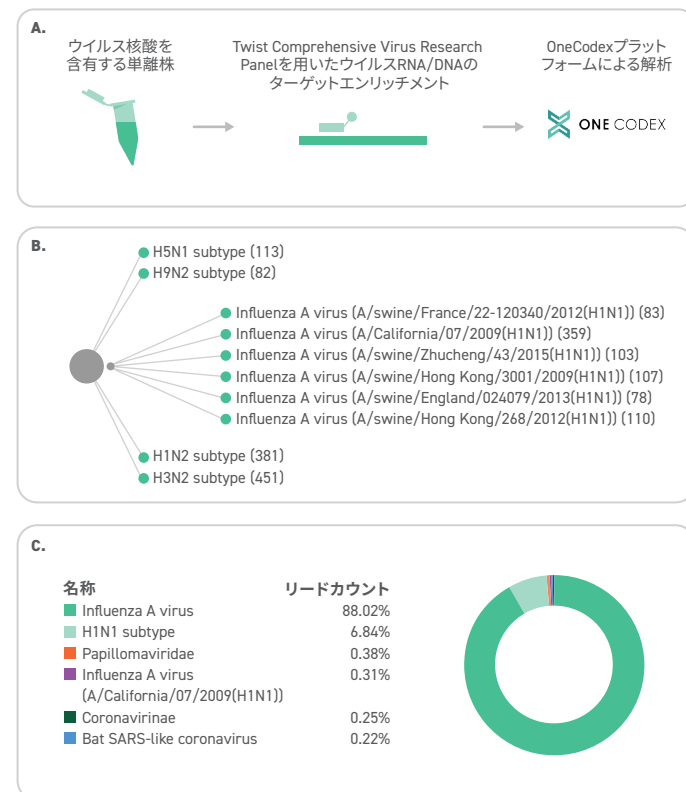


図 5: (A) Twist Comprehensive Viral Research Panel 及び One Codex プラットフォームを用いたウイルス検体のターゲットエンリッチメント及び解析に関する end-to-end のワークフロー。(B) 濃縮された新規ブタインフルエンザ株の One Codex 解析の系統樹。(C) One Codex プラットフォームにより検出された全種の要約



16 時間のハイブリダイゼーションキャプチャ反応ごとに 500 ng のライブラリを使用しました。濃縮後、NextSeq500/550 High Output キットを用いて Illumina NextSeq プラットフォーム上で 2x75 bp のペアエンドリードでライブラリシーケンスしました。パネル内の各ウイルスのリファレンス配列と結合したヒトゲノム (build hg38) で構成されるカスタムゲノムインデックスに対して、BWA でアラインメントを実施しました。特に指定がない限り、全てのデータを、1 検体あたりマッピングされたリード数 100 万にダウンサンプリングしました。

参考文献

- Chiu CY. (2013). Viral pathogen discovery. *Current opinion in microbiology*, 16(4), 468–478. <https://doi.org/10.1016/j.mib.2013.05.001>
- Chomel BB, Belotto A, & Meslin FX. (2007). Wildlife, exotic pets, and emerging zoonoses. *Emerging infectious diseases*, 13(1), 6–11. <https://doi.org/10.3201/eid1301.060480>
- Holland J et al. (1982). Rapid evolution of RNA genomes. *Science (New York, N.Y.)*, 215(4540), 1577–1585. <https://doi.org/10.1126/science.7041255>
- Kosik I & Yewdell JW. (2019). Influenza Hemagglutinin and Neuraminidase: Yin Yang Proteins Coevolving to Thwart Immunity. *Viruses*, 11(4), 346. <https://doi.org/10.3390/v11040346>
- Lam TT et al. (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*, 583(7815), 282–285. <https://doi.org/10.1038/s41586-020-2169-0>
- Mena I et al. (2016). Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. *eLife*, 5, e16777. <https://doi.org/10.7554/eLife.16777>
- Paskey AC et al. (2020). Detection of Recombinant Rousettus Bat Coronavirus GCCDC1 in Lesser Dawn Bats (*Eonycteris spelaea*) in Singapore. *Viruses*, 12(5), 539. <https://doi.org/10.3390/v12050539>
- Sun H et al. (2020). Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. *Proceedings of the National Academy of Sciences of the United States of America*, 117(29), 17204–17210. <https://doi.org/10.1073/pnas.1921186117>
- Woolhouse ME & Gowtage-Sequeria S. (2005). Host range and emerging and reemerging pathogens. *Emerging infectious diseases*, 11(12), 1842–1847. <https://doi.org/10.3201/eid1112.050997>