

# Seraseq ctDNA Mutation Mix v4 を用いた Agilent Avida DNA ワークフローによる cfDNA NGS 解析性能の評価

## 著者

Margherita Corioni, PhD,  
Xiaomu Chem, MS, Heng Wang,  
MS, Grace Zhao, PhD,  
R&D Scientists,  
Agilent Technologies, Inc.

## Collaborators from LGC Clinical Genomics

Andrew Anfora,  
Associate Director of Product  
Management

David Merriam,  
Commercial Director

Yves Konigshofer,  
Director of Technology  
Development

## はじめに

次世代シーケンシング (NGS) によるセルフリー DNA (cfDNA) の解析により、がん生物学の理解が大きく前進し、低侵襲性の分子診断の開発が促進されました。しかし、従来の NGS ワークフローは多くの場合、複雑で時間がかかります。特に精製のステップでは DNA が失われる可能性があるため、低インプット量のアプリケーションでは困難が伴います。

新しいハイブリダイゼーション技術を活用した Agilent Avida DNA ワークフローでは、ハイブリダイゼーションによるエンリッチメント前のライブラリ調製時に、増幅と精製を行う必要がありません。この簡素化されたワークフローにより、PCR で生じるバイアスを軽減し、回収率を損なわずに、ハイブリダイゼーションを 1 時間に短縮できました。これは、高い感度と特異性が要求される低インプット DNA の解析には不可欠です。さらに、PCR 増幅によるバイアスを最小限に抑えることができれば、より正確なコピー数変異 (CNV) 検出に好ましいと考えられます。

正確な対立遺伝子頻度を持つ、安定かつ包括的なリファレンス DNA の使用は、cfDNA NGS の解析性能の実証と、製品全体を通じた品質管理のモニタリングに不可欠です。このアプリケーションノートでは、がん関連のバリエーションを包括的にカバーし、従来のリファレンス ctDNA と比べてバックグラウンドノイズが少なく、サイズ分布とライゲーション効率の点で実際の cfDNA と非常に類似していることから、リファレンスに Seraseq ctDNA Mutation Mix v4 を使用しています。<sup>1</sup>

Avida テクノロジーは、ターゲット DNA のプラス (+) 鎖とマイナス (-) 鎖の両方をキャプチャし、エラー修正と正確な分子カウントに分子バーコード (unique molecular identifier : UMI) を使用します。このアプリケーションノートでは、UMI を使用し、

- それぞれ異なる深度でシーケンスされた 3 種類の Cancer パネル (26 kb、345 kb、2.7 Mb) で、様々なインプット量での DNA 分子の回収効率を調べました。
- 3 種類のパネルによるバリエーション検出、および中規模パネル (Avida DNA Expanded Cancer パネル : 345 kb) を用いて詳細な検出評価を行いました。これにより、腫瘍学研究関連バリエーションに対する、適切なシーケンス深度のバランスの目安になります。

このアプリケーションノートでは、特定のインプット量とシーケンス深度における検出可能な予想バリエーションアレル頻度 (variant allele frequency : VAF) と分子回収効率に関するシーケンスメトリクスを紹介しています。これらの情報は、他のAvidaカタログパネルやカスタムパネルに適用することができます。

## 方法

### サンプル

Seraseq ctDNA Mutation Mix v4 の 0 % (野生型)、0.5 %、および 5 % VAF は、製造元提供の原液 (製品番号 0710-3101、0710-3099、および 0710-3100、SeraCare) を使用しました。このリファレンス DNA に含まれる変異は、[SeraCare のウェブページ](#)で確認できます。0.2 %、1 %、2 % VAF サンプルは、0.5 % および 5 % VAF 原液を 0 % 野生型で希釈して調製しました。SeraCare ctDNA Mutation Mix v4 の濃度は、Agilent 4200 TapeStation System (製品番号 G2991A)、Agilent D1000 DNA ScreenTape (製品番号 5067-5582)、および対応する試薬 (製品番号 5067-5583) を使用して確認および定量化しました。

さまざまなインプット量のリファレンス DNA を、アジレントの 3 つの Avida Cancer カタログパネルを使用してターゲットエンリッチメントしました :

- Avida DNA Focused Cancer パネル (製品番号 5280-0050)
- Avida DNA Expanded Cancer パネル (製品番号 5280-0047)
- Avida DNA Discovery Cancer パネル (製品番号 5280-0044)

各条件で少なくとも 2 回のテクニカルレプリケートを実施しました。

### パネル詳細

- Avida DNA Expanded Cancer パネル : 105 種類のがん関連遺伝子をカバーしています。これには、SeraCare ctDNA v4 リファレンス DNA の転座 (TL)、および 7 個のコピー数変化 (CNV) が含まれています。このパネルは、サイズおよびリファレンス DNA との対応の面で、この検討に適していました。

- Avida DNA Focused Cancer パネル : 最小のがんパネル (26 kb) で、がん研究における 14 種類の主要な遺伝子から、主にホットスポットとエクソンに焦点を絞ったがんターゲットセットの高い回収率を実現するように最適化されています。
- Avida DNA Discovery Cancer パネル : Avida Cancer パネルの中で最大 (2.7 Mb) で、がん研究における探索と評価のために、680 種類を超える主要な遺伝子とバイオマーカーをカバーしています。

Focused パネルは 23 種類、Discovery パネルは 66 種類の、SeraCare ctDNA v4 リファレンス DNA の SNV およびインデルをそれぞれカバーしています。これらのパネルサイズでのアレル頻度 (allele frequency : AF) 検出の例を示します。

### ライブラリ調製とターゲットエンリッチメント

Focused、Expanded、または Discovery Cancer パネルを使用して、Avida DNA キット (G9418A、Agilent) 推奨のプロトコルに従って、リファレンス ctDNA を調製しました。インデックス PCR 産物は、プールの前に、D1000 ScreenTape を使用して定量しました。

### データ解析

Avida アダプタには UMI が含まれています。これは、PCR 増幅の前にサンプルライブラリ内の各分子をタグ付けするために使用されるランダムオリゴヌクレオチドの短いシーケンスです。UMI バーコードに基づくシーケンスアライメントは、偽陽性のバリエーションコールの割合を減らし、元のインプット DNA に存在する「真」の変異検出を改善します。この方法では、PCR 重複を利用してシーケンスリードをユニーク UMI ごとにグループ化し、リードの頻度と品質スコアを考慮して正確なコンセンサス配列を決定します。このプロセスでは、データの品質を向上させるため、シーケンスエラーの修正にシーケンスの重複を利用します。シーケンス量、重複率、およびバリエーション検出の関係については結果の項で説明します。

シーケンスリードは start/stop の位置情報とインライン UMI を使用してシングルモードおよびデュプレックスモードで重複除去し、bwa-mem を使用してゲノムにアラインメントしました。

リファレンス SeraCare ctDNA Mutation Mix v4 の SNV と indel は、VarDict (ver. 1.5.0) を使用して同定しました。CNV は CNVkit (ver. 0.9.8) を使用して解析し、融合は GeneFuse (ver. 0.6) を使用してコールしました。

この検討では、検出率 (感度) は、SeraCare ctDNA Mutation Mix v4 リファレンスおよびパネルでカバーされる共通したバリエーションのコールの成功率で定義されます。偽陽性率 (特異度) は、野生型 (0 %) リファレンスからの変異コールとして定義されます。

## 結果

### シーケンスのためのリード数選択とエラー修正のための UMI ベースの重複除去

Avida DNA ワークフローに基づくこの検討では、10 ~ 20 ng のリファレンス Seraseq ctDNA Mutation Mix v4 を使用しました。キャプチャパネルとして、Avida Cancer カタログパネル Focused (26 kb)、Expanded (345 kb)、Discovery (2.7 Mb) の各パネルを使用しました。シーケンス量は、パネルのサイズと DNA インプット量に基づいて調整しました。表 1 は、本検討におけるシーケンスリードペアの数とデザインコンテンツの平均カバレッジ中央値 (average raw median coverage) との相関関係の一例を示したものです。

シーケンスリードペア量は、バリエーション検出分析に不可欠なパラメータであり、特定のバリエーション検出カットオフに対して十分なカバレッジを確保するために適切に選択する必要があります。通常、インプット DNA 量が多く、低 AF を検出したい場合は、より多いリード数が必要になります。方法のセクションで説明したように、シーケンスリードの数は、回収されたユニーク UMI ごとに最も正確なコンセンサス配列を選択する UMI 重複除去プロセスの精度とも直接関連します。

重複除去前の重複率が 80 % に近づくか 80 % を超えるように、キャプチャ工程におけるインプット量とパネルサイズを考慮して、サンプルシーケンス深度を選択する必要があります (UMI シングルモード重複除去の場合)。これにより、サンプル内のユニーク UMI を網羅的にスクリーニングし、重複除去時に正確に UMI ユニークコンセンサス配列を割り当てることができます。

インライン UMI 重複除去を使用した低頻度バリエーションの検出では、重複率が 80 % を大きく下回ると、分子のかなりの部分がシーケンス決定されずバリエーションコールに寄与しないことがわかりました。つまり、シーケンスが不十分な場合、ライブラリの複雑さ (サンプル内の固有の分子数) が部分的にしか見えなくなり、レアなバリエーションが検出されない可能性が高くなります。重複率が 80 % を大幅に超えるようにシーケンスの深度を増やした場合、よりユニークな分子回収への効果は限定的で、結果としてライブラリ構成の理解への影響も限られています。しかし、追加のシーケンスは無意味ではなく、これにより、各 UMI ファミリーのサイズは大きくなるので、エラー修正をサポートできます。

表 1. Agilent Avida Cancer パネルを使用した、ユニーク分子コンセンサス補正 (UMI シングルモード重複除去) のデザインコンテンツの中央値カバレッジと重複率に対するシーケンス深度の影響

Panel	Size [kb]	Input [ng]	Sequencing Read Pairs [M]	Raw Median Coverage [k]	Duplication [%]
Avida DNA Focused Cancer	26	20	15	40	83
		20	10	27	76
		20	5	13	60
Avida DNA Expanded Cancer	345	10	50	14	88
		10	40	10.6	78
		10	20	5.3	62
	345	20	80	21.6	81
		20	50	14	72
		20	40	10.8	66
		20	20	5.4	47
Avida DNA Discovery Cancer	2700	20	400	19	80
		20	200	9	64
		20	100	4.7	45

シーケンス量はリードペアで表されています。総リード数はリードペア数の 2 倍です。

## 20 ng のリファレンスインプットによる 3 つの Cancer パネルの一般的な性能

Seraseq ctDNA Mutation Mix v4 リファレンス20 ng から Avida DNA Expanded Cancer パネルでキャプチャし調製したライブラリの場合、80 M のシーケンスリードペア（合計 160 M のリード）でパネル全体の平均カバレッジは 4000x でした（表 2）。カバレッジ深度が高いことに加えて、カバレッジも均一であり、標的塩基の 95 % 以上で重複除去後カバレッジが設計平均値の半分以上になっています。これは、エクソン領域とイントロン領域の両方を含むパネル全体の領域の 95 % 以上が 2000x を超えるカバレッジがあることを意味します。

つまり、20 ng のインプットで 0.2 % AF の場合、どの遺伝子座でも平均コピー数は 95 % の領域で少なくとも 4 コピー、平均でバリエーションのコピー数は 8 コピーになることが期待されます。Expanded パネルでカバーされているリファレンス Seraseq ctDNA Mutation Mix v4 の 60 種類の SNV と indel について調べたところ、20 ng のインプットにおける 0.2 % AF の検出率は 94 % となりました（表 2）。

0.2 % VAF リファレンス DNA（23 バリエーションをカバー）20 ng から Avida DNA Focused パネル（26 kb）でキャプチャし調製したライブラリの場合、10M リードペア量（総リード数 20 M）で SNV の検出率は 100 % でした（表 2）。平均 UMI カバレッジは 6000x（表 2）で、バリエーションコピー数は 12 個と予測され、100 % の検出率は予測と一致します。

0.5 % VAF リファレンス 20 ng を使用し、Avida DNA Discovery パネル（2.7 Mb）でキャプチャし調製したライブラリの場合では、約 170 M のリードペア量（総リード数 340 M）で SNV と indel（66 バリエーションをカバー）の検出率は 94 % でした（表 2）。Expanded および Focused の各パネル検出と同じ手順に従い、インプット量、パネルサイズ、およびシーケンス量を考慮すると、平均 UMI カバレッジ 2600x となり、平均 13 コピーのバリエーションが期待されました。この比較的大きなパネルでは、得られたシーケンス深度での重複率は 68 % にとどまっています（表 2）。そのため、より深くシーケンスすることで、UMI の回収率が増加し、感度が向上する可能性があります。

また、野生型（0 %）リファレンスを同じインプット量・同等のシーケンス深度で分析した場合、低い偽陽性率が示されました。Focused パネル・20 ng インプット・10 M リードペア（合計リード 20 M）の場合は 1 つの偽陽性コールが、Expanded パネル・20 ng インプット・80 M リードペア（合計リード 160 M）の場合は 4 つの偽陽性のコールがそれぞれありました。その他の組み合わせのパネル、インプット量、シーケンス深度では、偽陽性は検出されませんでした。

表 2. 20 ng の Seraseq ctDNA Mutation Mix v4 リファレンス DNA を使用した際の、Agilent Avida DNA Focused Cancer、Avida DNA Expanded Cancer、および Avida DNA Discovery Cancer パネルの一般的な性能メトリクス：UMI 重複除去後カバレッジとバリエーションの検出感度

Panel	Size [kb]	Input [ng]	SNV/Indels Covered	Expected VAF [%]	Read Pairs [M]	Raw Reads Coverage [k]	UMI-Deduplicated Coverage [k]	Duplication [%]	Detected VAF [%]	Detection Sensitivity [%]
Avida DNA Focused Cancer	26	20	23	0.2	10	27	6	78	0.2	100
Avida DNA Expanded Cancer	345	20	60	0.2	80	21	4.3	81	0.2	94
Avida DNA Discovery Cancer	2700	20	66	0.5	172	7.5	2.6	68	0.5	95

ここで示されているシーケンス量（リードペア）は、十分なエラー修正とユニークな分子の回収を可能にし、適切なサンプル解析（重複率）を行うために設定されました。シーケンス量はリードペアで表されています。総リード数はリードペア数の 2 倍です。

## Avida DNA Expanded Cancer パネルにおけるインプット量とシーケンス量の各種組み合わせによる検出

低頻度のバリエーションを探する場合、正確に必要なシーケンス量を決定することが必要です。これは、インプット量、目的とする AF、キャプチャパネルのサイズの組み合わせに基づいて行われます。理想的には、重複率の割合が 80 % 以上になるよう、平均カバレッジ中央値を十分に高くすることが求められます。これは、網羅的な数のユニークな分子をスクリーニングし、回収されたユニークな分子が正確に同定されることで、対象のサンプルを正確に調べることができるということを意味します (表 2 を参照)。

このアプリケーションノートでは、シーケンス深度は様々な要因によって制限される可能性があることを考慮し、ダウンサンプリングデータ例でそれぞれの重複率と AF 検出限界への影響を示しています。これにより、リード数がバリエーション検出率に与える影響を示し、可能な限り精度の高い解析計画を立案するのに役立ちます (表 1 および 3A を参照)。

アッセイへの DNA インプット量が制限されていたり、シーケンス量を見積もるのが困難である場合もありますが、これらはすべて、メソッドの検出能力に影響する変数になり得ます。表 3A および 3B は、Avida DNA Expanded Cancer パネルにおける様々な組み合わせのインプット量・シーケンス量での VAF 検出率のガイドラインを示しています。

表 3.

A. 20/40 ng の Seraseq ctDNA Mutation Mix v4 リファレンスから、Avida DNA Expanded Cancer パネルを使用して得られたバリエーションアレル頻度 (VAF) 検出

Input [ng]	Sequencing Read Pairs [M]	Duplication [%]	VAF [%]		
			0.1	0.2	0.5
40	160	82	92	98	
	50	26	86	96	
	25	13	73	91	
	13	6	51	76	
20	80	81		94	100
	50	51		92	100
	25	25		87	98
	13	13		69	96

実際のシーケンス量がエラー修正と重複率に与える影響、および 0.1 ~ 0.5 % の低 VAF でのバリエーション検出感度

B. 1 ~ 10 ng の超低インプット Seraseq ctDNA Mutation Mix v4 リファレンスから、Avida DNA Expanded Cancer パネルを使用して得られたバリエーションアレル頻度 (VAF) 検出

Input [ng]	Sequencing Read Pairs [M]	Duplication [%]	VAF [%]			
			0.5	1	2	5
10	46	81	96			
	32	80		98		
	39	78			99	
5	30	86	82			
	57	93		97		
	68	93				100
3	39	92		98		
1	33	92			98	
	11	94				93

重複率とエラー修正に必要なシーケンスリードの例と 0.5 ~ 5 % の範囲の VAF の検出感度 : Seraseq ctDNA Mutation Mix v4 リファレンスには、Expanded パネルでカバーされる 7 個の CNV と 7 個の転座が含まれています。両方のインプット量での CNV 検出率は 100 % でした (表 3C)。CNV の平均 fold change は、0.2 % の場合で約 1.1 倍、0.5 % の場合で約 1.3 倍でした。転座 (TL) 検出率はそれぞれ 93 % と 79 % でした。野生型 (0 %) リファレンス DNA の場合、同じインプット量とシーケンス深度では偽陽性は検出されませんでした。

C. CNV と転座 (TL)

Variant Type	VAF [%]	Input [ng]	Sequencing Read Pairs [M]	Duplication [%]	Detection Sensitivity [%]
CNV	0.2	20	80	90	100
CNV	0.5	10	66.6	81	100
Translocation	0.2	20	80	90	93
Translocation	0.5	10	66.6	81	79

## Avida DNA Expanded Cancer パネルの ターゲットカバレッジ統計

Avida DNA Expanded Cancer パネルを使用した、リファレンス DNA のエンリッチメント性能の一般的な概要について、ターゲットごとのカバレッジの統計を示します。図 1 は 20 ng のリファレンスインプットを、Expanded パネルでエンリッチメントし、80 M のリードペア（総リード数 160 M）でシーケンスした場合のターゲット領域のカバレッジです。ターゲット領域の 100% で 1000x のカバレッジ、97% で 2000x、87% で 3000x を超えるカバレッジがありました（図 1A）。パネル全体の平均カバレッジは 4425x でした。

図 1B は、Expanded パネル内の主要ながん関連遺伝子の平均カバレッジです。エクソン領域は全体で、少なくとも平均 3000x のカバレッジがありました。

図 1C は、図 1B の Expanded パネルの遺伝子リスト中で、リファレンスに含まれるバリエーションの検出の例を示しています。20 ng のインプット量、80 M リードペア（総リード数 160 M）とした場合の 0.5% および 0.2% の期待アレル頻度（AF）について検証しました。

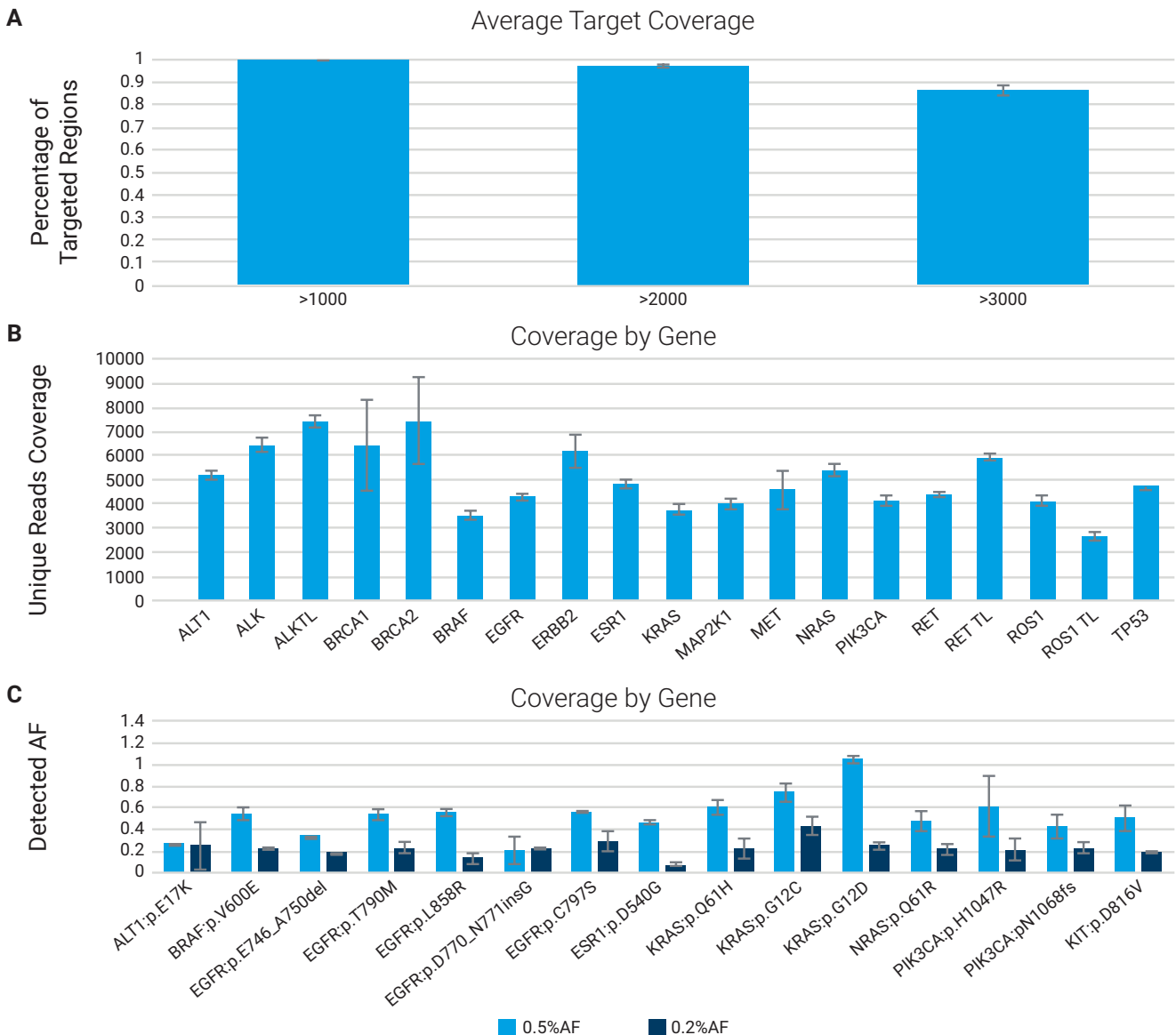


図 1. 20 ng Seraseq ctDNA Mutation Mix v4 リファレンス DNA を用いた Agilent Avida DNA Expanded Cancer パネルにおけるエンリッチメントによる遺伝子の回収数  
**A.** 80 M のシーケンスリードペア（総リード数 160 M）から計算されたターゲット領域別のカバレッジ中央値統計（UMI 重複除去後のユニークリードカバレッジ）。カバレッジ中央値は 6 回のテクニカルレプリケートの結果です。**B.** 80 M のシーケンスリードペア（総リード数 160 M）から計算された主要なシグネチャがん関連遺伝子の遺伝子毎に集計されたユニークリードカバレッジ（「TL」は、転座検出を強化するために含めた非エクソンターゲットコンテンツを示します）。**C.** 80 M のシーケンスリードペア（総リード数 160 M）における、期待値 AF が 0.5% および 0.2% のリファレンスのバリエーション検出：平均 AF は、20 ng インプット量からの 2 回の繰り返しを行い算出しました。

## まとめ

cfDNA を用いた NGS の分析性能評価は重要なステップであり、このプロセスには高品質のリファレンス DNA の使用が不可欠です。Seraseq ctDNA Mutation Mix v4 は、多様なバリエーションと低いバックグラウンドノイズにより、Avida DNA ワークフローによるバリエーション検出の分析性能を評価するための効率的で包括的なツールであることが明らかとなりました。

実証されたように、Avida DNA ワークフローは、リファレンス Seraseq ctDNA では 10 ~ 20 ng という柔軟なインプット範囲で、0.5 % および 0.2 % VAF で信頼性の高いバリエーションコールを可能にするのに十分なユニークな DNA 分子を安定して回収できました。

シーケンスデータをダウンサンプリングし、推奨されるリードペア数よりも少ない数でバリエーション検出を行うことで、Avida テクノロジーが正確で有益なサンプルの測定値を提供することが実証されました。これにより、インプット量とシーケンス量の見積りに柔軟性を持たせることができ、貴重で限られたサンプルの分析が実施しやすくなります。

この分析で示されているように、Agilent Avida システムは、プロトコルの変更や、性能の低下を招くことなく、柔軟な範囲のインプット量に容易に対応することができます (表 3 参照)。そのため、インプット量が限られ、分子の回収効率が重要となるリキッドバイオブシーアアプリケーションなどの検討に最適なソリューションとなります。

## 参考文献

1. Ruminski, L.; et al. Next generation liquid biopsy reference material performance across NGS assays and platforms. Presented at AACR Annual Meeting, **2024**. Available for download at: <https://digital.seracare.com/aacr-liquid-biopsy>

### [お問い合わせ窓口]

アジレント・テクノロジー株式会社

本社 / 〒192-8510 東京都八王子市高倉町9-1

●カスタムコンタクトセンター ☎ 0120-477-111

mail: email\_japan@agilent.com

※仕様は予告なく変更する場合があります。

※本資料掲載の製品はすべて試験研究用です。

診断目的にご利用いただくことはできません。

G250730

[www.agilent.com/genomics/genomics-jp](http://www.agilent.com/genomics/genomics-jp)

© Agilent Technologies, Inc. 2024

本書の一部または全部を画面による事前の許可なしに複製、改変、翻訳することは、著作権法で認められている場合を除き、法律で禁止されています。

Printed in Japan, December 12, 2024

5994-7953JAJP