

NGSを用いたターゲットキャプチャによる新型コロナウイルス (SARS-CoV-2) の検出と分析

はじめに

一本鎖RNAウイルスであるSARS-CoV-2は、世界的流行中の新型コロナウイルス感染症(COVID-19)の原因です。プライマーを用いてSARS-CoV-2ウイルスの対象領域を増幅するRT-PCR (逆転写PCR) は、COVID-19感染症を検出するための迅速な検査法です。しかし、迅速検査に留まらず、ウイルスの経時的な広がりや進化を監視するために、変異を含むウイルスゲノム全体を解析できる高感度検査も必要です。

Twistの核酸ハイブリダイゼーションキャプチャベースのアッセイは、新興ウイルス病原体を検出および追跡する手法として研究者に用いられています。以前、Twistは米国陸軍感染症研究所 (USAMRIID) およびイルミナ社と共同でNGS用に、すべてのヒトウイルスのターゲットエンリッチメントのためのキャプチャプローブからなる大規模パネルPan Viral 1.0を開発しました。これには、1000種以上のヒトウイルスのターゲットエンリッチメント用の60万を超えるプローブが含まれています。Pan Viral 1.0 NGSアッセイパネルは、2015年に西アフリカで大流行したエボラウイルスをキャプチャするためにUSAMRIIDが使用しました。また、2017年にナイジェリアで大流行したサル痘ウイルスを分析するために、セネガルのダカール・パスツール研究所が使用しました。

現在進行中のCOVID-19パンデミックの間、当社はSARS-CoV-2の検出と分析のために、同様でより特異的な手法を用いることにしました。SARS-CoV-2 (MN908947.3)参照ゲノム配列全体をターゲットとするプローブを含む、キャプチャパネルを設計しました。当社はまた、SARS-CoV-2 (MT007544.1)の全長を重複しない6つの断片で合成し、その後RNAに転写しました。SARS-CoV-2 (MT007544.1)ゲノムにはパネルゲノムと比較して、3か所のSNPと1か所のインデル領域が含まれます。合成したウイルス一本鎖RNAを、キャプチャアッセイの陽性コントロールとして、ヒトリファレンスRNAにスパイクイン (添加) しました。このアプリケーションノートには、以下のことが示されています。

1. わずか数コピーのウイルス性物質を100万倍近い濃縮倍率で検出。
2. 濃縮により、1X以上でゲノムの99.9%超をカバー。
3. SARS-CoV-2コントロールMN908947.3中の変異を同定。

材料と方法

合成RNA SARS-CoV-2コントロールを作成するため、参照ゲノム配列MT007544.1 (ncbi.nlm.nih.gov/nuccore/MT007544)およびMN908947.3 (ncbi.nlm.nih.gov/nuccore/MN908947)を6つの非重複断片に分割しました。断片は二本鎖DNAとして合成し、その後RNAに転写されました。次に、SARS-CoV-2 (MT007544.1) RNAコントロールをキャプチャに用いました。

50 ngのヒトリファレンスRNA (Agilent Technologies (740000-41))のバックグラウンドにサンプルあたり1~100万ウイルスコピー数の範囲で合成SARS-CoV-2 RNAをスパイクイン (添加) しました (表1)。ヒトリファレンスRNAのみから構成されたネガティブコントロールも並行して処理しました。次に、RNAサンプルは、NEBのRandom Primer 6 (S1230S)、ProtoScript II First Strand cDNA合成キット (E6560S)、NEBNext Ultra II Non-Directional RNA Second Strand合成キット (E6111S)のランダムプライマーを使用して、cDNAに変換しました。cDNAサンプルは、酵素による断片化を用いたTwist Library Preparation Kit (PN 101059 and 100401)と、Unique Dual Indices (UDI) (PN 101307)を用いて、Illumina TruSeq-compatibleライブラリに変換しました。

濃縮はTwist SARS-CoV-2 Research Panel (PN 102016、PN 102017、またはPN 102018)を用いて、シングルプレックスキャプチャに500 ngのライブラリを使用し、16時間のハイブリダイゼーションで実施しました。濃縮したライブラリは、NextSeq500/550 High Outputキットを使用し、Illumina NextSeqプラットフォームで2x75bpのペアエンドリード法でシーケンスしました。BWAを用いて、SARS-CoV-2ゲノム配列およびヒト参照ゲノム(hg38)に対してアライメントを実行しました。特に記載のない限り、アライメントされたリードはサンプルあたり100万リードにダウンサンプリングされています。

結果

SARS-CoV-2 Research Panelは、RefSeqでMN908947.3として提供されているSARS-CoV-2ゲノムに対して設計されました。キャプチャのポジティブコントロールとして、当社は2つの参照SARS-CoV-2ゲノムから合成RNAコントロールを作成しました。ここでは、Twist SARS-CoV-2 Research Panelと比較して3か所のSNPと1か所のインデルを含むMT007544.1をキャプチャしました。ターゲットエンリッチメントの変異に対する許容性を評価するために、意図的にこれらの変異を合成コントロールに含めました。

キャプチャ前のウイルスの割合は、ウイルススパイクインRNA量をトータルインプット量で割ることで計算しました。キャプチャ後のウイルスの割合は、100万リードまでダウンサンプリングし、duplicate s を削除せずにウイルスリード数を合計リード数で割ることで決定しました。濃縮倍率は、キャプチャ後のウイルス割合をキャプチャ前のウイルス割合で割ることで計算しました (表1)。

ウイルスのコピー数	ウイルス(PG)	キャプチャ前のウイルス割合	合計リード	ウイルスリード (ユニーク%)	キャプチャ後のウイルス割合	濃縮倍率
1,000,000	20	0.04000000%	1,000,000	977,796 (87%)	97.8%	2,444
1,000	0.02	0.00004000%	1,000,000	241,173 (37%)	24.1%	602,933
10	0.0002	0.00000040%	1,000,000	3,506 (18%)	0.351%	876,500
1	0.00002	0.00000004%	1,000,000	394 (33%)	0.039%	985,000
ネガティブコントロール	0	0.00000000%	1,000,000	26 (24%)	0.003%	N/A

表1: ヒトRNAへの異なるスパイクイン (添加) コピー数でのウイルスリードの検出。

ウイルス由来のリードはすべてのポジティブサンプルで検出されました。濃縮範囲は2,444X~985,000Xで、初期スパイクイン濃度に逆相関しました (表1)。低ウイルス力価での少ないユニークリードによる高いduplicate率は、出発物質の量が非常に少ないことから、予想外ではありませんでした。これらの結果は、わずか25000リードを用いて、数コピーのウイルスRNAを検出できることを示しています (表2)。ネガティブコントロール中にも極微量のウイルス性物質を検出したことから、非常に感度の高いアッセイを行う際には、サンプルを物理的に分離することが特に重要となります。コンタミネーションのレベルは、1コピーで見られたものよりも10倍以上低いので、1コピーの陽性結果は単なるバックグラウンドによるものではないと私たちは確信しています。

ウイルスのコピー数	1Xでカバーされるゲノムの割合					
	25Kリード	100Kリード	200Kリード	500Kリード	1Mリード	8Mリード
1,000,000	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
1,000	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
10	17.4%	51.2%	69.5%	83.2%	86.5%	91.5%
1	1.4%	7.8%	16.9%	21.7%	25.3%	27.2%
ネガティブコントロール	0.0%	0.0%	0.7%	1.3%	1.3%	1.4%

表2: 複数の合計リード数を用いた、さまざまなウイルスコピー数でのSARS-CoV-2ゲノムのカバレッジ。

中程度のウイルス力価 (1,000コピー) では、サンプルあたりわずか25,000シーケンスリードで (内在性のポリA鎖を除く) ゲノム全体を1xでカバーできました。低力価 (10コピー) であっても、500kのリードでゲノムの83%が最低1xカバーされ、サンプルあたり8Mのリードではゲノムの90%以上が最低1xの深さでカバーされました (表2)。ネガティブコントロールでもゲノムのごく一部 (8Mリードで最大1.4%) がカバーされていました。ゲノムの1コピーではキャプチャはかなり高く (最大27%)、この力価でのキャプチャはサンプル間のバックグラウンドコンタミネーションを反映したものではないことを示しています。

変異は高効率でキャプチャされました。MT007544.1にはパネルのゲノムと比較して、3か所の1塩基置換が含まれ、そのすべてがマップされたリードの99%以上で検出されました。MT007544.1の10塩基欠失では、欠失サイトにまたがるリードの99%が、正しく欠失を検出したことが分かりました (図1)。

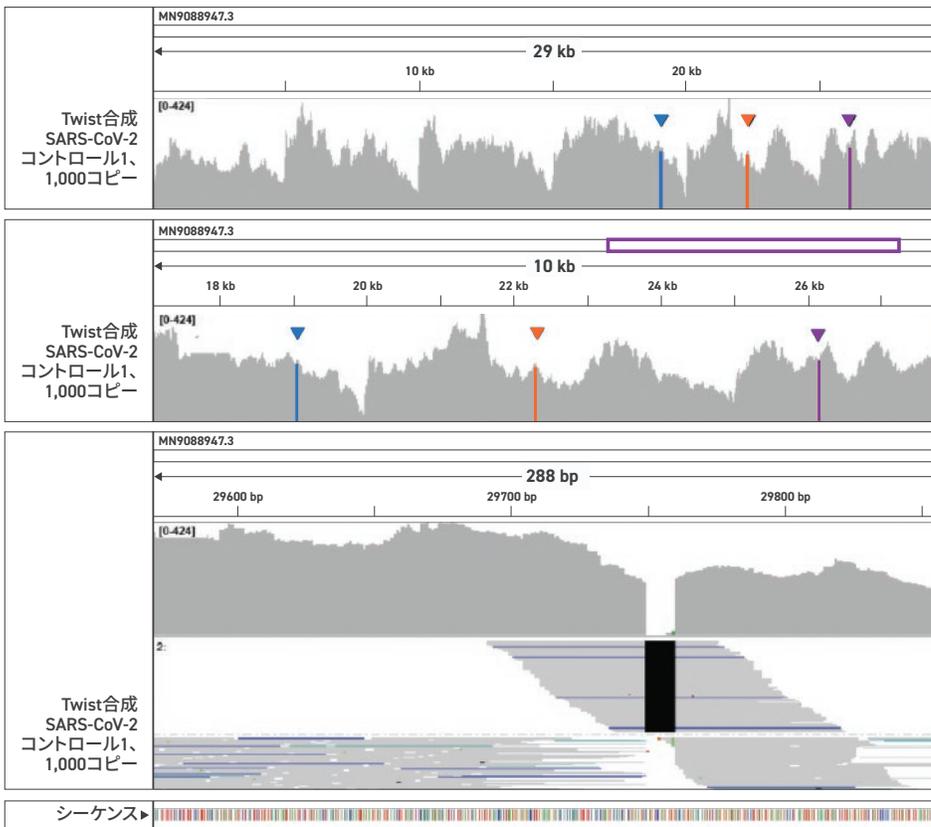


図1：MT007544.1参照配列に基づく、合成コントロールに存在する変異検出。一番上と中央のパネルの1塩基置換を三角形で示しています。一番下のパネルは、パネルゲノムとMT007544.1合成コントロール間の既知の欠失を強調表示しています。

考察

このアプリケーションノートでは、ターゲットキャプチャが非常に高感度で、わずか数コピーのウイルスでも検出できることを示しました。さらに、完全にウイルスゲノム配列をカバーすることで、系統解析を行い、SARS-CoV-2ウイルスの系統と進化の研究が可能になります。しかも、このキャプチャベースの手法は、ウイルス変異に許容性を示し、異なるタイプの変異を同定することに成功しました。NGSを用いたターゲットキャプチャ手法は、SARS-CoV-2ウイルスの検出および解析を同時に行うことができるので、RT-PCR手法の強力な代替手段となるだけでなく、ウイルスの進化と広がりを監視し、人口集団の監視のための貴重なツールになります。

今回は特にSARS-CoV-2を対象としましたが、パネル設計とキャプチャの原理は、より広いターゲットやアプリケーションに応用できます。例えば、当社は呼吸器疾患パネル（PN 103066、103067、103068）やComprehensive Viral Researchパネルを開発しました。呼吸器疾患パネルは、インフルエンザ、コロナウイルス、ライノウイルス、アデノウイルスなどの多くの呼吸器病原体に対応しており、類似した臨床症状を持つ呼吸器病原体を同定し、区別するのに役立ちます。すべてのヒトのウイルス性病原体をターゲットとするように更新されたコンテンツとワークフローを持つ網羅的ウイルス研究パネルは、さらなる検出を進め、新たな脅威に対して迅速に分析できるようにします。

主要製品

製品番号	名称	ストレージ
102019	Twist Synthetic SARS-CoV-2 RNA Control 1 (MT007544.1)	-90~-70°C
102024	Twist Synthetic SARS-CoV-2 RNA Control 2 (MN908947.3)	-90~-70°C
102016：2反応 102017：16反応 102018：96反応	Twist SARS-CoV-2 Research Panel	-25~-15°C

以下の商標の表示は正確であると考えられますが、それを保証するものではありません。Agilent Technologies®、Illumina®、NEB®、NEBNext®、NextSeq®、ProtoScript®、TruSeq®。登録済みまたは登録申請中の名称、商標などは、特に商標マークが付されていない場合でも、適用される法律で保護されていないとは見なされません。

これらの製品は研究専用であり、Twistの供給規約と条件で規定されている追加の使用制限が適用されます。 <https://www.twistbioscience.com/supply-terms-and-conditions>

DOC-001158 REV 3.0 | 最終更新日：2021年2月22日

改定	日付	説明
3.0	2021年2月22日	ページ3:Pan Viral 2.0を網羅的研究パネルに置き換えました。